PREMIER COLLOQUE IMAGE

Traitement, Synthèse, Technologie et Applications

BIARRITZ – Mai 1984 –



# Recent Advances in Image Sequence Analysis

# Progres recent en analyse de sequences d'images

Hans-Hellmut Nagel

Fraunhofer-Institut für Informations- und Datenverarbeitung, Sebastian-Kneipp-Str. 12-14, D - 7500 Karlsruhe 1 and Fakultät für Informatik der Universität Karlsruhe

## RESUME

Cette publication a pour but de montrer que des methodes de minimisation présentées récemment sont un outil mathematique pour la recherche systematique de solutions aux problèmes posés par l'estimation de champs de vecteurs de déplacement. Nous présentons comme perspective d'avenir l'étude de problèmes tels que le choix d'une échelle appropriée pour la description de structures de niveaux gris et un meilleur traitement des changements dans le temps. Nous soulignons les possibilites de faire une symbiose entre l'approche basée sur l'extraction de caracteristiques et celle basée sur l'utilisation d'une analyse des gradients de niveau gris en vue de l'estimation de vecteurs de deplacement.

## SUMMARY

It is argued that recently proposed minimization approaches offer mathematical tools for systematic investigations of problems connected with the estimation of displacement vector fields. Open problems such as the choice of a proper scale factor for the description of gray value structures and a better treatment of the time dependency are put into perspective. Possibilities to bridge the gap between feature-based and gradient-based approaches towards the estimation of displacement vector fields are outlined.

# Recent Advances in Image Sequences Analysis

# Progres recent en analyse de sequences d'images

H. NAGEL

## 1. Introduction

During the past few years, the analysis of image sequences has established itself firmly as a new area of scientific research. As Nagel 81 has shown, it is rooted in numerous application areas such as image coding, cloud motion analysis as one possibility to estimate wind velocity distributions, surveillance of outdoor as well as indoor scenes, "visual" feedback in robotics, target tracking -just to mention a few examples. The book by Ullman 79 represents an early link between this emerging technical discipline and fundamental research about how a biological system may analyze the changes in its visually perceptible environment.

Such a broad scope for image sequence analysis cannot be fully covered in a contribution of restricted size. The subsequent discussion will therefore be limited by the following - admittedly arbitrary - interpretation of 'recent advances'. Publications which have appeared prior to fall 1982 will be mentioned only to illustrate a particular point. The surveys of Aggarwal and Martin 83 as well as of Nagel 83b may provide access to the earlier literature about image sequence analysis by technical systems. Biological aspects are covered by Ullman 81 as well as Ullman and Hildreth 83. Numerous, more specialized contributions about these topics can be found in the recent books edited by Braddick and Sleigh 83 and by Huang 83. The word 'advances' will be restricted to refer to approaches for the frame-to-frame correspondence based on a mathematical formulation which facilitates analysis and comparisons. The concentration on what may be considered to represent a lopsided selection hopefully contributes to a discussion about the emergence of a theory from which even investigations towards more qualitative rather than quantitative analysis shall benefit in the long run.

The conceptual background for the subsequent discussion will be outlined in the next section. The notion of 'displacement vector field' will be introduced as a tool to make important information explicit that is captured implicitly by the raw data of a digitized image sequence. A minimization problem will be formulated in section three in order to estimate displacement vector fields from image sequences. This formulation will serve as a starting point to discuss various alternative approaches. One specific approximative solution to this estimation problem will be used to illustrate a relation between so-called token- or feature-matching approaches on one hand and gradient-based or differential approaches on the other hand in section four.

## 2. Conceptual background

A digital image sequence can represent spatio-temporal samples of the radiant energy flux - for example in the visible range of the electromagnetic spectrum - impinging on the surface of an image sensor.

This flux originates either at light sources or - usually - at surfaces which reflect light. The radiant flux density at the sensor surface carries information about the illumination, the material and the orientation of the reflecting surface. The digitized gray value sample recorded at a specified frame-time at a certain raster location on the sensor surface represents this flux density or irradiance -corrupted, however, by sensor noise. The variation of gray value samples from an image sequence as a function of the position vector $\vec{x}$ on the image sensor surface and of the frame time t thus carries information about the spatial arrangement of surface material, its illumination and the temporal variation of these entities.

In addition to the problems of static image analysis, the analysis of image sequences has

- to detect changes between image frames of a sequence

- and to interpret detected changes.
In order to reliably detect changes in the depicted environment, one has to separate the structural gray value variations related to the light reflecting surfaces from the stochastic variations associated with sensor noise. This requires adequate means to describe the structural gray value variation. Modeling the image as a mosaic of regions with constant gray values separated by step edge transitions is generally inadequate.

The more important hypotheses for the explanation of changes detected between image frames are the following:
- relative motion between the image sensor and (components of) the depicted scene;
- changes of light sources regarding their position, strength, number, spectral and beam formation characteristics;
- changes in shape and material properties of the scene or its components.

Among these, relative motion between sensor and - at least some - surfaces in the scene is the most frequently encountered hypothesis. Provided the images are sampled fast enough, it appears to be an acceptable approximation to assume that the radiant flux recorded from a specific surface element does not change significantly between two consecutive frames. Under this assumption, most changes can be described by a space variant mapping $\vec{u}(\vec{x})$ of image position $\vec{x}$ at frame time $t_2$ to image position $\vec{x} - \vec{u}(\vec{x})$ at frame time $t_1$:

$$g(\vec{x},t_2) = g2(\vec{x}) \simeq g1(\vec{x}-\vec{u}(\vec{x})) \simeq g(\vec{x}-\vec{u}(\vec{x}),t_1) \qquad (1)$$

This does not apply to image areas not visible at frame time $t_1$ - for example image areas corresponding to background uncovered or objects entering the field of view between frame times $t_1$ and $t_2$. If such image areas are small, $\vec{u}(\vec{x})$ is practically defined within the entire image frame.
Based on all these simplifying assumptions, equation (1) implies the so-called correspondence problem to estimate the displacement vector field $\vec{u}(\vec{x})$. It should be noted that we do not distinguish at this time between various possibilities:

- The sensor is stationary (with respect to a major part of the scene) with only one moving object;
- stationary sensor with several moving objects;
- the sensor moves relative to a stationary scene;
- the sensor moves relative to a scene containing several moving objects.

