Aus K. Linkwitz, U. Hangleiter (Eds.): "High Precision Navigation", Springer-Verlag 1989

Motion Estimation in Image Sequences

Norbert Diehl*, **Hans Burkhardt**
Arbeitsbereich Technische Informatik I, Technische Universität
Hamburg-Harburg

Abstract

This contribution presents an efficient method to estimate the motion parameters of planar rigid objects in space. The problem is formulated as a model-based estimation problem which is solved by an iterative minimization strategy based on ordinary grey-scale images. Therefore no a priori knowledge as corresponding points or the displacement vector fields are required.

The advantages of the algorithm - a so called modified Newton-Raphson algorithm and its combination with Quasi-Newton methods - are the large region of stability, the high, image-bandwidth-adaptive convergence rate and the low numerical expense within each iteration step. Therefore the algorithm may be used for near realtime applications.

Furthermore the algorithm is relatively insensitive to additive noise because whole image areas are used. The variance of the motion parameters can be reduced further if the problem of motion estimation is formulated as maximum-likelihood estimation. Under a few assumptions the maximum-likelihood estimation can be interpreted as a generalized correlation, i.e. the correlation of the appropriately prefiltered images.

1. Introduction

The problem of motion estimation has become of great interest in several areas of image processing such as motion compensated image coding, remote sensing, robotics or autonomous vehicles [Huang 1981, Huang 1983, Nagel 1981, Nagel 1983, Nagel 1985, Musmann et al. 1985, Bähr 1985]. So far most algorithms for three-dimensional motion estimation use corresponding points [Roach und Aggarwal 1980, Huang und Tsai 1981, Tsai et al. 1982, Tsai and Huang 1984a, Tsai and Huang 1984b, Fang and Huang 1984, Longuet-Higgins 1981, Yen and Huang 1983] which have to be known a priori or they use the displacement vector field [Prazdny 1980, Adiv 1985, Kanatani 1986]. Both methods have a very high numerical complexity and are very sensitive to noise.

In this contribution the motion estimation problem is solved using model-based estimation techniques based on ordinary grey scale images. The advantage of these iterative algorithms is the combination of good performance features such as a large region of stability and a high, image-bandwidth-adaptive convergence rate with low numeric expense within each iteration step. Furthermore the algorithms are relatively insensitive to additive noise because whole areas of the

* Now with: AEG Forschungsinstitut, Sedanstraße 10, D-7900 Ulm, West-Germany

images are used. Similar algorithms may be found in [Spoer 1987, Netravali and Salz 1985, Lenz 1986].

The problem of motion estimation may be devided into two subproblems. First the scene has to be segmented into the different moving objects and the background where also effects of occlusion have to be taken into consideration. In a second step the motion parameters of the objects have to be estimated. This contribution deals with this second problem under the simplifying assumption of only one moving object. To achieve quite general results methods to estimate the parameters of arbitrary motion and therefore of general coordinate transforms are developed. Three-dimensional motion of a rigid planar patch is treated as a special example.

The parameter estimation algorithm is based on a simplified signal model. A camera looks at a rigid moving object in the three-dimensional space. Two grey-scale images $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$ describe the motion of this object in front of a uniform background where no occlusion effects occur:

$$I_1(\mathbf{x}) = S(\mathbf{x}) + N_1(\mathbf{x}), \qquad I_2(\mathbf{x}) = S(\mathbf{x}') + N_2(\mathbf{x})\,. \tag{1.1}$$

$S(\mathbf{x})$ und $S(\mathbf{x}')$ are the projections of the moving object onto the image plane. $N_1(\mathbf{x})$ and $N_2(\mathbf{x})$ are additive noise terms. Thus equation 1.1 describes the connection between the projection of the object in position 1 ($I_1(\mathbf{x})$) and the same object in position 2 ($I_2(\mathbf{x})$) (cf. Fig. 1.1).

Generally the connection between the coordinates $\{\mathbf{x}\}$ and $\{\mathbf{x}'\}$ is given by a transform

$$\mathbf{x}' = \mathbf{h}(\mathbf{x}, \mathbf{T})\,, \tag{1.2}$$

with the motion vector $\mathbf{T}$. The transform $\mathbf{h}(\mathbf{x}, \mathbf{T})$ should be unique and invertible in the considered image region. For simplicity it should also be $\mathbf{x}' = \mathbf{h}(\mathbf{x}, \mathbf{T} = 0) = \mathbf{x}$ for $\mathbf{T} = \mathbf{0}$. If the image formation is described by central projection under the assumption of a rigid planar patch the coordinate transform is given by

$$x' = \frac{(1 + a_1)\,x + a_2\,y + a_3}{a_7\,x + a_8\,y + 1}\,, \qquad y' = \frac{a_4\,x + (1 + a_5)\,y + a_6}{a_7\,x + a_8\,y + 1}\,, \tag{1.3}$$

with the motion vector $\mathbf{T} = \mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8)^T$. For parallel projection the coordinate transform may be described as

$$x' = (1 + c_1)\,x + c_2\,y + c_3\,, \qquad y' = c_4\,x + (1 + c_5)\,y + c_6\,, \tag{1.4}$$

with $\mathbf{T} = \mathbf{c} = (c_1, c_2, c_3, c_4, c_5, c_6)^T$.

The parameters of the coordinate transforms in the image plane are unique. But it is very hard to answer the question how to get the true three-dimensional motion parameters of the object. Using the central projection and two successive images there are normally two solutions [Tsai and Huang 1982]. Often one solution can be excluded using slight a priori knowledge such as a rough estimate of the object position at the beginning of motion. If three successive images are used the motion parameters are unique up to a common scale factor of the translation vector [Tsai and Huang 1984b].

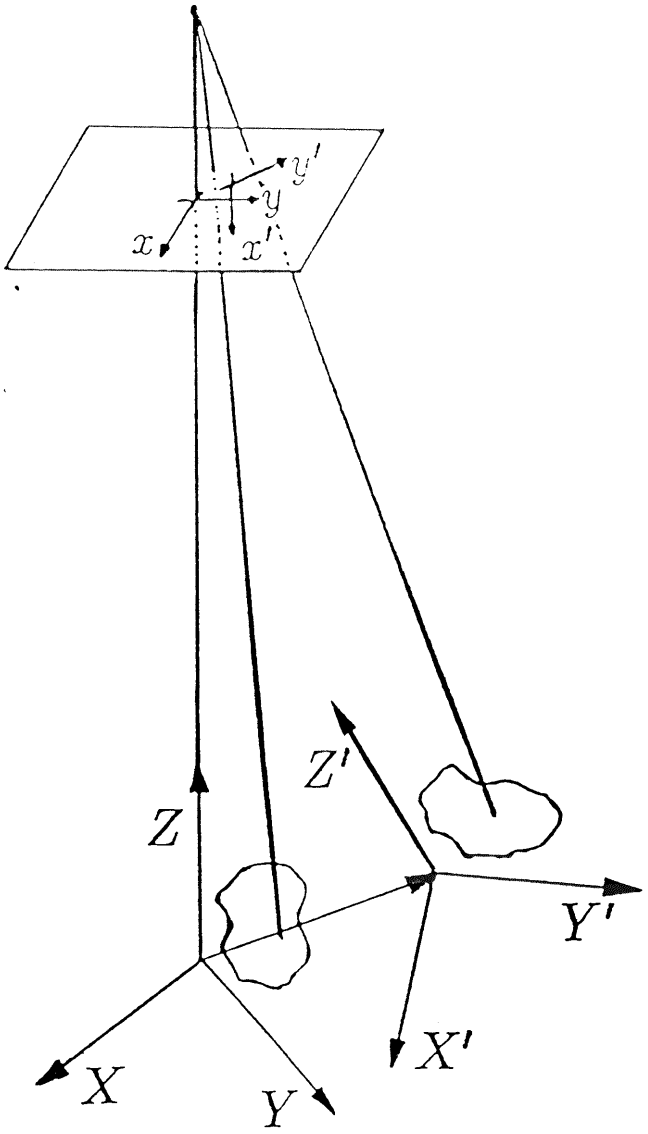

Fig. 1.1: Moving of an object in the three-dimensional space
and central projection onto the image plane

