(C) IEEE-PAMI (transactions on Pattern Analysis and Machine Intelligence),

vol. 13, No. 8, pp. 775–790, 1991.

Multidimensional orientation estimation with applications to texture analysis and optical flow

Josef Bigün,
(Corresponding author)
Laboratoire de Traitement des Signaux (DE)
Ecole Polytechnique Fédérale de Lausanne, Ecublens
CH-1015 Lausanne, Switzerland

Gösta H. Granlund and Johan Wiklund
Computer Vision Laboratory
Department of Electrical Engineering
Linköping University
S-581 83 Linköping, Sweden

Abstract

The problem of detection of orientation in finite dimensional Euclidean spaces is solved in the least squares sense. In particular, the theory is developed for the case when such orientation computations are necessary at all local neighborhoods of the n–dimensional Euclidean space. Detection of orientation is shown to correspond to fitting an axis or a plane to the Fourier transform of an n–dimensional structure. The solution of this problem is related to the solution of a well–known matrix eigenvalue problem. Moreover, it is shown that the necessary computations can be performed in the spatial domain without actually doing a Fourier transformation. Along with the orientation estimate, a certainty measure, based on the error of the fit, is proposed. Two applications in image analysis are considered: texture segmentation and optical flow. An implementation for 2–D (texture features) as well as 3–D (optical flow) is presented. In the case of 2–D, the method exploits the properties of the complex number field to by-pass the eigenvalue analysis, improving the speed and the numerical stability of the method. The theory is verified by experiments which confirm accurate orientation estimates and reliable certainty measures in the presence of noise. The comparative results indicate that the proposed theory produces algorithms computing robust texture features as well as optical flow. The computations are highly parallelizable and can be used in realtime image analysis since they utilize only elementary functions in a closed form (up to dimension 4) and Cartesian separable convolutions.

1 Introduction

Since the processing time is decreasing due to hardware development, n–D image processing with applications to time sequences of 2–D images, volume images, time sequences of volume images and segmentation of 2-D images using many feature dimensions are becoming feasible. In many applications realtime image processing is already a reality. The problem of orientation detection arises in many situations in 2–D as well as in higher dimensions in image processing, notably in texture segmentation and optical flow computations. In texture segmentation, the local dominant orientation can be used as a texture feature. Moreover the border of two textures can be locally modelled as an edge to be detected in an n–dimensional texture feature space. In image sequence analysis, the local spatiotemporal spectrum energy is concentrated to a tilted plane when the optical flow is possible to determine unambiguously. To compute the optical flow in this case, the tilt of the plane must be known.

Recent developments in modeling of the human visual system suggest the existence of frequency and orientation selective channels representing the local spatial spectrum, [32, 14, 10, 38], as well as the local spatiotemporal spectrum, [32, 46, 52, 41, 2, 47, 57]. In the latter case an energy concentration in the local spectrum into a particular orientation and frequency channel means a velocity vector with a particular direction and coarseness range. Adelson and Bergen [1] as well as Watson and Ahumada [56] and others have proposed models of human motion perception based on the local spectrum as estimated by spatiotemporal separable filters, e.g. Gabor in the spatial domain and some two lobed derivative filters in the time domain. The models differ qualitatively with respect to the extraction of motion estimates and how to represent the local spectrum. These studies advocate by-passing the point correspondence problem which arises if motion estimation is based on matched features in an image sequence, [53, 58]. They also stress the importance of having certainties attached to the orientation estimates in case the local image spectrum cannot be modeled as an oriented concentration of the energy.

In parallel with these studies much research effort is invested in computationally economic methods which mimick the essence of the discoveries about the human visual system. The spatial frequency decomposition is shown to be achieved efficiently using a pyramidal data structure in combination with separable spatial filtering [51, 12, 16] resulting in the *Laplacian pyramid.* To achieve orientation selectivity by using linear filters in combination with the Laplacian pyramids, Fleet and Jeppson [21] proposed filter design criteria.

Traditionally, the approximation of the local spectrum energy has been dealt with through directional linear filtering in a number of directions (usually about 5 in 2-D and about 10-15 in 3-D). Even if a reasonable approximation of the local spectrum can be obtained, for example by means of a Gabor decomposition with octave center frequency and band width progression, the problem of the analysis and compaction of the energy distribution of the filter responses remains. With support from the psychophysical experiments, [36, 45, 1, 56, 29] propose this analysis to be done with respect to the dominant orientation.

Having this and computational efficiency in mind, we will formulate a general (n–D) least mean square (LMS) problem, the solution of which facilitates the extraction

of dominant orientation and the modelling error by using very few orientation sensitive filters, for a given isotropic frequency channel. We assume that the frequency channels are obtained by utilizing the Laplacian pyramid, [12]. As a particular case of this theory we present the solutions for 2–D and 3–D problems with applications to texture and motion analysis, and show that the same mathematical concept can be used in both cases. A similar approach analyzing the symmetries, can be found in [9, 8, 27].

In Section 2 we will define the n–dimensional linear symmetry which will serve as a model in the subsequent sections. By minimizing the deviation from the model, the local images will be assigned an orientation along with a certainty. The approach is based on the solution of a problem similar to the minimum inertia axis problem of rigid bodies in mechanics, but applied to the Fourier domain [40, 7]. This is shown to be equal to an eigenvalue problem of an $n \times n$ matrix in the spatial domain. Consequently, the computations can be performed entirely in the spatial domain. In Section 3 we will present a discrete implementation technique for the results obtained in the continuous case. We show that the computations involve simple Gaussian and first derivatives of Gaussian convolutions and closed algebraic formulas (up to 4–D) applied to the filter responses. This property, along with the fact that the Gaussians and their partial derivatives are separable, can be utilized to decrease processing time and make the methods suitable for realtime processing. In Section 4 we will apply the methods to 2–D images and present experimental results and comparisons. Here we show that averaging the square of the complex gradient automatically solves the optimization problem. In section 5 we present our comparative results for the 3–D case, applied to the optical flow problem. An algorithm based on spatiotemporal filtering in analogy with previous sections is proposed. Sections 6 and 7 are devoted to the discussion and conclusions.

2 Orientation detection in n-dimensional Euclidean space using linear symmetry.

The approach chosen here to detect orientation in an n–dimensional Euclidean space is based on the Fourier transform[1] although the detection is carried out in the spatial domain.

Let E_n be the Euclidean space with dimension n and f is a real function defined on E_n with F being its Fourier transform. We will sometimes call this function an image and its values as gray values. Later f will be interpreted as a local image obtained by multiplying a window function with the original image.