Provided we can estimate $\vec{u}(\vec{x})$, all these situations should be distinguished by a segmentation of the displacement vector field into areas each of which can be described by its specific set of parameters that characterize the relative motion between the sensor and the depicted surfaces.

## 3. Estimation of displacement vector fields

In general, equation (1) does not provide enough information by itself for a complete specification of $\vec{u}(\vec{x})$. Since both g1($\vec{x}$) and g2($\vec{x}$) are assumed to be corrupted by noise, the approximate equality of equation (1) is substituted by a minimization requirement: choose $\vec{u}(\vec{x})$ in such a manner that

$$\iint d\vec{x} \left\{ g2(\vec{x}) - g1(\vec{x}-\vec{u}(\vec{x})) \right\}^2 \Rightarrow min \qquad (2)$$

The expression within braces is well-known in the image coding literature as 'displaced frame difference'.

Even equation (2) is still insufficient to define $\vec{u}(\vec{x})$ in general because both images may contain curves or areas with practically constant gray values. A human observer will intuitively complement the information obtainable from the local gray value structure according to equation (2) by a smoothness postulate.

Generalizing an idea by Horn and Schunck 81, Nagel 83c postulated an 'oriented smoothness requirement': the displacement vector field $\vec{u}(\vec{x})$ should vary smoothly in the direction of small or no gray value variation. In a sense, the 'oriented smoothness' requirement extends the "no news is good news" approach of Grimson 83 to image sequences: if the gray value distribution does not exhibit any sizable discontinuity, there is no reason to assume that the displacement vector field should have a discontinuity at such a location. Following the arguments presented in Nagel 83c, equation (2) is replaced by:

$$\iint d\vec{x} \left\{ \left( g2(\vec{x}) - g1(\vec{x}-\vec{u}) \right)^2 + \alpha^2 \, trace\left( (\nabla\vec{u})^T W (\nabla\vec{u}) \right) \right\} \Rightarrow min \quad (3)$$

where the weight matrix W represents the influence of the local gray value structure on the variation of the displacement vector field, and the superscript T indicates transposition.

$$W = F / det\, F \quad (4a)$$

with

$$F = \begin{pmatrix} g_y^2 & -g_x g_y \\ -g_x g_y & g_x^2 \end{pmatrix} + a^2 \begin{pmatrix} g_{xy}^2 + g_{yy}^2 & -g_{xy}(g_{xx}+g_{yy}) \\ -g_{xy}(g_{xx}+g_{yy}) & g_{xx}^2 + g_{xy}^2 \end{pmatrix} \quad (4b)$$

and

$$\vec{u} = \begin{pmatrix} u \\ v \end{pmatrix} \qquad \nabla\vec{u} = \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} \quad (4c)$$

Subscripts are used to denote the partial derivative, i.e. $g_x = \partial g/\partial x$. Recent investigations (Nagel and Enkelmann 84a+b) indicate that it may be advantageous to drop the normalization factor det F in equation (4a) and to use the matrix F of equation (4b) directly in the place of W in equation (3).

<u>3.1 The approach of Horn and Schunck</u>

The following steps allow to transform equation (3) into the equation suggested by Horn and Schunck 81:

(i) The weight matrix W is replaced by the unit matrix I:

$$trace\left( (\nabla\vec{u})^T I (\nabla\vec{u}) \right) = trace\left( \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} \right) \quad (5a)$$

$$= u_x^2 + u_y^2 + v_x^2 + v_y^2 \quad (5b)$$

(ii) With

$$g1(\vec{x}-\vec{u}) \simeq g1(\vec{x}) - (\nabla g1)^T \vec{u} \quad (6a)$$

and

$$g2(\vec{x}) - g1(\vec{x}) \simeq \frac{\partial g1}{\partial t} = g1_t \quad (6b)$$

one obtains instead of equation (3)

$$\iint d\vec{x} \left\{ \left( (\nabla g1)^T \vec{u} + g1_t \right)^2 + \alpha^2 \left( u_x^2 + u_y^2 + v_x^2 + v_y^2 \right) \right\} \Rightarrow min \quad (7)$$

There are two problems with this formulation. Near prominent gray value transitions, the first order Taylor expansion of equation (6a) is insufficient. Schunck and Horn 81 derive the relation expressed by the first term in equation (7) in a manner which extends its validity beyond that expected according to the derivation given in equation (6).

A more important argument against the use of equation (7) is based on the undiscriminating smoothness requirement even across gray value transitions which could be images of occluding boundaries and thus curves of marked discontinuities of the displacement vector field.

In order to avoid that displacement estimates spill across potential discontinuities, Cornelius and Kanade 83 deactivated the smoothness requirement in the neighborhood of zero-crossing contours.

Yachida 83 took displacement vectors estimated at gray value corners and propagated them into neighboring areas with large gray value gradients, based on the method of Horn and Schunck 81. In order to suppress the propagation of displacement estimates with large local variations, his iterative improvement scheme used the inverse variance of displacement estimates as a weight.