If parallel projection is used the ambiguity is larger. There are usually four solutions up to the common scale factor [Lenz 1986]. These ambiguities mainly result from the fact that the parallel projection cannot describe the magnification of object parts coming closer and the reduction of those parts moving away.

2. Model-based parameter estimation

In this contribution the main aspect will be the efficient estimation of the parameter vector $\mathbf{T}$ of the coordinate transform $\mathbf{h}(\mathbf{x}, \mathbf{T})$. Therefore the idea of model-based parameter estimation will be used and a so called modified Newton-Raphson algorithm and its combinations with Quasi-Newton methods will be developed. To get an appropriate estimate $\widehat{\mathbf{T}}$ of the parameter vector $\mathbf{T}$ the motion is modelled by a model image $I_m(\mathbf{x}, \widehat{\mathbf{T}})$. This image results from $I_1(\mathbf{x})$ by the same transform $\mathbf{h}(\mathbf{x}, \mathbf{T})$ like $I_1(\mathbf{x})$ but now with the model parameter vector $\widehat{\mathbf{T}}$. In a second step the

difference between $I_m(\mathbf{x}, \widehat{\mathbf{T}})$ an $I_2(\mathbf{x})$ is described by an appropriate error criterion $J\{e(\widehat{\mathbf{T}}\}$ which is minimized to get the optimal estimate $\widehat{\mathbf{T}} = \mathbf{T}^*$.

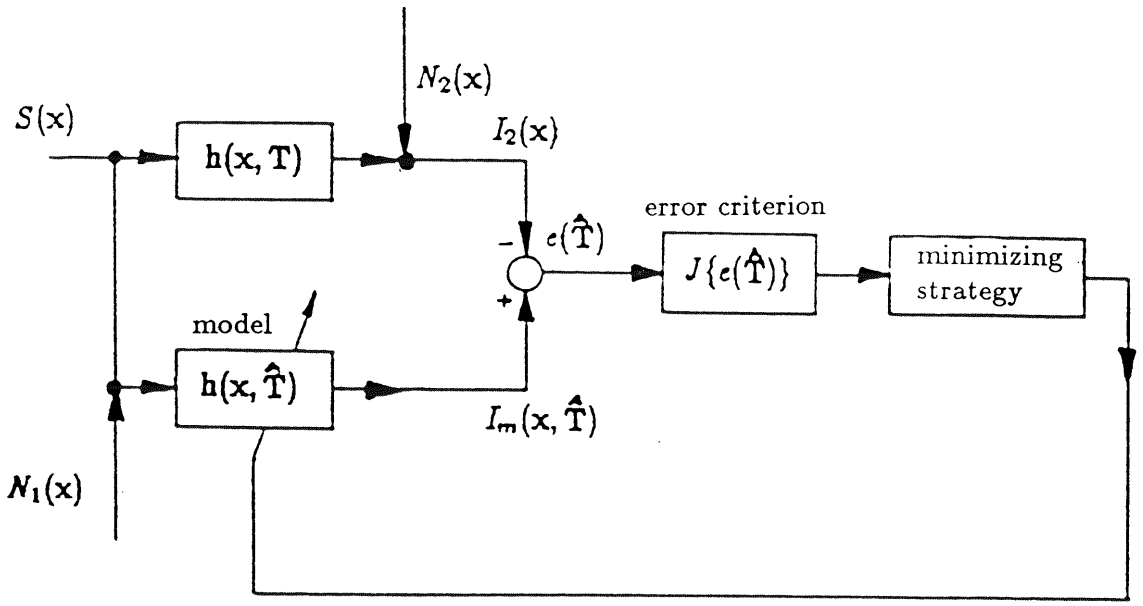

Fig. 2.1: Structure of model-based parameter estimation

2.1. The error criterion

The error criterion specifies the quality of the estimate. Therefore knowledge of the signal and noise statistics should be incorporated. The most general approach is given by a maximum likelihood estimate. At the moment the simpler error function

$$J\{e(\widehat{\mathbf{T}})\} = \frac{1}{2} E\{(I_m(\mathbf{x}, \widehat{\mathbf{T}}) - I_2(\mathbf{x}))^2\}, \tag{2.1}$$

i.e. the variance of the model error is chosen. The expectation value is realized by summing over the interesting image region. For stationary signals $J\{e(\widehat{\mathbf{T}})\}$ equals the negative cross correlation funktion of $I_m(\mathbf{x}, \widehat{\mathbf{T}})$ and $I_2(\mathbf{x})$ plus an additive constant.

This error criterion has to meet some requirements. First the contents of the images $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$ must have a sufficiently high similarity such that the error function has a unique minimum. Second it is assumed that $J\{e(\widehat{\mathbf{T}})\}$ is at least twice differentiable at the optimum, i.e. the Hessian of the error function must exist. For ill conditioned images where the criterion has no dominant parabolic shape near the optimum appropriate filtering may be used.

2.2. The minimizing strategy

The minimizing strategy determines the velocity and robustness of the estimate. The region of stability, the convergence rate and the numerical complexity of the minimizing algorithm are very important and decide if near realtime applications are possible.

A global search for the optimal parameter vector $\mathbf{T}^*$ is unrealistic because it results in the calculation of a multidimensional cross-correlation function in conjunction with a tremendeous numerical complexity. Instead local, iterative methods are used, especially a so called modified Newton-Raphson algorithm and its combination with Quasi-Newton methods.

2.2.1. The modified Newton-Raphson algorithm

The modified Newton-Raphson algorithm originally has been developed for one-dimensional problems such as time delay estimation and stereoscopic range finding [Mesch 1982, Burkhardt and Moll 1978, Burkhardt and Moll 1979]. Then it was used to estimate simultaneously rotation and translation in image sequences [Burkhardt and Diehl 1986, Diehl and Burkhardt 1986] before it has been extended to general transforms $\mathbf{h}(\mathbf{x}, \mathbf{T})$ [Diehl 1987, Diehl 1988]. The modified Newton-Raphson algorithm differs from the normal Newton algorithm at one important point. Instead of using the Hessian $\mathbf{H}(\widehat{\mathbf{T}})$ at the actual iteration point $\widehat{\mathbf{T}} = \widehat{\mathbf{T}}_k$ the Hessian is always used at the optimum $\widehat{\mathbf{T}} = \mathbf{T}^*$. For pure translation ($\mathbf{T} = (d_1, d_2)^T$) the algorithm is given by the iteration

$$\widehat{\mathbf{T}}_{k+1} = \widehat{\mathbf{T}}_k - \mathbf{H}^{-1}(\mathbf{T}^*)\, \mathbf{g}(\widehat{\mathbf{T}}_k). \tag{2.2}$$

The special advantage of this algorithm is the possibility to compute the Hessian $\mathbf{H}(\mathbf{T}^*)$ only once before starting the iteration.