In the following, if f has an orientation then the locus of f's isogray values, points which have the same values, will consist of either parallel lines or parallel hyperplanes. We note that a line is a linear, 1–D space embedded in n-D, while a plane is a linear $(n-1)$–D space embedded in n–D, as usual. In the case of parallel lines the orientation is represented by the common direction while in the case of parallel planes it is represented by the common normal. Step edges in 3–D (volume segments with different gray values)

[1] An equivalent formulation in the spatial domain utilizing the variational calculus is possible but then the geometric interpretation in the frequency domain is obscured.

and lines in 3-D are examples of images having orientation. Later we will show more explicitly that the problems of describing the orientation in these two cases are reduced to the same problem. Therefore we will first be concerned with the case of planes since it has some notational advantages.

Definition 1 *We will call an image, f, linearly symmetric if its isogray values consist of parallel hyperplanes of dimension $n-1$. That is*

$$f(\bar{r}) = g(\bar{k}^t\bar{r})$$

where g is a one dimensional function ($g: \quad E_1 \to E_1$) and $\bar{k}$ is a constant vector in E_n.

It should be noted that allthough $g(x)$ is a one dimensional function, $g(\bar{k}^t\bar{r})$ is a multi-dimensional function. This property assures that the linearly symmetric images have the same gray values at all points, $\bar{r}$, satisfying $\bar{k}^t\bar{r} = Constant$.

Example 1 *A step edge, $\sigma_n(\bar{r})$, defined in n–D as*

$$\sigma_n = \begin{cases} 1, & \text{if } \bar{k}_0^t\bar{r} \geq 0; \\ 0, & \text{otherwise} \end{cases}$$

is equal to $\sigma(\bar{k}_0^t\bar{r})$, where σ is the usual one dimensional step function

$$\sigma(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Thus σ_n is linearly symmetric.

Lemma 1 *A linearly symmetric image, $f(\bar{r}) = g(\bar{k}_0^t\bar{r})$, has a Fourier transform concentrated to a line through the origin:*

$$F(\bar{\omega}) = G(\bar{\omega}^t\bar{k}_0)\delta(\bar{\omega}^t\bar{u}_1)\delta(\bar{\omega}^t\bar{u}_2)\cdots\delta(\bar{\omega}^t\bar{u}_{n-1})$$

$\bar{k}_0, \bar{u}_1 ... \bar{u}_{n-1}$ are orthonormal, and δ is the dirac distribution. G is the one dimensional Fourier transform of g.

To detect linearly symmetric objects is consequently the same thing as to check the existence of energy concentration to a line in the Fourier domain. This lemma further states that the function $g(\bar{k}_0^t\bar{r})$, which is in general a "spread" function, is compressed to a line. This is a property which will be exploited to construct algorithms checking whether a neighbourhood is linearly symmetric or not.

We will fit an axis through the origin of the Fourier transform domain of a general image, f, which is not necessarily linearly symmetric. Fitting an axis to a finite set of points is classically performed by minimizing the error function:

$$\min_{\|\bar{k}\|=1} e(\bar{k}) = \sum_j d^2(\bar{\omega}_j, \bar{k}) \tag{1}$$

where $d(\bar{\omega}_j, \bar{k})$ is the Euclidean distance between a point, ω_j, in the set and a candidate axis $\bar{k}$. Since in general we have a continuous Fourier transform function, F, defined on

the entire E_n instead of a finite point set, this error function does not suit our purposes as it stands. We use instead the following error function which obtains contributions from all frequencies in the Fourier domain according to their energy in a continuous manner:

$$\min_{\|\bar{k}_l\|=1} e_l(\bar{k}_l) = \int_{E_n} d_l^2(\bar{\omega}, \bar{k}_l)|F(\bar{\omega})|^2 dE_n \tag{2}$$

where dE_n is $d\omega_1 d\omega_1 \cdots d\omega_n$ and the label l represents the line fitting case. If the energy of the Fourier transform is interpreted as the mass-density, then $e_l(\bar{k}_l)$ is the inertia of a mass with respect to the axis $\bar{k}_l$. This error function has a screening effect, that is, F is concentrated to a line if and only if $e_l(\bar{k}_{lmin})$ vanishes in analogy with the finite set case, (1). The distance function, Figure 1, is given by:

$$\begin{aligned} d_l^2(\bar{\omega}, \bar{k}_l) &= \|\bar{\omega} - (\bar{\omega}^t \bar{k}_l)\bar{k}_l\|^2 \\ &= (\bar{\omega} - (\bar{\omega}^t \bar{k}_l)\bar{k}_l)^t (\bar{\omega} - (\bar{\omega}^t \bar{k}_l)\bar{k}_l) \end{aligned}$$

Using matrix multiplication rules and remembering that $\bar{\omega}^t \bar{k}_l$ is a scalar and identical to $\bar{k}_l^t \bar{\omega}$ and $\|\bar{k}_l\|^2 = \bar{k}_l^t \bar{k}_l = 1$, the quadratic form

$$d_l^2(\bar{\omega}, \bar{k}_l) = \bar{k}_l^t (\mathbf{I}\bar{\omega}^t \bar{\omega} - \bar{\omega}\bar{\omega}^t)\bar{k}_l$$

is obtained. Thus equation (2) is expressed in a quadratic form

$$e_l(\bar{k}_l) = \bar{k}_l^t \mathbf{J} \bar{k}_l \tag{3}$$

with

$$\mathbf{J} = \begin{pmatrix} J_{11} & -J_{12} & \dots & -J_{1n} \\ -J_{21} & J_{22} & \dots & -J_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -J_{n1} & -J_{n2} & \dots & J_{nn} \end{pmatrix}$$

where J_{ij}'s are given as

$$J_{ii} = \int_{E_n} \sum_{j \neq i} \omega_j^2 |F(\bar{r})|^2 dE_n \tag{4}$$

and

$$J_{ij} = \int_{E_n} \omega_i \omega_j |F(\bar{\omega})|^2 dE_n \quad \text{when} \quad i \neq j. \tag{5}$$

The minimization problem formulated in (2) is solved by $\bar{k}_l$ corresponding to the least eigenvalue of the inertia matrix, $\mathbf{J}$, of the Fourier domain, [60]. All eigenvalues are real and non-negative and the smallest eigenvalue is the minimum of e_l. The matrix $\mathbf{J}$, contains sufficient information to allow computation of the optimal $\bar{k}_l$ in the least square error sense, LSE, given by equation (2).

This error will be exactly zero if and only if f is a linearly symmetric image. The obtained orientation will be unique if the least eigenvalue has the multipicity 1. When the multiplicity of the least eigenvalue is larger than 1, there is no unique axis $\bar{k}$, by which the image can be described as $g(\bar{k}_l^t \bar{r})$ for some one–dimensional function g. Instead, the energy in the Fourier domain is distributed in such a way that there are plenty of axes which give the least square error. More exactly, these axes are any axes given by a linear

combination of the eigenvector space belonging to the least eigenvalue. Here it should be observed that the dimension of this space is equal to the multiplicity of the eigenvalue it corresponds to (the least one). This is due to the fact that $\mathbf{J}$ is positive semi-definite and symmetric by definition, (2) and (3). In other words, there is no optimal and unique axis passing through the origin but a unique $(n-m)$-D linear vector space when the multiplicity of the least eigenvalue, m, is greater than one.