Wu et al. 82, Davis et al. 83 propagated a displacement estimate only along a contour line between corner points. At each new contour point, they combined the estimated displacement vector from a previous contour point with new estimates of the contour direction and of the displacement component perpendicular to the contour in order to update the tangential component of the displacement vector.

Compared with all these approaches, the one expressed by equation (3) has the advantage that it does not require the explicit determination of gray value transition fronts such as edge lines or zero-crossing contours. Moreover, the local gray value structure has influence on the displacement estimate directly rather than indirectly through a weight factor given by the inverse variance of the displacement estimate as in the case of Yachida 83.

### 3.2 The approach of Hildreth

Hildreth 83a+b minimized the sum of two terms integrated along a zero-crossing contour. The first term is the squared difference between the estimated and the 'measured' displacement component perpendicular to the contour. The latter one is determined by the following expression (Hildreth 83b, p. 62):

$$u^{\perp} = -\left(g2' - g1'\right) / |\nabla g'| \qquad (8)$$

where g' represents the convolution of the gray value $g(\vec{x})$ with the Laplacian of a Gaussian

$$g' = g * \nabla^2 G = g * \left(G_{xx} + G_{yy}\right) \qquad (9a)$$

and

$$G(\vec{x}) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (9b)$$

The second term - expressing the smoothness requirement - is given by the squared derivative of the displacement vector field with respect to the arclength along the zero-crossing contour.

If $\vec{n}(s)$ represents the normal to the zero-crossing contour for the arclength s, the integral to be minimized according to Hildreth 83b - equ. (25) on page 46 - can be written

$$\oint_{\substack{\text{zero-crossing}\\ \text{contour}}} ds \left\{ \beta\left(\vec{u}^T\vec{n} - u^{\perp}\right)^2 + \left(\frac{\partial u}{\partial s}\right)^2 + \left(\frac{\partial v}{\partial s}\right)^2 \right\} \Rightarrow min \qquad (10)$$

Equation (3) with W given directly by the matrix F - see equ. (4b) - can be related to equation (10) by the following steps:

(i) The convolution of $g(\vec{x},t)$ with the Gaussian G is replaced by $g(\vec{x},t)$ in equation (8) and (10).

(ii) Since we now integrate along the zero-crossing contour of $\nabla^2 g$ instead of $\nabla^2 g'$, we have

$$\nabla^2 g = g_{xx} + g_{yy} = 0 \quad \text{or} \quad g_{xx} = -g_{yy} \qquad (11)$$

This simplifies the matrix F to

$$F_{\substack{\text{zero-}\\ \text{crossing}}} = \begin{pmatrix} g_y^2 & -g_x g_y \\ -g_x g_y & g_x^2 \end{pmatrix} + a^2 \begin{pmatrix} g_{xy}^2 + g_{yy}^2 & 0 \\ 0 & g_{xx}^2 + g_{xy}^2 \end{pmatrix} \qquad (12)$$

Since $g_{xx}{}^2 = g_{yy}{}^2$ at a zero-crossing contour of $\nabla^2 g$, it can be shown by a short calculation that

$$g_{xx}^2 + g_{xy}^2 = g_{xy}^2 + g_{yy}^2 = \varkappa^2 \qquad (13)$$

where $\varkappa$ represents the eigenvalue of the diagonalized Hessian at a zero-crossing contour

$$\begin{pmatrix} g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{pmatrix} \Rightarrow \begin{pmatrix} \varkappa & 0 \\ 0 & -\varkappa \end{pmatrix} \qquad (14)$$

Using equation (12) and (13), one obtains for the oriented smoothness term in equation (3):

$$\mathrm{trace}\left(\left(\nabla\vec{u}\right)^T F\left(\nabla\vec{u}\right)\right) = \left(g_y u_x - g_x u_y\right)^2 + \left(g_y v_x - g_x v_y\right)^2 + a^2 \varkappa^2\left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right) \quad (15)$$

If the zero-crossing contour is straight, then $\varkappa^2 = 0$ and the gradient is perpendicular to the zero-crossing contour (see Nagel 83a). In this case, the vector $(g_y, -g_x)^T$ is tangential to the zero-crossing contour. If the zero-crossing contour is given by $\vec{x}(s)$, the tangent to the zero-crossing contour is $(dx/ds, dy/ds)^T$. We thus may write along a straight-line zero-crossing contour

$$\left(g_y u_x - g_x u_y\right) = |\nabla g| \left(u_x \frac{dx}{ds} + u_y \frac{dy}{ds}\right) = |\nabla g| \frac{\partial u}{\partial s} \quad (16)$$

If we now assume that, even for curved parts of the zero-crossing contour, the gradient $\nabla$g will be virtually perpendicular to the zero-crossing contour, we may write instead of equation (15)

$$\mathrm{trace}\left(\left(\nabla\vec{u}\right)^T F\left(\nabla\vec{u}\right)\right) = |\nabla g|^2\left(\left(\frac{\partial u}{\partial s}\right)^2 + \left(\frac{\partial v}{\partial s}\right)^2\right) + a^2\varkappa^2\left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right) \quad (17)$$

(iii) Significant contributions to the area integral of equation (3) with F instead of W will only occur when the first and second derivatives of $g(\vec{x})$ are large, i.e. in the vicinity of zero-crossing contours. Neglecting the contributions to the integral (3) in areas with small gradients will essentially reduce the area integral to a curve integral along a zero-crossing contour - implying, of course, a 'reasonable' gray value structure.