An interpretation will be given in the following if additive noise is neglected. Imagine that the error criterion equals the negative cross-correlation of $I_m(\mathbf{x}, \widehat{\mathbf{T}})$ and $I_2(\mathbf{x})$. At the optimum when $I_m(\mathbf{x}, \widehat{\mathbf{T}})$ is identical to $I_2(\mathbf{x})$ the cross-correlation equals the autocorrelation of $I_2(\mathbf{x})$. Therefore the Hessian can be calculated at the beginning as the curvature of the autocorrelation instead of the cross-correlation at the unknown optimum. Thus the numerical complexity is quite low. Only the gradient vector $\mathbf{g}(\widehat{\mathbf{T}}_k)$ has to be updated within the iteration.

If these results are extended to the general transform $\mathbf{h}(\mathbf{x}, \mathbf{T})$ a few small changes have to be made. Especially a moving coordinate system $\{\mathbf{x}_k\}$ and an additional motion vector $\widetilde{\mathbf{T}}$ have to be introduced [Diehl 1987, Diehl 1988]. This vector connects the coordinate system $\{\mathbf{x}_k\}$ of iteration step k with the coordinate system $\{\mathbf{x}_{k+1}\}$ of iteration step $k+1$:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{h}(\mathbf{x}, \widehat{\mathbf{T}}_k) \\ \mathbf{x}_{k+1} &= \mathbf{h}(\mathbf{x}_k, \widetilde{\mathbf{T}}_{k+1}) = \mathbf{h}(\mathbf{h}(\mathbf{x}, \widehat{\mathbf{T}}_k), \widetilde{\mathbf{T}}_{k+1}) = \mathbf{h}(\mathbf{x}, \widehat{\mathbf{T}}_{k+1}). \end{aligned} \tag{2.3}$$

Thus the coordinate system $\{\mathbf{x}_{k+1}\}$ results directly from $\{\mathbf{x}\}$ by the vector $\widehat{\mathbf{T}}_{k+1}$ or indirectly from $\{\mathbf{x}_k\}$ by $\widetilde{\mathbf{T}}_{k+1}$. This is only possible if the set of transforms $\mathbf{h}(\mathbf{x}, \mathbf{T})$ is closed.

Now the model image $I_m(\mathbf{x}, \widehat{\mathbf{T}}_{k+1})$ is generated indirectly :

$$I_m(\mathbf{x}, \widehat{\mathbf{T}}_{k+1}) = I_1(\mathbf{x}_{k+1}) = I_1(\mathbf{h}(\mathbf{h}(\mathbf{x}, \widehat{\mathbf{T}}_k), \widetilde{\mathbf{T}}_{k+1})) \tag{2.4}$$

and the error criterion has to be minimized with respect to $\tilde{\mathbf{T}}$. Then the new iteration is given by

$$\tilde{\mathbf{T}}_{k+1} = -\tilde{\mathbf{H}}^{-1}\,\tilde{\mathbf{g}}(\hat{\mathbf{T}}_k) \tag{2.5}$$

for the innovation at step $k+1$ and

$$\hat{\mathbf{T}}_{k+1} = \mathbf{f}(\hat{\mathbf{T}}_k, \tilde{\mathbf{T}}_{k+1}) \tag{2.6}$$

connecting $\tilde{\mathbf{T}}_{k+1}$ and $\hat{\mathbf{T}}_k$ to the whole motion vector $\hat{\mathbf{T}}_{k+1}$.

The Hessian $\tilde{\mathbf{H}}$ at the optimum is given by

$$\tilde{\mathbf{H}} = E\left\{\left(\frac{\partial \mathbf{h}(\mathbf{x},\mathbf{T})}{\partial \mathbf{T}}\right)^T \frac{\partial I_1(\mathbf{x})}{\partial \mathbf{x}} \left(\frac{\partial I_1(\mathbf{x})}{\partial \mathbf{x}}\right)^T \frac{\partial \mathbf{h}(\mathbf{x},\mathbf{T})}{\partial \mathbf{T}}\right\}\Bigg|_{\mathbf{T}=\mathbf{0}} \tag{2.7}$$

if additive noise is neglected. This expression does not explicitely depend on the actual motion vector. Therefore the Hessian actually can be calculated before starting the iteration. The gradient vector is given by

$$\begin{aligned}\tilde{\mathbf{g}}(\hat{\mathbf{T}}_k) = E\Big\{ &\Big(I_m(\mathbf{x},\hat{\mathbf{T}}_k) - I_2(\mathbf{x})\Big) \\ &\times \left(\left(\frac{\partial I_2(\mathbf{x})}{\partial \mathbf{x}}\right)^T \left(\frac{\partial \mathbf{x}_k}{\partial \mathbf{x}}\right)^{-1} \left(\frac{\partial \mathbf{x}_{k+1}}{\partial \tilde{\mathbf{T}}_{k+1}}\right)\Bigg|_{\tilde{\mathbf{T}}_{k+1}=\mathbf{0}}\right)^T\Big\},\end{aligned} \tag{2.8}$$

which has to be updated within each iteration step.

To get a better idea of the algorithm the formulas for a planar rigid patch using central projection are given in detail. With

$$\frac{\partial \mathbf{x}'}{\partial \mathbf{T}} = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 & -x^2 & -xy \\ 0 & 0 & 0 & x & y & 1 & -xy & -y^2 \end{pmatrix}, \tag{2.9}$$

and

$$\begin{aligned}\frac{\partial I_1(\mathbf{x}')}{\partial \mathbf{T}}\Bigg|_{\mathbf{T}=\mathbf{0}} = \Bigg(&\frac{\partial I_1(\mathbf{x})}{\partial x}x, \frac{\partial I_1(\mathbf{x})}{\partial x}y, \frac{\partial I_1(\mathbf{x})}{\partial x}, \frac{\partial I_1(\mathbf{x})}{\partial y}x, \\ &\frac{\partial I_1(\mathbf{x})}{\partial y}y, \frac{\partial I_1(\mathbf{x})}{\partial y}, -\frac{\partial I_1(\mathbf{x})}{\partial x}x^2 - \frac{\partial I_1(\mathbf{x})}{\partial y}xy, -\frac{\partial I_1(\mathbf{x})}{\partial x}xy - \frac{\partial I_1(\mathbf{x})}{\partial y}y^2\Bigg)^T\end{aligned} \tag{2.10}$$

the Hessian can be calculated according to

$$\tilde{\mathbf{H}} = E\left\{\left(\frac{\partial I_1(\mathbf{x}')}{\partial \mathbf{T}}\right)\left(\frac{\partial I_1(\mathbf{x}')}{\partial \mathbf{T}}\right)^T\right\}\Bigg|_{\mathbf{T}=\mathbf{0}}. \tag{2.11}$$