At this point the question whether the degenerated solutions of the line fitting problem are at the same time solutions to other similar problems arises. Suppose that we are intending to approximate the function F by a hyperplane with the dimension $n-1$ under the condition that the hyperplane passes through the origin. Representing the normal of the hyperplane by $\bar{k}_p$, we must minimize

$$\min_{\|\bar{k}_p\|=1} e_p(\bar{k}_p) = \int_{E_n} d_p^2(\bar{\omega}, \bar{k}_p)|F(\bar{\omega})|^2 dE_n \tag{6}$$

where the label p represenents the plane fitting case. The difference here is the distance function $d_p(\bar{\omega}, \bar{k}_p)$ which is the perpendicular distance of the point $\bar{\omega}$ to the hyperplane represented by its normal vector $\bar{k}_p$

$$d_p^2(\bar{\omega}, \bar{k}_p) = (\bar{\omega}^t \bar{k}_p)^2 = \bar{k}_p^t \bar{\omega}\bar{\omega}^t \bar{k}_p \tag{7}$$

Thus (6) reduces to minimize

$$e_p(\bar{k}_p) = \bar{k}_p^t \mathbf{A} \bar{k}_p \tag{8}$$

where

$$A_{ij} = \int_{E_n} \omega_i \omega_j |F(\bar{\omega})|^2 dE_n \tag{9}$$

Comparing equation (9) with (4) and (5) yields an algebraic relationship between these two different problems:

$$\mathbf{J} = \text{Trace}(\mathbf{A})\mathbf{I} - \mathbf{A} \tag{10}$$

where "Trace($\mathbf{A}$)" is the sum of all eigenvalues of $\mathbf{A}$ and can be computed by summing up $\mathbf{A}$'s diagonal elements. In analogy with $\mathbf{J}$, $\mathbf{A}$ is also positive semi definite and the solution to the plane fitting problem is given by the least eigenvalue of $\mathbf{A}$ and its corresponding eigenvector(s). The matrix $\mathbf{A}$ defined by (9) is often referred to as a *scatter matrix* in the literature and arises, [44], when the most economic basis is to be found to represent a given set of n dimensional data, $\{\bar{\omega}_i\}$. This problem then leads to Karhunen-Loewe expansion which is also known as principle value decomposition.

Lemma 2 *The matrices* $\mathbf{J}$ *and* $\mathbf{A}$ *have common eigenvectors that is*

$$\mathbf{J}u = \lambda' u \quad \Leftrightarrow \quad \mathbf{A}u = \lambda u \tag{11}$$

with

$$\lambda' = \mathit{Trace}(\mathbf{A}) - \lambda. \tag{12}$$

Thus fitting a line to F *is algebraically equivalent to fitting a plane to it.*

The lemma is a consequence of the fact that **J** and **A** commute but can also be proved immediately by utilizing the relationship (10) and operating with **J** on u which is assumed to be an eigenvector of **A**. Given this Lemma one has the following conjecture which assumes that the eigenvalues of **A** are enumerated in ascending order, $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$.

Conjecture 1 *If and only if the eigenvalues of the scatter matrix related to F fulfill the relationships*

$$\lambda_1 = \lambda_2 \cdots = \lambda_{n-1} > 0 \qquad and \qquad \lambda_n = 0 \tag{13}$$

then the energy of F is concentrated to a plane *through the origin. The normal of this plane is given by u_n. Similarly if and only if*

$$\lambda_1 > 0 \qquad and \qquad \lambda_2 = \lambda_3 \cdots = \lambda_n = 0 \tag{14}$$

is fulfilled then the energy of F is concentrated to a line *through origin. The direction of this line is given by u_1.*

The conjecture will be exploited later in Section 5 to construct a certainty measure.

3 Discrete approximation

Up to this section the theory for detection of orientation in multidimensional Euclidean space has been based on continuous signals. The method examines whether an image consists of an $n-1$ dimensional hyperplane or a line, by eigenvalue analysis of the matrix **A**. The computation of the matrix elements themselves are given by equation (9) for the continuous case in the Fourier domain.

In practical situations, however, we would like to have a good approximation of the matrix **A** by measurements based on discrete data, and this should, in many applications , be done quite often. For example we would like to do it for local images around every point. For this reason an economic computation scheme providing the elements of **A** is necessary.

The first step in this direction is to eliminate the need of Fourier transformation. This can be done by utilizing the Parseval theorem which translates the computation of matrix elements from the Fourier domain to the spatial domain:

$$A_{ij} = \frac{1}{4\pi^2} \int_{E_n} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} dE_n \tag{15}$$

The next step is to find a discrete approximation of A_{ij} given by equation (15). A good approximation can be obtained using finite element methods, [4], yielding a function dependent non uniform mesh. Szelski, [50], proposed an interpolation method for computer vision applications based on hierarchical interpolation functions yielding uniform meshes. Recently Jähne, [35], reported accurate estimates of (15) on a uniform mesh utilizing extrapolation techniques. Here we will choose a simple method yielding uniform discretization and allowing fast implementation with a high degree of parallelism.

Assuming $f(\bar{r})$ now being the original image instead of the local image, we can, without loss of generality, choose the center of the neighborhood (the examined point) to be the origin of the coordinate system. That is, we approximate (15) by utilizing

$$\frac{\partial f(\bar{r})}{\partial x_i}\frac{\partial f(\bar{r})}{\partial x_j} \approx \sum_l \frac{\partial f(\bar{r}_l)}{\partial x_i}\frac{\partial f(\bar{r}_l)}{\partial x_j}\mu_l(\bar{r})w(\bar{r}) \tag{16}$$

where $\bar{r}_l$ is the n dimensional coordinate vector of a mesh point, $w(\bar{r})$ is the window function and $\mu_l(\bar{r})$ is the continuous interpolation function. This approximation is exact if μ_l is chosen as the interpolation function corresponding to the sampling of the original image with an over sampling factor of 2 in each dimension and w is chosen as a band limited window function, [19]. However the disadvantages of such a choice include large filters which are either rotationally symmetric but not separable or separable but not rotationally symmetric. We will choose both μ_l and w as two Gaussians which are the only functions being rotationally symmetric and separable in Cartesian coordinates. This alternative yields filters which are small in size since a Gaussian is concentrated both in the spatial and the Fourier domain. Thus we have