(iv) Using a substitution analogous to equation (6), we thus obtain

$$\iint d\vec{x} \left\{ \left(g2(\vec{x}) - g1(\vec{x}-\vec{u})\right)^2 + \alpha^2\, \mathrm{trace}\left(\left(\nabla\vec{u}\right)^T F\left(\nabla\vec{u}\right)\right) \right\}$$

$$\Rightarrow \oint_{\substack{\text{zero-crossing}\\ \text{contour}}} ds \left\{ \left(g2(\vec{x}) - g1(\vec{x}-\vec{u})\right)^2 + \alpha^2 |\nabla g1|^2 \left(\left(\frac{\partial u}{\partial s}\right)^2 + \left(\frac{\partial v}{\partial s}\right)^2\right) + \alpha^2 a^2 \varkappa^2 \left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right) \right\} \quad (18a)$$

$$\Rightarrow \oint_{\substack{\text{zero-crossing}\\ \text{contour}}} ds\, |\nabla g1|^2 \left\{ \left(\vec{u}^T\vec{n} - \frac{g2-g1}{|\nabla g1|}\right)^2 + \alpha^2\left(\left(\frac{\partial u}{\partial s}\right)^2 + \left(\frac{\partial v}{\partial s}\right)^2\right) + \alpha^2 \frac{a^2\varkappa^2}{|\nabla g1|^2}\left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right)\right\} \Rightarrow \min \quad (18b)$$

A comparison between equations (10) and (18b) now allows to illustrate the difference between the approaches of Hildreth 83 and Nagel 83c.

If we replace in equation (8) the convolution of the gray value g with the Laplacian of a Gaussian by g itself, we obtain

$$u^{\perp} = -\left(g2 - g1\right) / |\nabla g1| \quad (8')$$

The difference between equations (10) and (18b) thus reduces to two aspects:

(a) Equation (18b) derived from equation (3) weighs the integrand by the square of the gray value gradient: the larger the gray value gradient across the zero-crossing contour, the more stringent is the local smoothness requirement. The approach of Hildreth according to equation (10) does not know such an arclength-depending weight for the integrand.

(b) At sharp corners of the zero-crossing contour - i.e. at locations with large values of $\varkappa^2$ - the integrand of equation (18b) requires that the derivatives of both components of the displacement vector remain small in order to minimize the integral. In other words, the approach of Nagel 83c implies that the displacement vector field is locally (almost) constant at gray value corners. This is gratifying because it justifies a posteriori an approach to estimate the displacement vector for gray value corners based on the assumption that the displacement vector is constant in a small environment around a gray value corner - see Nagel 83a.

### 3.3 The approach of Prager and Arbib

Prager and Arbib 83 have developed a heuristic approach for the estimation of a displacement vector field which is based on feature-matching but incorporates a smoothness requirement. According to their notation, an assertion P for having found a feature at image position PX with attributes PT is denoted by the tupel (PX, PT). An analogous assertion for the next frame is denoted by Q = (QX, QT). These authors introduce a heuristic distance function to grade the match between tupel Q(i) from one frame and tupel P(j) from the preceding frame, with a displacement

$$DX(i) = QX(i) - PX(j) \qquad (19)$$

Based on these concepts, they directly introduce a heuristic iteration formula for the estimation of DX(i):

$$DX(i)_{k+1} = DX(i)_k + \lambda B_1(i) + \mu B_2(i) \qquad (20)$$

with

$$B_1(i) = \frac{1}{\sum_{j \in N} w_{ij}} \sum_{j \in N} w_{ij} DX(j) - DX(i) \qquad (21a)$$

where the sums extend over all tupels j in a neighborhood around tupel i, and $w_{ij}$ denote heuristic weight factors. This term represents the smoothness requirement: it will contribute to a correction for $DX(i)_{k+1}$ in the $(k+1)^{th}$ iteration unless $DX(i)_k$ is given by a weighted mean of surrounding displacement estimates.

The remaining term

$$B_2(i) = QX(i) - A[E(i); P] - DX(i) \qquad (21b)$$

contains a function A which describes the position of a feature assertion P from the preceding frame most compatible with a constructed assertion E(i) = (QX(i)-DX(i),T) characterizing an expected match for Q(i). Whenever the current displacement $DX_k(i)$ deviates from the vector connecting the position of Q(i) with its current best match P according to A[E(i); P], a correction $B_2(i)$ will modify $DX_k(i)$ in order to obtain $DX_{k+1}(i)$. Due to the numerous parameters in this entire approach, it is difficult to evaluate it for a comparison with the minimization approach given by equation (3).

The features are corners, edge elements, and points of high contrast. These features are characterized by heuristically introduced masks which are convolved with each image frame to obtain assertions P.

### 3.4 An example from image coding

Movement compensation for image coding requires the estimation of displacement vector fields. Basically, the square of the displaced frame difference given by the integrand of equation (2) has to be minimized for this purpose. An approach based on the Newton-Raphson method will provide faster convergence than a gradient descent approach - see, for example, Bergmann 83. It requires, however, a starting estimate which has to be closer to the solution than in the case of a gradient descent approach.

Cafforio and Rocca 83 illustrated the theoretical background for such approaches and studied a smoothness requirement in this context. They assumed that a displacement vector $\hat{\vec{u}}(i)$ has to be estimated for the i-th block of pixels. Let $\hat{\vec{u}}(i-1)$ denote the estimate derived for the preceding block. $\hat{\vec{v}}(i)$ should represent the estimate derived from the data of the current block only. If the block size is small compared to the image regions expected to exhibit essentially the same displacement, it is reasonable to combine the preceding estimate $\hat{\vec{u}}(i-1)$ with $\hat{\vec{v}}(i)$ for the current block in order to obtain a more reliable estimate $\hat{\vec{u}}(i)$ for the current block:

$$\hat{\vec{u}}(i) = A \cdot \hat{\vec{u}}(i-1) + B\,\hat{\vec{v}}(i) \qquad (22)$$

The weight matrices A and B in equation (22) have been derived by Cafforio and Rocca 83. They depend on the gray value distribution within the two blocks and on the autocorrelation function of the true displacement vector field. The authors develop a stochastic model describing the displacement vector field as a mosaic of areas each with an independent constant randomly oriented displacement vector. Its magnitude is considered to be another random variable with zero mean and a probability density function uniformly decreasing for large arguments.