The gradient vector may be written in the simplified form

$$\tilde{\mathbf{g}}(\hat{\mathbf{T}}_k) = \mathbf{B}_k^{MNR}\,\mathbf{z}(\hat{\mathbf{T}}_k) \tag{2.12}$$

with the simple weighting matrix $\mathbf{B}^{MNR}$ which is independent of the image content and the coordinates. The vector $\mathbf{z}(\widehat{\mathbf{T}}_k)$ is given by

$$\mathbf{z}(\widehat{\mathbf{T}}_k) = E \left\{ \begin{array}{c} \left(I_m(\mathbf{x},\widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \frac{\partial I_2(\mathbf{x})}{\partial x} x^2 \\ \left(I_m(\mathbf{x},\widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \frac{\partial I_2(\mathbf{x})}{\partial x} xy \\ \left(I_m(\mathbf{x},\widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \frac{\partial I_2(\mathbf{x})}{\partial x} x \\ \left(I_m(\mathbf{x},\widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \frac{\partial I_2(\mathbf{x})}{\partial x} y \\ \left(I_m(\mathbf{x},\widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \frac{\partial I_2(\mathbf{x})}{\partial x} \\ \left(I_m(\mathbf{x},\widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \frac{\partial I_1(\mathbf{x})}{\partial y} x^2 \\ \left(I_m(\mathbf{x},\widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \frac{\partial I_1(\mathbf{x})}{\partial y} xy \\ \left(I_m(\mathbf{x},\widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \frac{\partial I_1(\mathbf{x})}{\partial y} x \\ \left(I_m(\mathbf{x},\widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \frac{\partial I_1(\mathbf{x})}{\partial y} y \\ \left(I_m(\mathbf{x},\widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \frac{\partial I_1(\mathbf{x})}{\partial y} \end{array} \right\} . \tag{2.13}$$

To compute this vector the image difference between $I_m(\mathbf{x},\widehat{\mathbf{T}})$ and $I_2(\mathbf{x})$ has to be built and has to be multiplied by the weighted partial derivatives. Further details ar given in [Diehl 1988].

The transform

$$\widehat{\mathbf{T}}_{k+1} = \mathbf{f}(\widehat{\mathbf{T}}_k, \widetilde{\mathbf{T}}_{k+1}), \tag{2.14}$$

which connects $\widehat{\mathbf{T}}_k$ and $\widetilde{\mathbf{T}}_{k+1}$ is given by

$$\begin{pmatrix} \widehat{a}_1 & \widehat{a}_2 & \widehat{a}_3 \\ \widehat{a}_4 & \widehat{a}_5 & \widehat{a}_6 \end{pmatrix}_{k+1} = \begin{pmatrix} \widetilde{a}_1 & \widetilde{a}_2 & \widetilde{a}_3 \\ \widetilde{a}_4 & \widetilde{a}_5 & \widetilde{a}_6 \end{pmatrix}_{k+1} \begin{pmatrix} \widehat{a}_1 & \widehat{a}_2 & \widehat{a}_3 \\ \widehat{a}_4 & \widehat{a}_5 & \widehat{a}_6 \\ \widehat{a}_7 & \widehat{a}_8 & 1 \end{pmatrix}_k \tag{2.15}$$

and

$$(\widehat{a}_7, \widehat{a}_8)_{k+1} = (\widetilde{a}_7, \widetilde{a}_8, 1)_{k+1} \begin{pmatrix} \widehat{a}_1 & \widehat{a}_2 \\ \widehat{a}_4 & \widehat{a}_5 \\ \widehat{a}_7 & \widehat{a}_8 \end{pmatrix}_k . \tag{2.16}$$

Thus using these formulas and

$$\widetilde{\mathbf{T}}_{k+1} = -\widetilde{\mathbf{H}}^{-1}\, \widetilde{\mathbf{g}}(\widehat{\mathbf{T}}_k) \tag{2.17}$$

the modified Newton-Raphson algorithm is described completely.

The advantages of the modified Newton-Raphson algorithm are manifold. Because the Hessian is always used at the optimum the advantages of the gradient algorithm and those of the normal Newton algorithm are combined. The modified Newton-Raphson algorithm has the large region of stability like the gradient algorithm. Like the normal Newton algorithm it has good signal adaptive properties. The Hessian which can be interpreted as the curvature of the autocorrelation function of the images adjusts to the image bandwidth. Therefore the convergence rate of the

Euclidean error norm $\|\widehat{\mathbf{T}}_k - \mathbf{T}^*\|$ is at least of second order, independently of the chosen signals. Another important advantage is the low numerical complexity because of the ability to calculate the Hessian before starting the iteration. Therefore only the gradient vector has to be updated.

Nevertheless there are a few problems. If there is significant noise the Hessian calculated according to eq. 2.7 does not exactly represent the second derivatives of the error function. Therefore the convergence rate may be worse. Furthermore it is sometimes disadvantageous that the set of transforms $\mathbf{h}(\mathbf{x}, \mathbf{T})$ has to be closed. Quasi-Newton algorithms and especially their combination with the modified Newton-Raphson algorithm can succesfully deal with these problems.

2.2.2. Quasi-Newton methods

Quasi-Newton methods [Gill et al. 1981] try to build up the Hessian at the actual iteration point $\widehat{\mathbf{T}}_k$ only using values of the gradient vector which have already been calculated. Thus the curvature of the error criterion is approximated only by first derivatives. The iteration scheme is given by

$$\widehat{\mathbf{T}}_{k+1} = \widehat{\mathbf{T}}_k - \mathbf{C}_k^{-1}\, \mathbf{g}(\widehat{\mathbf{T}}_k)\,. \tag{2.18}$$

The Quasi-Newton matrix $\mathbf{C}_k$ converges to the true Hessian at the optimum if for example the BFGS-update [Gill et al. 1981] is used:

$$\mathbf{C}_{k+1} = \mathbf{C}_k + \frac{1}{\mathbf{g}(\widehat{\mathbf{T}}_k)^T \mathbf{s}_k} \mathbf{g}(\widehat{\mathbf{T}}_k)\mathbf{g}(\widehat{\mathbf{T}}_k)^T + \frac{1}{\mathbf{y}_k^T \mathbf{s}_k} \mathbf{y}_k \mathbf{y}_k^T\,. \tag{2.19}$$