$$\mu_l(\bar{r})w(\bar{r}) = \exp\,(\,-\frac{1}{2\sigma_p^2}\|\bar{r}-\bar{r}_l\|^2)\exp\,(\,-\frac{1}{2\sigma_w^2}\|\bar{r}\|^2) \tag{17}$$

where σ_p and σ_w are two constants which control the effective width of the two Gaussian functions. Substitution in equation (15) yields

$$A_{ij} \approx \frac{1}{4\pi^2}\sum_l \frac{\partial f(\bar{r}_l)}{\partial x_i}\frac{\partial f(\bar{r}_l)}{\partial x_j}m_l \tag{18}$$

where m_l is

$$\begin{aligned} m_l &= \int_{E_n} \mu_l(\bar{r})w(\bar{r})dE_n \\ &= \frac{2\pi}{4}\frac{\sigma_p^2\sigma_w^2}{\sigma_p^2+\sigma_w^2}\exp\,(\,-\frac{1}{2(\sigma_p^2+\sigma_w^2)}\|\bar{r}_l\|^2). \end{aligned} \tag{19}$$

Equation (18) is equivalent to a discrete convolution by a Gaussian if A_{ij} is computed for every point in the original image. Since m_l's, the filter coefficients, decrease rapidly outside of a circle with radius $\sqrt{\sigma_p^2+\sigma_w^2}$, we can truncate the infinite filter when its coefficients are sufficiently small. In our experiments this is done when the coefficients have decreased to about 1% of their maximum in magnitude. Thus equation (18) implies

$$\mathbf{A} \approx \frac{1}{4\pi^2}\sum_l (\nabla f_l)(\nabla f_l)^t m_l. \tag{20}$$

where ∇f_l is the gradient of $f(\bar{r})$ at the discrete image position $\bar{r}_l$.

4 Orientation detection in 2–D

In the 2-D case, there have been solutions to the problem of finding the dominant local orientation, for example [34, 17], by projecting the neighbourhood onto a number of fixed orthogonal functions. The projection coefficients are then used to evaluate the orientation parameter of the model. When the number of filters used is increased, the local image is described better and better in terms of the filters. But the inverse function, mapping the coefficients to the optimal orientation, if it exists at all, increases greatly in complexity when mapping the obtained projection coefficients (filtering results) to orientations. Another drawback of these methods is that their generalization to higher dimensions is not obvious.

The proposed computations in the previous section are simplified further by utilizing some fundamental properties of the complex **z**-plane, implying that the matrix eigenvalue analysis is replaced by its equivalent: a complex convolution. Moreover, because of the low dimensionality, the "fitting a plane" problem as discussed in Section 2 becomes identical to "fitting a line" problem simplifying the analysis further. We present two certainty measures and an orientation estimate which are based on the eigenvalues and eigenvectors of the scatter matrix of the local spectrum, **A**.

We consider $t\bar{k}$, $t \in E_1$, which represents a line and observe that $-t\bar{k}$ is an equivalent representation of the same axis. Thus *the orientation of a line* $t\bar{k}$ *can be defined as* 2θ, if θ is the direction angle of $\bar{k}$ given by

$$\bar{k} = (\cos\theta, \sin\theta)^t \tag{21}$$

since the function 2θ maps θ and $\theta + \pi$ to the same angle. Let $\bar{x}$ be a two–dimensional vector:

$$\bar{x} = (x, y)^t \tag{22}$$

Define $\mathbf{x}$ to be the complex number obtained by:

$$\mathbf{x} = x + iy \tag{23}$$

Then the relationship

$$\bar{x}_1^t \bar{x}_2 = \mathrm{Re}(\mathbf{x}_1^* \mathbf{x}_2) = \frac{1}{2}[\mathbf{x}_1^* \mathbf{x}_2 + (\mathbf{x}_1^* \mathbf{x}_2)^*] \tag{24}$$

holds. Since $\|\bar{k}\| = 1$ and **A** is positive semi definite, $\bar{k}$ maximizing $\bar{k}^t \mathbf{A} \bar{k}$ is the eigenvector belonging to the largest eigenvalue of **A**, [60]. By using (24) and (20) we obtain

$$\begin{aligned}
\bar{k}^t \mathbf{A} \bar{k} &= \frac{1}{4\pi^2} \sum_j \bar{k}^t (\nabla f_j)(\nabla f_j)^t m_j \bar{k} \\
&= \frac{1}{4\pi^2} \sum_j (\mathrm{Re}(\mathbf{k}^* \nabla \mathbf{f}_j))^2 m_j \\
&= \frac{1}{4\pi^2} \sum_j \frac{1}{4} ((\mathbf{k}^2)^* (\nabla \mathbf{f}_j)^2 + [(\mathbf{k}^2)^* (\nabla \mathbf{f}_j)^2]^* + 2|\nabla \mathbf{f}_j|^2) m_j \\
&= \frac{1}{4\pi^2} \sum_j \frac{1}{2} |\nabla \mathbf{f}_j|^2 m_j + \frac{1}{2} \mathrm{Re}((\mathbf{k}^2)^* (\nabla \mathbf{f}_j)^2) m_j
\end{aligned} \tag{25}$$

Here, as defined by (22) and (23), $\nabla \mathbf{f}_j$ and $\mathbf{k}$ are the complex interpretations of the real 2-D vectors ∇f_j and $\bar{k}$. The complex number $\mathbf{k}^2$ maximizing (25) is the same as the one which maximizes:

$$\sum_j \mathrm{Re}((\mathbf{k}^2)^*(\nabla \mathbf{f}_j)^2)m_j = \mathrm{Re}((\mathbf{k}^2)^* \sum_j (\nabla \mathbf{f}_j)^2 m_j) = \mathrm{Re}((\mathbf{k}^2)^* \mathbf{z}) \tag{26}$$

where

$$\mathbf{z} = \sum_j (\nabla \mathbf{f}_j)^2 m_j. \tag{27}$$

By remembering that $|\mathbf{k}^2| = 1$ and interpreting (26) geometrically by means of equation (24) as a scalar product of two vectors, $\mathbf{k}^2_{max}$ is obtained as:

$$\mathbf{k}^2_{max} = \mathbf{z}/|\mathbf{z}|. \tag{28}$$

Consequently the complex number, $\mathbf{k}^2_{min}$, minimizing (25) is given by $-\mathbf{k}^2_{max}$. Substituting $\mathbf{k}^2_{min}$ and $\mathbf{k}^2_{max}$ in (25) yields:

$$\lambda_0 = e(\bar{k}_{min}) = \frac{1}{8\pi^2}(\sum_j |\nabla \mathbf{f}_j|^2 m_j - |\mathbf{z}|) \tag{29}$$

$$\lambda_1 = e(\bar{k}_{max}) = \frac{1}{8\pi^2}(\sum_j |\nabla \mathbf{f}_j|^2 m_j + |\mathbf{z}|). \tag{30}$$