Although such an approach yields quite acceptable results for image coding applications, it does not appear to be appropriate if the displacement vector field should be evaluated, for example, in order to obtain a three-dimensional description of moving objects and their trajectories.

### 3.5 The scale problem

In order to evaluate an equation which contains derivatives of the gray value function $g(\vec{x},t)$ with respect to x, y, or t, one has to specify an environment for the computation of these derivatives from the digitized samples. If the mask size is too small, the estimate for the derivative will be influenced by noise. Too large a mask size will not allow to estimate the local gray value structure. Most mask sizes so far have been chosen ad hoc.

A more systematic approach convolves the image with a Gaussian - see equation (9b). A spatial derivative of the convolved image can be computed by convolving the image with the corresponding spatial derivative of the Gaussian. A choice for the standard deviation $\sigma$ of the Gaussian guides the choice of the mask size for the derivative operator: it should be large enough so that the functional value for the derivative of the Gaussian can be safely neglected outside this mask.

Since the extent of characteristic gray value structures may vary within an image, the exclusive use of a single value for the scale parameter $\sigma$ will be insufficient in general. One solution to this problem could be seen in the systematic analysis of an image with a set of different scale parameters - in analogy to approaches studied for the estimation of stereo disparities (see, e.g., Marr 82). Results obtained for different scale parameters could be combined in a hierarchical manner. A strictly top-down approach from large scale parameters to smaller ones may lead to difficulties because displacement estimates obtained at a larger scale may give inappropriate start values at smaller scales, for example around discontinuities of the displacement vector field. A multi-grid approach whereby intermediate estimates obtained at several scales are communicated in an iterative manner up as well as down along a scale hierarchy appears to be the method of choice. It is well known that multigrid methods offer a significant computational advantage for the numerical solution of partial differential equations. Terzopoulos 83 provides examples for the reconstruction of smooth surfaces from sample points.

Using a small set of different values for the scale parameter $\sigma$ may appear to be mainly a computational technique employed in order to obtain faster convergence. There is, however, another aspect which appears worth studying it in more detail. Witkin 83 began to study how the convolution of a function with a Gaussian behaved as a function of the scale parameter $\sigma$. This may offer a method for the systematic investigation of gray value structures in images. The implication for the estimation of displacement vector fields have to be evaluated.

The idea to emphasize spatial gray value variations at a certain scale by convolving an image with a bivariate Gaussian has been extended by Buxton and Buxton 83 to the temporal axis:

$$g''(\vec{x},t) = \int d\tau \iint d\vec{\xi} \;\; g(\vec{\xi},\tau) \cdot c\left(\frac{\alpha}{\pi}\right)^{3/2} e^{-\alpha\left[(\vec{x}-\vec{\xi})^2 + c^2(t-\tau)^2\right]} \qquad (23)$$

The standard deviation for the spatial variables corresponds to

$$\sigma_{spatial} = \frac{1}{\sqrt{2\alpha}} \qquad (24a)$$

and for the temporal variable to

$$\sigma_{temporal} = \frac{1}{c\sqrt{2\alpha}} \qquad (24b)$$

It appears advantageous to interpret the constant c with the dimension of a velocity as the factor determining the relation between the scale factors for spatial and temporal coordinates. These authors analyzed the zero-crossings of $\nabla^2 g''$ and found that the location of zero-crossings of $\nabla^2 g''$ shifts with time in a manner that depends on the scale parameter $\alpha$. It thus is no longer self-evident how to combine zero-crossings of $\nabla^2 g''$ obtained for different scale parameters $\alpha$. Buxton and Buxton 83 solved this problem by substituting the d'Alambertian operator for the Laplacian. They investigated the zero-crossings of

$$-\Box^2 \left\{ G(\vec{x},t) * g(\vec{x},t) \right\} \qquad (25)$$

where $\Box^2$ is given by

$$\Box^2 = \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \qquad (26)$$

Apart from the important property that a gray value transition will have zero-crossings at the same position for different scale parameters, it turned out that another type of zero-crossing shows up. It is related to the relative motion between the sensor and the (surface, material, illumination) discontinuity in space which causes a gray value transition to appear in the image. In the periphery of the field of view, these two types of zero-crossings may interfere with each other. According to Buxton and Buxton 83, observation of such an interference or 'cross-over' effect would offer an alternative way to infer the distance of the spatial discontinuity from a moving sensor. It still has to be investigated to what extent these ideas can be exploited for the analysis of image sequences. It clearly represents a more systematic approach towards the problem of relative scaling between the temporal and the spatial coordinates. A rather heuristic approach to temporal smoothing for the benefit of more reliable displacement estimates has already been described by Yachida 83 who averaged displacement estimates iteratively across three consecutive image frames.

## 4. The 'feature' problem

Many approaches towards the frame-to-frame correspondence problem first attempt to extract features from each image frame. In a subsequent step, matching of such features is attempted. The work of Prager and Arbib 83 discussed in section 3.3 provides just one of numerous examples for such an approach. Rather than portraying a strict dichotomy between feature-based and gradient-based correspondence methods, this section will attempt to show a common mathematical basis for the most important features: corners, edges, and isolated gray value extrema.