$\mathbf{s}_k = \widehat{\mathbf{T}}_{k+1} - \widehat{\mathbf{T}}_k$ describes the difference of two succesive estimates and $\mathbf{y}_k = \mathbf{g}(\widehat{\mathbf{T}}_{k+1}) - \mathbf{g}(\widehat{\mathbf{T}}_k)$ the change in the gradient. The gradient vector is given by

$$\mathbf{g}(\widehat{\mathbf{T}}_k) = E\left\{ \left(I_m(\mathbf{x}, \widehat{\mathbf{T}}) - I_2(\mathbf{x})\right) \left(\left(\frac{\partial I_2(\mathbf{x})}{\partial \mathbf{x}}\right)^T \left(\frac{\partial \mathbf{x}'}{\partial \mathbf{x}}\right)^{-1} \left(\frac{\partial \mathbf{x}'}{\partial \widehat{\mathbf{T}}}\right) \right)^T \right\} \Bigg|_{\widehat{\mathbf{T}}=\widehat{\mathbf{T}}_k} \tag{2.20}$$

which again can be simplified for the case of central projection to

$$\mathbf{g}(\widehat{\mathbf{T}}_k) = \mathbf{B}_k^Q\, \mathbf{z}(\widehat{\mathbf{T}}_k) \tag{2.21}$$

with $\mathbf{z}(\widehat{\mathbf{T}}_k)$ of equation 2.13. The weighting matrix $\mathbf{B}^Q$ is similar to B^{MNR}. It is important to note that only the gradient vector has to be calculated whithin each iteration step. Therefore the numerical complexity - characterized by the number of operations on each pixel - is as low as for the gradient or modified Newton-Raphson algorithm. Because the Hessian is approximated the convergence rate only is superlinear. Furthermore the region of stability is quite low.

2.2.3. A combined algorithm

A very efficient algorithm can be realized if the modified Newton-Raphson and the Quasi-Newton algorithm are combined. The iteration is given by

$$\widehat{\mathbf{T}}_{k+1} = \widehat{\mathbf{T}}_k - \mathbf{M}_k^{-1}\, \mathbf{g}(\widehat{\mathbf{T}}_k)\,. \tag{2.22}$$

Here a matrix $\mathbf{M}$ is used as a combination of the Hessian $\widetilde{\mathbf{H}}$ at the optimum and a Quasi-Newton matrix $\mathbf{M}^{BFGS}$

$$\mathbf{M}_{k+1} = \frac{1}{2}\left(\widetilde{\mathbf{H}} + \mathbf{M}_{k+1}^{BFGS}\right). \tag{2.23}$$

The matrix $\mathbf{M}_{k+1}^{BFGS}$ results from $\mathbf{M}_k$ of the preceeding iteration step k by the BFGS-Quasi-Newton update. If the estimate $\widehat{\mathbf{T}}$ is far away from the true value $\mathbf{T}^*$ and therefore the Quasi-Newton method would diverge $\mathbf{M}_k^{BFGS} = \mathbf{0}$ is used. Thus the algorithm keeps stable and the innovation is twice the innovation of the modified Newton-Raphson algorithm, which improves the far away convergency. Near the optimum $\mathbf{M}_{k+1}^{BFGS}$ converges to the Hessian $\widetilde{\mathbf{H}}$ of the optimum such that the combined algorithm turns into the modified Newton-Raphson algorithm. An additional improvement can be reached if at the beginning only the translation parameters a_3 and a_6 are estimated under the assumption that the other parameters are zero. More deatails are given in [Diehl 1988].

Fig. 2.2: Picture of a woman digitized by 512 × 512 pixel

2.3. Tests using real image data

The algorithm has been tested with real image data. Therefore different images $I_1(\mathbf{x})$ have been transformed in the computer by a definite eight parametric motion vector $\mathbf{T} = \mathbf{a}$ to generate $I_2(\mathbf{x})$. Using two corresponding images $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$ the motion vector $\mathbf{T}$ was estimated. Fig. 2.2, 2.3 and 2.4 show a typical experiment. The motion vector connecting $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$ was $\mathbf{T} = (0.166, -1736, -5.0, 0.2409, -0.0152, -7.0, -0.00577, 0.0)^T$. This describes a rotation of 10° around the camera main axis, an inclination of 30° around the y-axis and a translation assuming that the object is very close to the camera. Fig 2.4 illustrates how a_1, a_2, a_3 and a_7 of the estimation vector $\widehat{\mathbf{T}}$ converge to the the true vector $\mathbf{T}$ using the combined algorithm. Further results are given in [Diehl 1988].

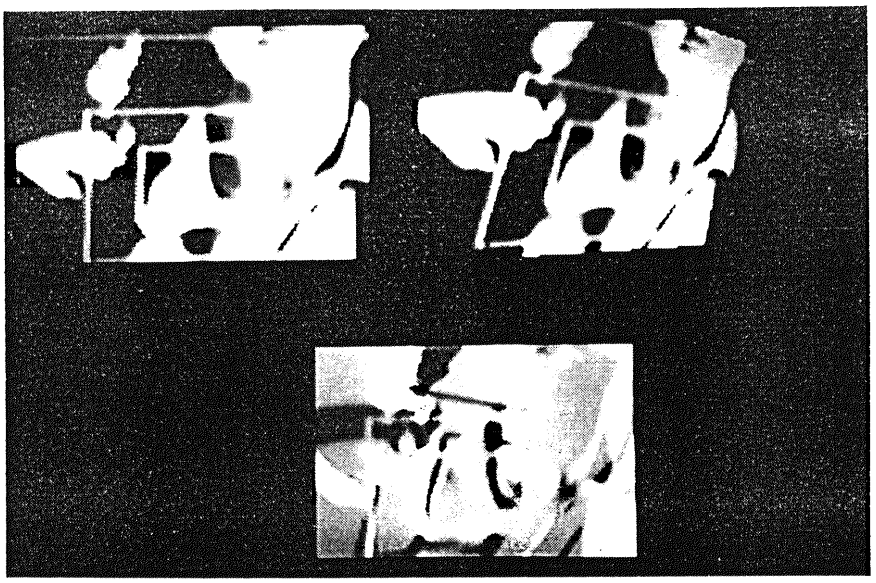

Fig. 2.3: Picture of a woman with $I_1(\mathbf{x})$, $I_2(\mathbf{x})$ and the difference image before starting the iteration