That is:

$$|\mathbf{z}| = 4\pi^2(\lambda_1 - \lambda_0) \tag{31}$$

$$\arg \mathbf{z} = 2\theta_0. \tag{32}$$

Thus only computation of $\mathbf{z}$ is sufficient for knowing whether the considered neighbourhood is linearly symmetric or not. The argument of $\mathbf{z}$ will be a representation of the optimal orientation given the data and the interpolation function. Hu, [31], has presented a pattern reconition technique based on the complex moments of the gray values of the image using moment invariants. Second order complex spatial moments of his method are similar to those presented here for the 2-D local spatial frequency spectrum. Computation of $\mathbf{z}$ can be done by an averaging of the *complex derivative image,* according (27). Performing this for every neighbourhood of an image is possible through a convolution and results in the first scheme, Figure 2. The certainty in this approximation of the local orientation depends on the behaviour of the eigenvalues λ_0 and λ_1, according to the discussion in Section 2. An existence of a definite orientation in the neighbourhood is probable if λ_0 is small relative to λ_1. An 'ideal' case occurs when $\lambda_0 = 0$ and $\lambda_1 >> 0$. The magnitude of the resulting image measures this very property, and will be referred to as a certainty measure, C_{f1}.

$$C_{f1} = 4\pi^2(\lambda_1 - \lambda_0) \tag{33}$$

A high C_{f1} indicates that the orientation found is reliable.

C_{f1} is an energy dependent certainty measure, that is, the certainty decreases as the contrast of the neighbourhood decreases. In many applications however there is a need for an energy independent orientation detection. The measure

$$C_{f2} = (\frac{\lambda_1 - \lambda_0}{\lambda_1 + \lambda_0})^c \tag{34}$$

has such a property where c is a positive constant, for the purpose of controlling the dynamic range. C_{f2} is defined as 0 when $\lambda_1 + \lambda_0 = 0$. This happens only when we have a constant image in the neighbourhood and thus no unique orientation is present. C_{f2} attains the maximum value 1 if and only if $\lambda_0 = 0$, since both λ_0 and λ_1 are non–negative. It decreases when the difference between the eigenvalues decreases. This property effectively tests whether the multiplicity of λ_0 is 2, in which case there is not a unique orientation minimizing the error. Utilizing (29) and (30) provides

$$4\pi^2(\lambda_1 + \lambda_0) = \sum_j |\nabla \mathbf{f_j}|^2 m_j \tag{35}$$

and results in the second scheme, Figure 2, implementing the ideas given in Sections 2 and 3. The choice of C_{f2} as in (34) can be replaced by any other energy independent relation of the eigenvalues fitting to one's purpose since these eigenvalues are provided through convolutions given by equations (29) and (30).

These two schemes can be summarized by the compact notation

$$\begin{aligned} 2\theta_0 &= \arg((\nabla \mathbf{f})^2 * m) \\ C_{f1} &= \left|(\nabla \mathbf{f})^2 * m\right| \\ C_{f2} &= (\frac{|(\nabla \mathbf{f})^2 * m|}{|\nabla \mathbf{f}|^2 * m})^c \end{aligned} \tag{36}$$

where the symbol $*$ represents the usual convolution operation and $\arg(\cdot)$, $|\cdot|$ and $(\cdot)^c$ operations are being applied pointwise to their arguments which are complex or real images. The first scheme computes $2\theta_0$ and C_{f1} and the second scheme computes $2\theta_0$ and C_{f2}.

4.1 Accuracy evaluation

In this section, by using a specially constructed test image, we evaluate the extent of the validity of the 2–D implementation technique proposed previously. However, because of the nature of the tests, these results concerning accuracy are indicative for dimensions higher than 2 as well. Figure 3 shows the test image (512×512 and 256 gray levels) which contains sine waves with exponentially increasing frequencies in the radial direction, including all possible frequencies and directions. Gaussian uncorrelated white noise is added to the right half of the image with the proportion 1:3, that is $0.25 f_i + 0.75 Y_i$, where f_i is the image intensity and Y_i is the stochastic variable with the distribution of N(0,32).

Figure 4 shows the local orientation, $2\theta_0$, estimate of the test image along the line shown in Figure 3. Signal 1 is obtained by means of a 9×9 complex derivative filter (partial derivatives of a Gaussian) and a 15×15 averaging filter (a Gaussian). Signal 2

will be dealt with later. Ideally Signal 1 should be constant. At the left part of Signal 1 where no noise is added this is perfectly the case, while at the right part fluctuations due to the noise are observed. Signal 1 deviates from its true, constant value by ± 0.01 radians in the average in the noiseless (except for the quantization noise) region, while in the noisy region the measurement of the corresponding deviation needs some care since the certainty measures at these points are far from uniform, Signal 3 Figure 5. As a consequence it is possible to know when the orientation estimate is reliable and hence such an estimation is meaningfull. Measuring the orientation deviation at those parts of the signal where the energy independent certainty measure claims a moderate degree of confidence, is one way to judge the the orientation accuracy quantitatively. By using this technique, in particular when $C_{f2} > 0.5$ with $c = 1$ in Signal 3 of Figure 5, the orientation deviation in the noisy part is found to be ± 0.33 radians on the average. The original image, as well as the intermediate and final images were all quantized to 256 quantization levels. Noting that the error measurements were based on the quantized results and 75 % of the signal amplitude consisted of noise in one part of the signal, one can infer that the orientation estimates are accurate and robust both in presence of quantization noise and Gaussian additive noise. The quantization error robustness is particularly important when an algorithm is intended to be implemented on simple hardware using integer arithmetic processors for the purpose of speed and economy.

The signals in Figure 5 illustrate the certainty measures C_{f1} and C_{f2} corresponding to the orientation estimate signal commented previously (Signal 1 in Figure 4). Of these curves 1 and 3 correspond to C_{f2} with c=6 and c=1 respectively. Increasing c from 1 to 6 suggest a dramatical change of the dynamic range of C_{f2}. However, this is only observable for the noisy part of the image as illustrated by the corresponding signals. The certainty level of the noiseless part is very close to 1 so that the mentioned change of the dynamic range has a very small effect. C_{f2} is given by Signal 1. Since C_{f1} is not a relative measure like C_{f2} is, it has a high degree of frequency dependence which can be observed as a peak in Signal 2 (left part). In comparison with the signals 1 and 3, Signal 2 is considerably more "suspect" outside of an effective pass band. The frequency sensitivity band of the certainty parameter C_{f1} is due to the derivation in the first step, and averaging in the second step. In the Fourier domain, at low frequencies, this roughly corresponds to a multiplication of an increasing and a decreasing function. The center frequency can thus be varied by varying the scale of the filters at the two steps, Figure 6. Thus, the method permits to perform orientation measurements corresponding to a given frequency band. However, if many frequency bands are to be studied with respect to orientation at the same time, it is computationally more efficient to decompose the original in a Laplacian pyramid and apply the orientation estimation technique described here to each level of the Laplacian pyramid separately. The second alternative will be utilized in the texture analysis application below.