Our discussion starts from equation (2) - i.e. we neglect the smoothness issue for the moment. Let us assume that an initial estimate $\vec{u}_o$ is available which should be refined by the determination of a small correction rector $D\vec{u} = (Du, Dv)^T$:

$$\vec{u}(\vec{x}) = \vec{u}_0(\vec{x}) + D\vec{u}(\vec{x}) \qquad (27)$$

As shown by Nagel 83a+c, one can derive the following equation for $D\vec{u}$:

$$C\,D\vec{u} = -\overline{\left(g2(\vec{x}) - g1(\vec{x}-\vec{u}_0)\right)\nabla g1(\vec{x}-\vec{u}_0)} \qquad (28)$$

The averaging process denoted by the overbar on the right hand side of equation (28) is extended over an environment around the position $\vec{x}_0$ which is required to estimate the spatial derivatives of $g1(\vec{x}_0)$ occurring in equation (28) - see section 3.5. The coefficient matrix C is given by (writing g instead of g1):

$$C = \begin{pmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{pmatrix} + a^2 \begin{pmatrix} g_{xx}^2 + g_{xy}^2 & g_{xy}(g_{xx}+g_{yy}) \\ g_{xy}(g_{xx}+g_{yy}) & g_{xy}^2 + g_{yy}^2 \end{pmatrix} \qquad (29)$$

It should be noted that $C^{-1} = W = F/\det F$ as given by equations (4a + b). This result has been derived by approximating the gray value distribution of $g1(\vec{x})$ around $\vec{x}_o$ by a Taylor series expansion retaining terms up to the second order in $(\vec{x}-\vec{x}_o)$. It is illustrating to see that this coefficient matrix captures essential information about the gray value structure around $\vec{x}_o$.

4.1 Gray value corners

Nagel 83a - see also Kitchen and Rosenfeld 80 + 82 - has shown that a gray value corner can be characterized as the location of maximum planar curvature in the locus curve of steepest gray value slope. If the local coordinate system is aligned with the principal curvature directions at $g(\vec{x}_o)$ -so that $g_{xy} = 0$ -, a gray value corner may be defined by the following requirements:

$$g_x = max \qquad g_y = 0 \qquad (30a)$$

$$g_{xx} = 0 \qquad g_{yy} = max \qquad (30b)$$

Equations (30a) express the fact that the maximum gradient is oriented in one principal curvature direction - here taken to be x. The zero-crossing of $g_{xx}$ implies that $g_x$ is indeed a maximum. The non-zero value of $g_{yy}$ implies that the locus line of zero-crossings has maximum curvature at $\vec{x}_o$. One thus obtains the following coefficient matrix at a gray value corner:

$$C_{Corner} = \begin{pmatrix} g_x^2 & 0 \\ 0 & a^2 g_{yy}^2 \end{pmatrix} \qquad (31)$$

Since according to equations (30) the coefficient matrix is non-singular at $\vec{x}_o$, it can be inverted to obtain a closed form solution for the correction vector $D\vec{u}$ at such a location. An experimental evaluation of this approach by Nagel and Enkelmann 82 has yielded satisfactory results - see also Dreschler-Fischer et al. 83.

The position of maximum planar curvature in contour lines has been selected as a feature point by Yam and Davis 81. More recently, Lawton 83 employed corners in zero-crossing contours as feature points for a heuristic interframe match. As has been pointed out by Nagel 83a and Dreschler and Nagel 82, the point of maximum curvature in a zero-crossing contour of $\nabla^2 g = g_{xx} + g_{yy}$ does not coincide with the location of a gray value corner defined by equations (30). It appears, however, that both approaches pick predominantly the same gray value structure selected intuitively by a human observer - provided, of course, that the derivatives are determined at the same scale. Recently, Zuniga and Haralick 83 reported some variations on the gray value corner concept described by Kitchen and Rosenfeld 82 and Nagel 83a and first used by Nagel 83c for closed-form frame-to-frame matching.

4.2 Isolated gray value extrema

Local maxima or minima of the gray value distribution offer an additional well defined feature to be matched from frame to frame. Local extrema can be characterized by vanishing gradient magnitude with nonzero values for the corresponding second derivatives. One thus obtains

$$C_{extremum} = a^2 \begin{pmatrix} g_{xx}^2 & 0 \\ 0 & g_{yy}^2 \end{pmatrix} \qquad (32)$$

Again as in the case of the gray value corner, the conditions of a gray value extremum ensure the nonsingularity of C so that it may be inverted to obtain a closed form solution for both components of the displacement estimate correction $D\vec{u}$. Prager and Arbib 83 employed isolated extrema as feature points for interframe matching.

4.3 Edge lines

In the case of a straight line gray value transition front, the second derivative with respect to y will vanish in an entire environment around the zero-crossing of $g_{xx}$ which characterizes the curve of steepest slope; i.e. maximum $\nabla g = (g_x, o)$. In this case, the matrix C degenerates to:

$$C_{straight\ line} = \begin{pmatrix} g_x^2 & 0 \\ 0 & 0 \end{pmatrix} \qquad (33a)$$

reflecting the well known fact that one cannot determine both components of $\vec{u}$ uniquely at a straight line gray value transition. This remains true even for the area around the zero-crossing of $g_{xx}$:

$$C_{\substack{straight\ line \\ neighborhood}} = \begin{pmatrix} g_x^2 + a^2 g_{xx}^2 & 0 \\ 0 & 0 \end{pmatrix} \qquad (33b)$$