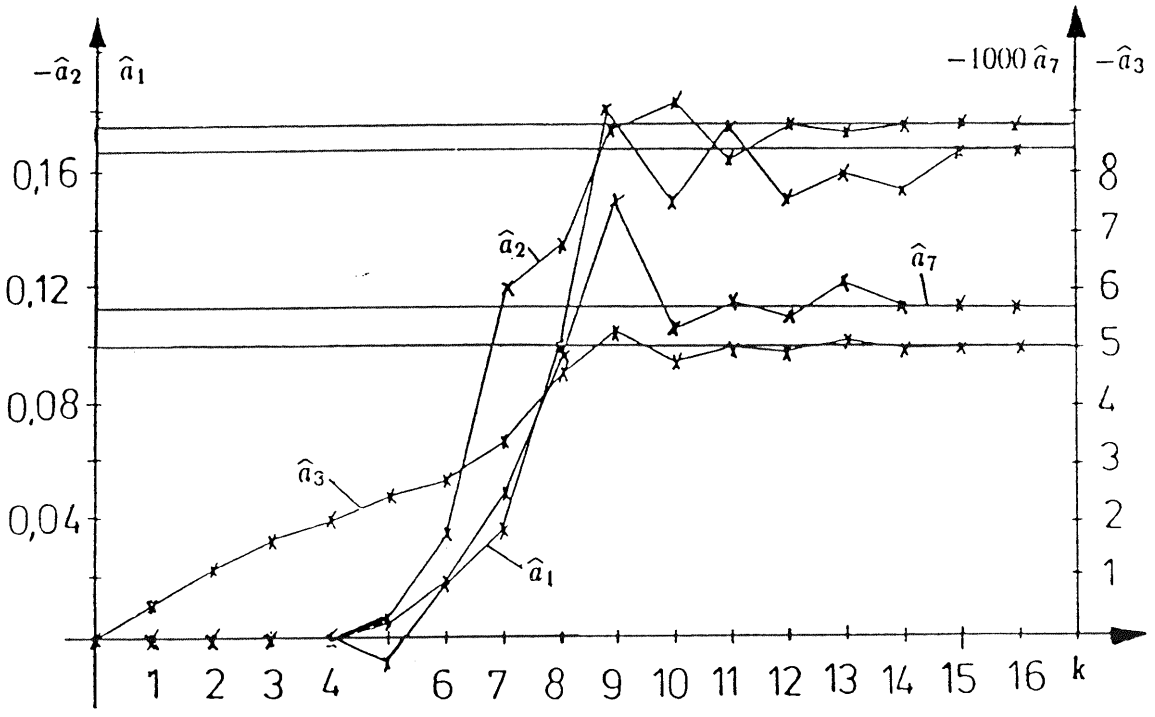

Fig. 2.4: The estimation values $\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$ and $\hat{a}_7$ using the combined method

3. The maximum-likelihood estimation of the motion parameters

3.1. The maximum-likelihood estimation as generalized correlation

So far the ad hoc error criterion $J\{e(\widehat{\mathbf{T}})\}$ of eq. 2.1, i.e. the variance of the model error was minimized. In this part the results of the maximum-likelihood estimation of the motion parameters will be given. To deal with this problem a few simplifying assumptions have to be made. The image $\mathrm{S}(\mathbf{x})$ and the additive noise terms $N_1(\mathbf{x})$ and $N_2(\mathbf{x})$ are assumed to be stationary, zero mean gaussian processes which are not correlated to each other. Furthermore the coordinate transform $\mathbf{h}(\mathbf{x},\mathbf{T})$ may be a linear transform, e.g. the affine transform $\mathbf{x}' = \mathbf{C}\,\mathrm{x} + \mathrm{d}$. Under these assumptions the likelihood-function $p(I_1(\mathbf{x}), I_2(\mathbf{x})|\mathbf{T})$ may be calculated. The value $\mathbf{T}$ which maximizes this function is called the maximum-likelihood estimate $\widehat{\mathbf{T}}_{ML}$.

It can be shown that the maximum-likelihood estimation can be interpreted as a generalized correlation near the optimum [Diehl 1988]. Similar results are known from time delay estimation [Knapp and Carter 1976] or pure translation [Beyer 1985]. Thus the estimate $\widehat{\mathbf{T}}_{ML}$ is the value which maximizes the cross-correlation function of the two images $\widehat{I}_1(\mathbf{x})$ and $\widehat{I}_2(\mathbf{x})$ which result from the original images $\widehat{I}_1(\mathbf{x})$ and $\widehat{I}_2(\mathbf{x})$ by linear filtering using the impulse response $w(\mathbf{x})$ calculated from $|W_{ML}(\mathbf{f})|^2$:

$$\widehat{I}_1(\mathbf{x}) = w_{ML}(\mathbf{x}) \otimes I_1(\mathbf{x}), \quad \widehat{I}_2(\mathbf{x}) = w_{ML}(\mathbf{x}) \otimes I_2(\mathbf{x}). \tag{3.1}$$

$\otimes$ denotes convolution. The prefilter is given by

$$|W_{ML}(\mathbf{f})|^2 = \frac{G_{S'S'}(\mathbf{f})}{G_{S'S'}(\mathbf{f})\left(G_{N_1'N_1'}(\mathbf{f}) + G_{N_2N_2}(\mathbf{f})\right) + G_{N_1'N_1'}(\mathbf{f})\,G_{N_2N_2}(\mathbf{f})} \tag{3.2}$$

$G_{S'S'}(\mathbf{f})$ is the power-spectral density of the image content $S'(\mathbf{x})$ of image $I_2(\mathbf{x})$ and $G_{N_2N_2}(\mathbf{f})$ the power-spectral density of the noise $N_2(\mathbf{x})$. $G_{N_1'N_1'}(\mathbf{f})$ denotes the power-spectral density of $N_1(\mathbf{x})$ which has been moved in direction of $I_2(\mathbf{x})$ under the assumption that the matrix $\mathbf{C}$ of the affine transform may be approximately known. If $N_1(\mathbf{x})$ and $N_2(\mathbf{x})$ have the same statistics $G_{N_1'N_1'}$ may be replaced by $G_{N_2N_2}$. Fig. 3.1 shows the structure of the generalized correlation.

The interpretation of the maximum-likelihood estimation as generalized correlation has an interesting impact. Because the error criterion of eq. 2.1 may be interpreted as a cross-correlation the algorithms developed in chapter 2 can be used directly for the maximum-likelihood estimation. Only the original images $I_1(\mathbf{x})$ und $I_2(\mathbf{x})$ have to be replaced by the prefiltered images $\widehat{I}_1(\mathbf{x})$ and $\widehat{I}_2(\mathbf{x})$.

3.2. Properties of the maximum-likelihood estimation

The maximum-likelihood estimate has some interesting properties. For high signal-to-noise ratios

$$G_{SS}(\mathbf{f}) \gg G_{NN}(\mathbf{f})$$

$|W_{ML}(\mathbf{f})|^2$ is given by

$$|W_{ML}(\mathbf{f})|^2 \approx \frac{1}{G_{N_1'N_1'}(\mathbf{f}) + G_{N_2N_2}(\mathbf{f})}. \tag{3.3}$$

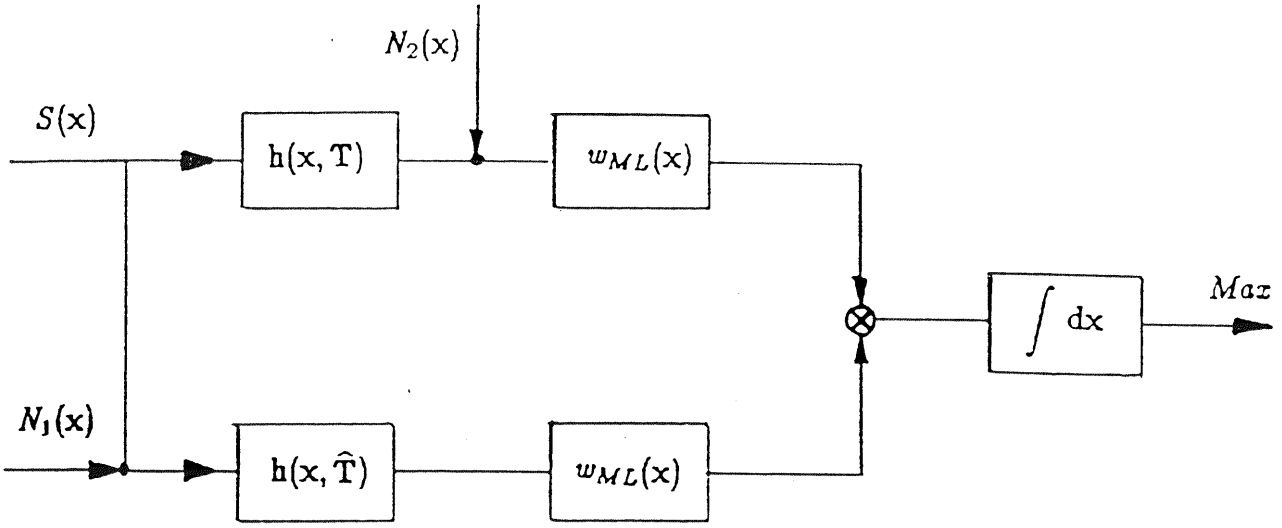

Abb. 3.1: The maximum-likelihood estimation as generalized correlation

If white noise, i.e.