It can be observed that the size of the partial derivative filters affects the accuracy of the orientation detection at high frequencies more than the size of the averaging filter if the original image, as in the case of our test image, is not over sampled by a factor of 2 in each dimension. Signal 2 in Figure 3 which illustrates the orientation estimate, shows this effect. The signal is obtained as a result of a 5×5 complex derivative and a 21×21 averaging filter configuration and should be compared to Signal 1, obtained as a

result of 9×9 and 15×15 filter configuration. It can be seen that the size of the complex derivative filter is not large enough to suppress the undesired high frequency components, and increasing the spatial extent of the averaging does not compensate for that failure. Accurate orientation estimate at high frequencies can be achieved by resampling the discrete data, [44], so that the over sampling condition mentioned is fulfilled or by using another approximation than the simple one proposed by equation (16).

4.2 Texture analysis application and evaluation

Texture analysis has been used in classification tasks concerning 2–D images in general, and in the segmentation of aerial images in particular. An important means in the course of the analysis is extraction of texture features. Given a set of texture features, the original image can be segmented using unsupervised or supervised segmentation techniques. These exploit the observation that in the feature space, a well chosen set of features induce well separated clusters corresponding to different classes.

Feature extraction has been effectuated by computing characteristics of the autocorrelation function, partitioning of the Fourier transform energy into radial and angular bins, [5], interpretation of co-occurrence matrices through their moments, [28], identification of markov random processes parameters [23], energy of real valued special texture masks [37], [54]. The first two methods exploit the fact that a texture is repetitive. A consequence of this is that peaks occur in the autocorrelation function and a concentration of the energy is observed in the frequency domain. These methods are able to take into account textures with repetition frequencies ranging from very low to very high values while the other methods are restricted to the size of a prescribed neighbourhood.

The textural structures which can be described within a neighbourhood are naturally limited to those which are observable within the size of the neighbourhood. Thus a feature, based on measurements within a neighborhood, fixed in size, has a poor discrimination power when applied to textures not observable within the neighborhood because of the wrong scale. But in general the size information is not available. The psychophysical experiments indicate, [10], [33], that there exists frequency and orientation selectivity in the human visual system, giving a hint of how this *scaling problem* can be solved. The scaling problem has, in image processing, been treated by filtering with Laplacian filters with different sizes, [39], [12] (octave band pass filters) or using different decompositions in Gabor, [22], Wavelet, [26], and Wigner, [59], basis functions. In [13], the moments of Laplacian pyramid images are proposed providing anisotropic information about the local image. In [21], orientation selectivity within a frequency band is proposed to be done by filtering a Laplacian image with a directional cosine filter. Such an approach results in spatial phase dependency, that is the response to a sinusoid is also a sinusoid. This is not desirable in texture analysis because one would like to have a uniform response at those regions of the image which have a dominant orientation. To achieve phase independency, the magnitudes of Gabor filter responses or quadrature filter responses can be used. To detect the orientation [36] proposes interpolation on the magnitudes of the filter responses of quadrature filters. The used filters however, are non-separable and numerous and therefore require dedicated hardware.

The original image, with the true borders superimposed for the purpose of elucidation,

which we used in the subsequent tests is presented in Figure 7. The texture patches in the original are cut from real aerial images. They were individually normalized with respect to illumination (mean) and contrast (variance) before they were pasted together to form the 256×256 original. The individual textures, 7 in total, represent some typical forest, field, and residential areas. They are repeated within a 4×4 configuration such that every texture is neighbour of any other at least once for the sake of evaluation. In contrast with the texture patches from Brodatz album [11], the presented textures are extracted from images formed during aerial imagery and are contaminated by a large amount of non artificial noise. They were moreover selected by a group of 3 people, independent of the authors, with experience in aerial images to decrease the unconscious bias in the choice of interesting textures.

We propose to use the linear symmetry along with the Laplacian pyramids approach, to produce the feature images. To be more explicit, we first decompose the original image into its components of different resolutions by means of the Laplacian pyramid, [12]. To each level in the Laplacian pyramid i.e. a frequency channel, we apply a linear symmetry measure. In the particular application below we have used the second scheme proposed in the previous section with 9×9 complex derivative, 17×17 Gaussian filter. The number of frequency channels were 4. The real and imaginary parts of the resulting images are used as features along with the average local energy (within the same 17×17 Gaussian window as before) resulting in 3 real features per frequency channel. The feature images are blown up to the size of the original using the technique discussed in [12]. The number of features are reduced using the local Karhunen-Loeve transform as suggested by [6]. The clustering algorithm proposed by [49] is applied to these to obtain the unsupervised segmentation result illustrated by Figure 8. The result predicts 6 classes of which 5 correspond to true classes and the remaining class correspond to a merging of two true classes. Almost all borders are found with a good accuracy. It should be noted that the classification as well as feature reduction is unsupervised. Thus the border quality should be compared with other results, like the subsequent ones to be presented, using unsupervised procedures.

Figure 9 illustrates the result when the texture energy measures proposed by [37] are utilized as features along with the same feature reduction and clustering technique as before. The texture energy planes are obtained by first convolving the original by 7 filters (E5l5, L5E5, R5R5, E5S5, S5E5, L5S5, S5L5 as labeled by [37]) and then applying the standard deviation filtering using a 7×7 window. We have also tried larger sizes which resulted in a poorer class and border performance compared to Figure 9. The feature dimensionality reduction process of [6], tested also on many other different feature sets, seems to be an adequate means to compress the information and increase the discrimination power of the texture features, a matter which is also dealt with by [55] recently. The result suggests that 4 classes exist of which two more or less represent two classes while the two other represent the merged classes. Many of the borders are identified with a good accuracy.

Figure 10 illustrates the segmentation result when the features, proposed by [54] are utilized. These features, totaling to 4, are obtained by first computing the discrete Hadamard transform within a running neighbourhood of 2×2 and then applying the variance filtering to these in a window of 8×8. In the result 4 classes are possible to distinguish and of these only one corresponds to a true class while the remaining being the

merged classes. Some of the borders are possible to identify with a reasonable accuracy.

We have also tried to obtain a segmentation result using the 13 features proposed by [28]. The used features represented certain, in [28] well defined, properties (angular second moment, contrast,... etc.) of the cooccurrence matrix with 64 quantization levels computed in 9×9 neighbourhoods using a distance vector of $(0,1)^t$. The result was a single class for the entire image and naturally no borders.