There are numerous examples for interframe matching schemes based on line elements. It is interesting to see how a slight bending of the gray value transition front will result in an ill-conditioned version of C. Let us assume that $g_{yy}$ is small. In this case the gradient will still be essentially perpendicular to the zero-crossing contour of $g_{xx}$, i.e. $g_x$ will be large. As a consequence, one eigenvalue of C - say $\varkappa$ - is much larger than the second one, say $\lambda$. We thus have

$$\det C = \varkappa \cdot \lambda \qquad (34a)$$

$$\operatorname{trace} C = \varkappa + \lambda \qquad (34b)$$

$$\frac{\det C}{\left(\frac{1}{2} \operatorname{trace} C\right)^2} = \frac{\varkappa \cdot \lambda}{\left(\frac{\varkappa + \lambda}{2}\right)^2} \simeq 4 \frac{\lambda}{\varkappa} \quad \text{for } \lambda \ll \varkappa \qquad (35)$$

Since the left hand side of equation (35) represents a measure invariant with respect to rotations of the coordinate system, ill-conditioned matrices C - i.e. insufficiently curved gray value transition fronts - can be detected without diagonalizing C. This has been demonstrated by Nagel and Enkelmann 83.

5. Conclusion

The preceding sections attempted to support the following hypotheses:

(1) As long as a strictly two-dimensional approach to the estimation of a displacement vector field is considered, a minimization approach according to equation (3) appears to offer an acceptable solution. It

incorporates in a natural manner the match between significant gray value structures as well as a smoothness criterion for the displacement vector field. The latter one leaves enough flexibility to account for discontinuities at prominent gray value transition fronts.

(2) The approach formulated by equation (3) incorporates in a natural manner the influence of the most basic features found in many token-based or feature-based approaches to the corresponding problem, namely corners, isolated gray value extrema, and line elements. It thus offers a road to unify feature-based and gradient-based approaches.

Equation (3) does not contain any clue how to cope with the scale problem. It treats, moreover, sampling along the temporal axis as a basically discrete problem whereas sampling along the two spatial image axes does not show up explicitly in the equation (3). It is hoped, however, that the discussion of this minimization approach could put these two problems into the proper perspective and to indicate directions how they could be attacked.

The overriding hypothesis is that tools emerge which facilitate a systematic investigation into the problem of estimating displacement vector fields. Such more systematic studies may provide the groundwork to justify the design and development of special purpose computer structures which could facilitate the estimation of displacement vector fields from image sequences in real-time.

The contents of this contribution represents a deliberate gamble on the part of the author to omit many very interesting aspects from discussion in order not to overcrowd the picture by too many details. There may be readers who are going to miss, for example, references to attempts at three-dimensional interpretations of changes observable in image sequences. Recent surveys by Ullman 84 and Nagel 84 are offered as a way to access literature about this topic. It should be pointed out that eventually a strictly two-dimensional approach to the estimation of displacement vector fields is inappropriate. Somehow those problems neglected for the derivation of equation (3) have to be attacked, especially the change in radiant flux from a surface element which moves relative to light source and sensor, the uncovering of background, the appearance or disappearance of scene components along the boundary of the field of view. These topics are left to the future.

## References

Aggarwal and Martin 83
Dynamic Scene Analysis, J.K. Aggarwal and W.N. Martin, in T.S. Huang (ed.) Image Sequence Processing and Dynamic Scene Analysis, NATO ASI Series F2, pp. 40-73, Springer-Verlag Berlin/Heidelberg/New York 1983.

Bergmann 83
Analysis of Different Displacement Estimation Algorithms for Digital Television Signals, H.C. Bergmann, in T.S. Huang (ed.) Image Sequence Processing and Dynamic Scene Analysis, NATO ASI Series F2, pp. 215-234, Springer-Verlag Berlin/Heidelberg/New York 1983.

Braddick and Sleigh 83
Physical and Biological Processing of Images, O.J. Braddick and A.C. Sleigh (eds.), Springer-Verlag Berlin/Heidelberg/New York 1983.

Buxton and Buxton 83
Monocular Depth Perception from Optical Flow by Space Time Signal Processing, B.F. Buxton and H. Buxton, Proc. Royal Society London B218 (1983) 27-47.

Cafforio and Rocca 83
The Differential Method for Image Motion Estimation, C. Cafforio and F. Rocca, in T.S. Huang (ed.), Image Sequence Processing and Dynamic Scene Analysis, NATO ASI Series F2, pp. 104-124, Springer-Verlag Berlin/Heidelberg/New York 1983.

Cornelius and Kanade 83
Adapting Optical-Flow to Measure Object Motion in Reflectance and X-Ray Image Sequences, N. Cornelius and T. Kanade, ACM SIGGRAPH/SIGART Interdisciplinary Workshop on Motion: Representation and Perception, April 4-6, 1983, Toronto/Canada, pp. 50-58.

Davis et al. 83
Contour-Based Motion Estimation, L.S. Davis, Z. Wu, and H. Sun, Computer Vision, Graphics, and Image Processing 23 (1983) 313-326.

Dreschler and Nagel 82
On the Selection of Critical Points and Local Curvature Extrema of Region

Boundaries for Interframe Matching, L. Dreschler and H.-H. Nagel, ICPR-82, pp. 542-544.

Dreschler-Fischer et al. 83
Lernen durch Beobachtung von Szenen mit bewegten Objekten: Phasen einer Systementwicklung, L. Dreschler-Fischer, W. Enkelmann und H.-H. Nagel, Proc. 5. DAGM-Symposium Mustererkennung 1983, Karlsruhe, 11. - 13. Oktober 1983, VDE-Fachberichte 35, pp. 29-34, VDE-Verlag GmbH Berlin-Offenbach 1983.

Grimson 83
Surface Consistency Constraints in Vision, W.E.L. Grimson, Computer Vision, Graphics, and Image Processing 24 (1983) 28-51.