$$G_{N_1'N_1'}(\mathbf{f}) = \sigma^2_{N_1} \qquad \text{and} \qquad G_{N_2N_2}(\mathbf{f}) = \sigma^2_{N_2} \tag{3.4}$$

is assumed, the prefilter is constant:

$$|W_{ML}(\mathbf{f})|^2 \approx \frac{1}{\sigma^2_{N_2} + \sigma^2_{N_1}} = const. \tag{3.5}$$

For this special case the generalized correlation is equal to the correlation of the original images. Thus the ad hoc error criterion of eq. 2.1 equals the maximum-likelihood criterion if stationary signals, white noise and a high signal-to-noise ratio are assumed.

Furthermore the maximum-likelihood estimate reaches the Cramer-Rao bound if the noise terms and the signal are not cross-correlated [van Trees 1976, Diehl 1988]. The Cramer-Rao bound which is the bound for the lowest variance of the estimation values can be expressed by the Hessian of the prefilterd signals:

$$var\{\mathbf{T}^*\} = \frac{1}{XY} E\{\hat{\mathbf{H}}\}^{-1} . \tag{3.6}$$

3.3. Tests of the maximum-likelihood estimation

The maximum-likelihood estimation has been tested using real image data and different additive noise. Details are given in [Diehl 1988]. Fig. 3.2 and 3.3 illustrate the standard deviation of a translation in x-direction. If the signal-to-noise ratio is low and the noise is highly correlated a

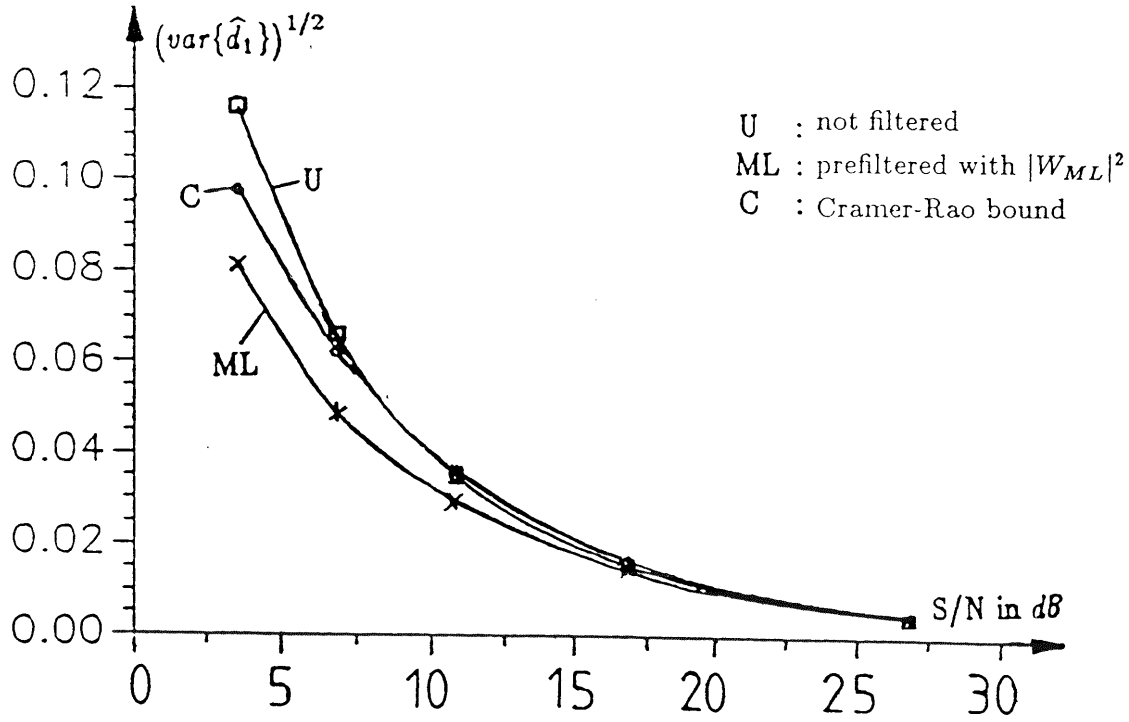

Abb. 3.2: Standard deviation of $\hat{d}_1$ using white noise

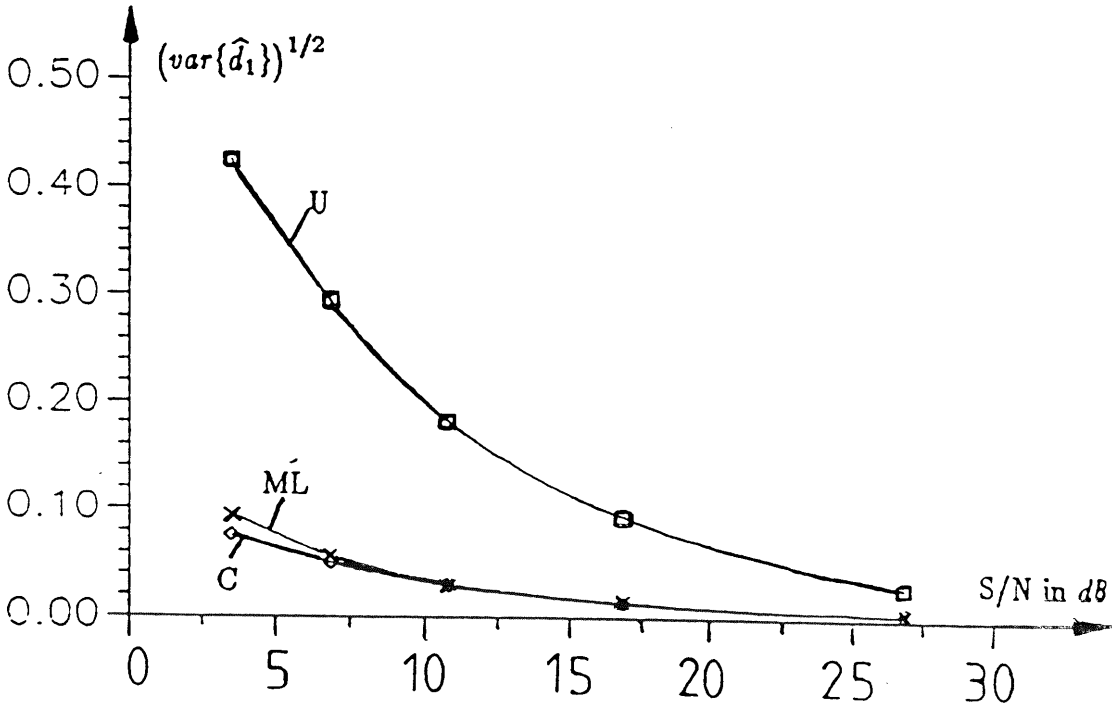

Abb. 3.3: Standard deviation of $\hat{d}_1$ using stronly correlated noise

significant reduction of the variance can be reached. For white noise and high signal-to-noise ratios there is according to eq. 3.5 no difference between the variance using the original or prefiltered images.