A general explanation for why the linear symmetry have performed well is that the textures in the test image differ in their "repetition frequency" e.g. scale, considerably while the used neighborhoods in the 3 compared methods were constant. On the other hand, a straight forward application of these techniques to every level of a multiresolution pyramid would result in too many texture features for further unsupervised processing. Moreover, the linear symmetry features are directly measuring the dominant orientation information in each frequency channel. This is a property which is in line with the current understanding of the human visual system. The results of these unsupervised experiments and others we have done using aerial images suggest that the linear symmetry features can be used successfully in texture modeling.

5 Application to image sequences

Optical flow computation has been a subject of an intensive research during the last decade because of its potential and realized applications in image sequence coding, object tracking, surveillance, robot control and many other domains. An exposure of the applications and the methods computing the optical flow with brief summaries can be found in [42]. We directly propose our method based on linear symmetry and then in the sequel we establish the relationship and differences with existing relevant methods. Further on we present our experimental results.

Two types of local motion will be analyzed within the same model: The pure translation of a line and the coherent translation of a random dot pattern. We think that the rotational components of the motion can be dealt with at a higher level of the motion analysis given the local velocity estimates (local linear approximations).

Let $f(x,y,t)$ represent a spatiotemporal local image registered by a visual sensor system. When f is generated by the translation of a line the velocity component of the translation along the line is not possible to determine locally. In general this problem occurs whenever the isogray values of f consists of parallel 2–D planes as in the case of the translation of the sine stripes used in the experiments of Adelson and Movshon [2]. This phenomenon is often referred to as the aperture problem. In this case the translation is implicitly given by the normal vector of the parallel planes of the spatiotemporal domain. If the normal of the planes is represented by the unit vector $\bar{k} = (k_x, k_y, k_t)^t$ then by using geometry, Figure 11, the components of the normal velocity vector $\bar{v}$, is determined by the formulas:

$$v_x = -\frac{k_x k_t}{k_x^2 + k_y^2} \tag{37}$$

$$v_y = -\frac{k_y k_t}{k_x^2 + k_y^2}. \tag{38}$$

In the linear symmetry model, Conjecture 1, the normal vector is given by the eigenvector belonging to the largest eigenvalue of the scatter matrix $\mathbf{A}$, under the condition that the smallest eigenvalue is zero and it has the multiplicity 2.

When the local intensity function f is generated by the coherent translation of a random dot pattern or more precisely when the isogray values of f consist of parallel lines, the motion can be determined unambiguously. When $(k_x, k_y, k_z)^t$ represents the direction of these parallel lines, the velocity components are given by

$$v_x = \frac{k_x}{k_z} \tag{39}$$

$$v_y = \frac{k_y}{k_z} \tag{40}$$

In the linear symmetry model, this direction is given by the eigenvector belonging to the smallest eigenvalue of the scatter matrix $\mathbf{A}$, under the conditions that the largest eigenvalue has the multiplicity 2 and the smallest eigenvalue is zero.

The value of $\bar{v}$, obtained by (37-38) or (39-40) is significant and to be computed only when a high certainty degree about the type of the motion exists. The relative positions of the three eigenvalues of $\mathbf{A}$ enable us to infer about the quality of the optical flow estimate as well as the type of the motion. Using Conjecture 1, We propose

$$C_f = \frac{2\lambda_2}{\lambda_1 + \lambda_3} - 1 \tag{41}$$

to be the measure to be used as a discriminator and a certainty degree. Remembering that the eigenvalues are related to inertia measurements with respect to the principal axes and having Conjecture 1 in mind, many other measures can be constructed to analyze the energy distribution in the frequency domain. If the sign of C_f is positive the optical flow estimate is unambiguous and of the type "translation of a random dot pattern"; if it is negative then the optical flow estimate is ambiguous and the estimate is of the type "translation of a line". The magnitude of C_f which has the upper bound 1, increases gradually as the the type of the motion becomes more certain. If and only if $C_f = 1$ then one has an unambiguous optical flow in the local image while if and only if $C_f = -1$ one has the ambiguous flow case. When $C_f = 0$ there is total uncertainty about the nature of the motion.

5.1 Relationship with some existing methods

Depending on the chosen mathematical model, the exact meaning of optical flow has been defined differently in the computer vision literature. In this section we will compare two different definitions with the definition contained in the concept of linear symmetry.

One of the most common models uses variational calculus, Horn and Schunk [30]. In this approach $\frac{df}{dt}$, which depends on optical flow is minimized over a region of E_2. On the solution a smoothness constraint assuring a smooth variation of the optical flow is imposed.

$$\min_{\bar{k} \in S_{hs}} = \int_{E_2} (\bar{k}^t \nabla f)^2 dxdy$$

$$+ \quad \lambda \int_{E_2} (\frac{\partial k_x}{\partial x})^2 + (\frac{\partial k_x}{\partial y})^2 + (\frac{\partial k_y}{\partial x})^2 + (\frac{\partial k_y}{\partial y})^2 dxdy \tag{42}$$

where $\bar{k}$ is a point in S_{hs} which represents a plane consisting of the points $(u, v, 1)^t$ with u and v being the x and y components of the optical flow vector. E_2 is an image region and λ is a heuristic positive coefficient regulating the importance of the smoothness in the solution. The coefficient is chosen large when the gray value measurements are noisy, small otherwise. It should be noted that there is no time integration in the expression above. The minimization is possible to perform by solving a second order elliptic partial differential equation related to this problem. A feature of the smoothness constraint is that when there is no unique optical flow in a region then unique solutions from other parts are propagated towards this region. The propagation direction is uniform in the original approach but [43] discusses a modified version of the smoothness constraint which depends on the gray value variations and hence allows oriented smoothness resulting in improved optical flow estimates along object boundaries. The equivalent minimization problem formulated in this paper by the linear symmetry concept takes the form

$$\min_{\bar{k} \in S_{ls}} e(\bar{k}) = \int_{E_3} (\bar{k}^t \nabla f)^2 dxdydt \tag{43}$$

where S_{ls} represents the surface of the unit sphere. Once $\bar{k} \in S_{ls}$ is obtained the optical flow vector is defined as the intersection point of $\bar{k}$ with S_{hs}. The intersection point may degenerate to a line in which case the optical flow is possible to compute only along the gradient and is represented by the point of the line closest to $(0, 0, 1)$. Putting $\lambda = 0$ makes both approaches possible to compare to each other conceptually, since we don't propose that the optical flow should be propagated from regions with good estimates to regions with poor estimates. We advocate, as in [25, 3, 1, 56, 29], that certainty measures, should be used to keep track of the regions with poor estimates and thereby to facilitate for making a proper use of them in the subsequent analysis. Such an analysis could be token tracking as in [48, 20].