Hildreth 83a
Computing the Velocity Field Along Contours, E.C. Hildreth, ACM SIGGRAPH/SIGART Interdisciplinary Workshop on Motion: Representation and Perception, April 4-6, 1983, Toronto/Canada, pp. 26-32.

Hildreth 83b
The Measurement of Visual Motion, E.C. Hildreth, Ph.D. Thesis (August 1983), Dept. Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge/MA.

Horn and Schunck 81
Determining Optical Flow, B.K.P. Horn and B.G. Schunck, Artificial Intelligence 17 (1981) 185-203.

Huang 83
Image Sequence Processing and Dynamic Scene Analysis, T.S. Huang (ed.), NATO Advanced Study Institute Series F2, Braunlage/FR Germany, June 21-July 2, 1982, Springer-Verlag Berlin/Heidelberg/New York, 1983.

Kitchen and Rosenfeld 80
Gray-Level Corner Detection, L. Kitchen and A. Rosenfeld, TR-887 (April 1980), Computer Science Center, University of Maryland, College Park/MD.

Kitchen and Rosenfeld 82
Gray-Level Corner Detection, L. Kitchen and A. Rosenfeld, Pattern Recognition Letters 1 (1982) 95-102.

Lawton 83
Processing Translational Motion Sequences, D.T. Lawton, Computer Vision, Graphics, and Image Processing 22 (1983) 116-144.

Marr 82
Vision, D. Marr, W.H. Freeman and Company, San Francisco/CA 1982.

Nagel 81
Image Sequence Analysis: What Can We Learn from Applications?, H.-H. Nagel, in Image Sequence Analysis, pp. 19-228, T.S. Huang (ed.), Springer-Verlag Berlin/Heidelberg/New York, 1981.

Nagel 83a
Displacement Vectors Derived from Second-Order Intensity Variations in Image Sequences, H.-H. Nagel, Computer Vision, Graphics, and Image Processing 21 (1983) 85-117.

Nagel 83b
Overview on Image Sequence Analysis, H.-H. Nagel, Proc. NATO Advanced Study Institute on Image Sequence Processing and Dynamic Scene Analysis, Braunlage/FR Germany, June 21-July 2, 1982, T.S. Huang (ed.), NATO ASI Series F2, pp. 2-39, Springer-Verlag Berlin/Heidelberg/New York 1983.

Nagel 83c
Constraints for the Estimation of Displacement Vector Fields from Image Sequences, H.-H. Nagel, IJCAI-83, pp. 945-951.

Nagel 84
Spatio-Temporal Modeling Based on Image Sequences, H.-H. Nagel, Proc. International Symposium on Image Processing and its Applications, University of Tokyo, January 18-21, 1984 M. Onoe (ed.) (in press).

Nagel and Enkelmann 82
Investigation of Second Order Gray Value Variations to Estimate Corner Point Displacements, H.-H. Nagel and W. Enkelmann, ICPR-82, pp. 768-773.

Nagel and Enkelmann 83
Iterative Estimation of Displacement Vector Fields in TV-Frame Sequences, H.-H. Nagel and W. Enkelmann, Proc. European Signal Processing Conference, EUSIPSO-83, pp. 299-302.

Nagel and Enkelmann 84a
Towards the Estimation of Displacement Vector Fields by "Oriented Smoothness" Constraints, H.-H. Nagel and W. Enkelmann, ICPR-84, (in press).

Nagel and Enkelmann 84b
An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences, H.-H. Nagel and W. Enkelmann (in preparation).

Prager and Arbib 83
Computing the Optical Flow: The MATCH Algorithm and Prediction, J.M. Prager and M.A. Arbib, Computer Vision, Graphics, and Image Processing 24 (1983) 271-304.

Schunck and Horn 81
Constraints on Optical Flow Computation, B.G. Schunck and B.K.P. Horn, IEEE PRIP-81, pp. 205-210.

Terzopoulos 83
Multilevel Computational Processes for Visual Surface Reconstruction, D. Terzopoulos, Computer Vision, Graphics, and Image Processing 24 (1983) 52-96.

Ullman 79
The Interpretation of Visual Motion, S. Ullman, The MIT Press, Cambridge/Mass., 1979.

Ullman 81
Analysis of Visual Motion by Biological and Computer Systems, S. Ullman, IEEE Computer 14, No. 8 (August 1981) 57-69.

Ullman 84
Computational Studies in the Interpretation of Structure and Motion: Summary and Extension, S. Ullman, AI Memo 706 (March 1983), Artificial Intelligence Laboratory MIT, Cambridge/MA, to appear in Beck and Rosenfeld (eds.) Human and Machine Vision, Academic Press.

Ullman and Hildreth 83
The Measurement of Visual Motion, S. Ullman and E.C. Hildreth, in O.J. Braddick and A.C. Sleigh (eds.), Physical and Biological Processing of Images, Springer-Verlag

Berlin/Heidelberg/New York 1983, pp. 154-176.

Witkin 83
Scale-Space Filtering, A.P. Witkin, IJCAI-83, pp. 1019-1022.

Wu et al. 82
Determining Velocities by Propagation, Z. Wu, H. Sun, and L.S. Davis, ICPR-82, pp. 1147-1149.

Yachida 83
Determining Velocity Maps by Spatio-Temporal Neighborhoods from Image Sequences, M. Yachida, Computer Vision, Graphics, and Image Processing 21 (1983) 262-279.

Yam and Davis 81
Image Registration Using Generalized Hough Transforms, S. Yam and L.S. Davis, IEEE PRIP-81, pp. 526-533.

Zuniga and Haralick 83
Corner Detection Using the Facet Model, O.A. Zuniga and R.M. Haralick, CVPR-83, pp. 30-37.