Acknowledgements

This work was supportet by the Deutsche Forschungsgemeinschaft (DFG).

References

Adiv, G. (1985):
Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects. IEEE Trans. Pattern Analysis and Machine Intelligence, PAMI-7, 4, pp. 384-401, 1985.

Bähr, H.-P. (1985):
Digitale Bildverarbeitung: Anwendung in Photogrammetrie und Fernerkundung. Wichmann Verlag, 1985.

Beyer, S. (1985):
Displacement-Schätzverfahren für Fernsehbildsignale mit minimaler Schätzfehlervarianz. Diss. Univ. Hannover, VDI-Fortschrittbericht: Reihe 10; Nr. 51, 1985.

Burkhardt, H.; Diehl, N. (1986):
Simultaneous Estimation of Rotation and Translation in Image Sequences. Proc. of the European Signal Processing Conference, EUSIPCO-86, Den Haag, pp. 821-824, 1986.

Burkhardt, H; Moll, H. (1978):
Ein modifiziertes Newton-Raphson-Verfahren zur modelladaptiven Identifikation von Laufzeiten. Mitteilungen aus dem Institut für Meß- und Regelungstechnik der Univ. Karlsruhe, 1978.

Burkhardt, H; Moll, H. (1979):
A modified Newton-Raphson-Search for the Model-Adaptive Identifications of Delays. In: Isermann, R. (ed.): Identification and System Parameter Estimation. Pergamon, pp. 1279-1286, 1979.

Diehl, N.; Burkhardt, H. (1986):
Planar Motion Estimation with a Fast Converging Algorithm. Proc. of the 8th Intern. Conf. on Pattern Recognition, 8-ICPR, Paris, pp. 1099-1102, 1986.

Diehl, N. (1987):
Schätzung dreidimensionaler Bewegungsparameter aus Bildfolgen. 9. DAGM-Symposium Mustererkennung, Braunschweig, Informatik Fachberichte, Bd. 149, pp. 272-276, 1987.

Diehl, N. (1988):
Methoden zur allgemeinen Bewegungsschätzung in Bildfolgen. Dissertation TU Hamburg-Harburg, Fortschritt-Bericht VDI Reihe, 10 Nr. 92, 1988.

Fang, J.-R.; Huang, T.S. (1984):
Solving Three-Dimensional Motion Equations: Uniqueness, Algorithms, and Numerical Results. Computer Vision, Graphics, Image Processing, 26, pp. 183-206, 1984.

Gill, P.E.; Murray, W.; Wright, M.H. (1981):
Practical Optimization. Academic Press, 1981.

Huang, T.S. (ed.) (1981):
Image Sequence Analysis. Springer, 1981

Huang, T.S. (ed.) (1983):
Image Sequence Processing and Dynamic Scene Analysis. Springer, 1983.

Kanatani, K.-I. (1986):
Structure and Motion from Optical Flow under Orthographic Projection. Computer Vision, Graphics, and Image Processing, 35, pp. 181-199, 1986.

Knapp, C.H.; Carter, G.C. (1976):
The Generalized Correlation Method for Estimation of Time Delay. IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-24, 4, pp. 320-327, 1976.

Lenz, R. (1986):
Ein Verfahren zur Schätzung der Parameter geometrischer Bildtransformationen. Diss. Techn. Univ. München, 1986.

Longuet-Higgins, H.C. (1981):
A Computer Algorithm for Reconstructing a Scene from Projections. Nature, 293, pp. 133-135, 1981.

Mesch, F. (1982):
Geschwindigkeits- und Durchflußmessung mit Korrelationsverfahren. Regelungstechnische Praxis 24, Heft 3, pp. 73-82, 1982.

Musmann, H.G.; Pirsch, P.; Grallert, H.J. (1985):
Advances in Picture Coding. Proceedings of the IEEE, 73, 4, pp. 523-548, 1985.

Nagel, H.H. (1981):
Image Sequence Analysis: What can we learn from Applications? In: Huang, T.S. (ed.): Image Sequence Analysis. Springer, pp. 19-228, 1981.

Nagel, H.H. (1983):
Overview on Image Sequence Analysis. In: Huang, T.S. (ed.): Image Sequence Processing and Dynamic Scene Analysis. Springer, pp. 2-39, 1983.

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

Nagel, H.H. (1986):
Image Sequences - Ten (octal) Years - From Phenomenology towards a Theoretical Foundation. Proc. of the 8th Intern. Conf. on Pattern Recognition, pp. 1099-1102, 1986.

Netravali, A.N.; Salz, J. (1985):
Algorithms for Estimation of Three-Dimensional Motion. AT & T Technical Journal, 64, 2, pp. 335-346, 1985.

Prazdny, K. (1980):
Egomotion and Relative Depth Map from Optical Flow. Biological Cybernetics, pp. 87-102, 1980.

Roach, J.W.; Aggarwal, J.K. (1980):
Determining the Movements of Objects from a Sequence of Images. IEEE. Trans. Pattern Analysis and Machine Intelligence, PAMI-2, 6, pp. 554-562, 1980.

Spoer, P. (1987):
Schätzung der 3-dimensionalen Bewegungsvorgänge starrer, ebener Objekte in digitalen Fernsehbildfolgen mit Hilfe von Bewegungsparametern. Diss. Univ. Hannover, 1987.

Tsai, R.Y.; Huang, T.S. (1981):
Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch. IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-29, 6, pp. 1147-1152, 1981.

Tsai, R.Y.; Huang, T.S. (1982):
Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch, II: Singular Value Decomposition. IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-30, 4, pp. 525-534, 1982.

Tsai, R.Y.; Huang, T.S. (1984a):
Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces. IEEE Trans. Pattern Analysis and Machine Intelligence, PAMI-6, 1, pp. 13-27, 1984.

Tsai, R.Y.; Huang, T.S. (1984b):
Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch, III: Finite Point Correspondence and the Three-View-Problem. IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-32, 2, pp. 213-220, 1984.

van Trees, H.L. (1968):
Detection, Estimation, and Modulation Theory - Part I: Detection, Estimation and Linear Modulation Theory. Wiley, 1968.

Yen, B.L.; Huang, T.S. (1983):
Determining 3-D Motion and Structure of a Rigid Body Using Straight Line Correspondence. In: Huang, T.S. (ed.): Image Sequence Processing and Dynamic Scene Analysis. Springer, pp. 365-394, 1983.