An important difference is that the linear symmetry approach also takes account of the errors made in the time direction contributing to the robustness of the approach in the presence of temporal noise. The solution spaces are normally different. However when zero error is reached by (43) for some optical flow solution, then the same error (zero) is attained by (42) as well. This corresponds to the ideal case when the energy of the 3–D Fourier transform of the local image is concentrated to a Dirac plane or to a Dirac line.

Adelson, [1], has proposed a model for human motion perception based on the energies of frequency tuned quadrature filters (Gabor like). Heeger, [29], has recently proposed an algorithm for detecting optical flow using similar filters. He analyses the Fourier transform of the image on the basis of the extracted energy responses using Gabor filters tuned to different frequencies. In the analysis the actual motion (the responses of the Gabor filters) is approximated by the motion of a translating random dot pattern (a dirac plane in the frequency domain). The approximation is realized by minimizing the squared error of the fit through varying the tilt, the intersection point of the normal of the Dirac plane and the S_{hs} plane.

$$\min_{\bar{k} \in S_{hs}} e(\bar{k}) = \sum_{i=1}^{12} (\bar{m}_i \frac{R_i(u, v)}{\bar{R}_i(u, v)} - m_i)^2 \tag{44}$$

where m_i is the magnitude of the i'th filter response, $\bar{m}_i$ the average of m_i over the filters with the same orientation as the i'th, R_i is the magnitude of the i'th filter response when the input is a dirac plane with the tilt represented by u and v, $\bar{R}_i$ is the averaged R_i over the same orientation. The function $\frac{\bar{m}_i}{\bar{R}_i(u,v)}$ is used as a factor of $R_i(u,v)$ to compensate for the energy differences of the responses of the Dirac plane and the input. The error function is a 2–D function defined on the S_{hs} plane and has a unique point at which the error reaches its minimum when there is a unique optical flow while if there is an aperture problem the optimum is reached along a line. The kind of motion is determined on the basis of whether the error function is elongated or rotationally symmetric by computing the curvature of the error function. The accuracy of the results depend on whether the range of the optical flow to be estimated is known a priori or not. Also the computational expenses may be very difficult to overcome due to necessary nonlinear optimization and the numerous filterings. Both in the linear symmetry approach and the restricted Horn and Shunk approach, ($\lambda = 0$), the error functions are quadrics for which the minimization is not more intricate than solving for the roots of a third order polynomial[2] or solving for a 2×2 linear equation system.

5.2 Experiments on motion estimation and evaluation

To test the proposed method we have implemented the scheme given in Figure 12. The flow chart consists of three steps: gradient estimate and auto outer product computation, averaging and eigenvalue analysis. To compute optical flow corresponding to several spatial frequency channels, the Laplacian pyramid filtering should be applied to each image in the original sequence before applying the optical flow scheme.

The gradient is estimated by convolving the image sequence with separable 3–D kernels, $7 \times 7 \times 7$ each, given by the partial derivatives of a Gaussian, one for dx, dy and dt respectively. This means that some averaging is done in the two directions perpendicular to the gradient direction of the kernel. The gradient estimates are then combined to the following quantities; $dxdx, dydy, dtdt, dxdy, dxdt, dydt$ (since the scatter matrix is symmetric only six components are needed).

The second step is to average the six components from the previous step. This is done by applying a 3–D Gaussian averaging kernel, $11 \times 11 \times 11$, to each component. That is, each component is smoothed in all three dimensions.

The third step takes the smoothed components and builds the matrix $\mathbf{A}$ (eqn. 20). The eigenvalues are computed and are used to obtain the certainty measure proposed by (41). One large eigenvalue and two small ones correspond to a moving line, the aperture problem. Two large eigenvalues and a small one correspond to a moving point, that is there is no aperture problem. The sign of the certainty vector is utilized to determine which type of optical flow vector to compute. When all eigenvalues are equal and zero the certainty measure is defined to be zero. The eigenvalue analysis at such points is avoided by assuring that $\text{Trace}(\mathbf{A}) > 0$ is fullfilled at the points where the analysis is carried out.

Two different image sequences have been used in the experiments, one synthetic and one natural. The synthetic sequence is a star shaped object with a known amount of

[2] Closed form solutions for the roots of the polynomials with degree less than 5 is available in most of the mathematical handbooks.

rotation and a translation. The signal to noise ratio is 0 dB. The other sequence is shot from a video camera mounted on the roof of a car, driving on a road. One frame from each of the sequences are shown in figures 13 and 14.

Figure 15 shows the magnitude of the optical flow vectors for the synthetic test sequence. The certainty of the optical flow vectors are shown in figure 16, where gray means total uncertainty, degree of whiteness is proportional to certainty for the case without aperture problem, and degree of blackness is proportional to certainty for the case with aperture problem. Figure 17 shows the original frame with a set of estimated optical flow vectors overlayed. White arrows indicate the case when there is no aperture problem, and black ones the case with aperture problem. For a detailed accuracy evaluation of the optical flow estimates obtained by the linear symmetry concept we refer to [35]. In figure 18 the same is shown for the outdoor sequence.

As comparison, the algorithm described by Horn and Schunk has been used. In particular, for a meaningfull comparison, equation (42) with $\lambda = 0$ and the same interpolation and window functions used in the linear symmetry approach is applied. Consequently, we have solved the resulting linear system of equation (2×2) whenever there was a unique solution (white arrows). In the case when there were an infinite number of solutions we have chosen the one with $\bar{k}$, closest to $(0, 0, 1)^t$ (the black arrows). For the synthetic image sequence, the resulting magnitude of the optical flow vectors are shown in figure 19. Figure 20 shows the original frame with a set of optical flow vectors estimated by the Horn and Schunk algorithm. The estimates are reasonably well in the areas with unique optical flow (white arrows) while they are poor in the areas with no unique flow (black arrows). One reason is that there is no certainty measure, i.e. it is not possible to know to what degree there is an aperture problem. Another is that there is no time integration. We have also tested the iterative version of the process minimizing (42) as it is originally proposed by [30]. The results obtained using various λ's in the range of $[0.1, 500]$ and iterations in the range of $[30, 500]$ were comparable to those illustrated by Figures 19 and 20 and therefore we have omitted them. In this case the SNR in combination with the sparcity of the points having a unique flow estimate disturbed the propagation of the true flow estimates towards the points with ambiguous flow estimates.

The true motion in the traffic sequence was not known. For this reason and also for the fact that the Horn and Schunck's method have a fundemantally different way of handling the aperture problem (we have certainty measures giving a continuous selection between black and white arrows) it is very difficult to give an objective judgement of the performance results in the case of the traffic sequence. For this reason we only include our velocity result in Figure 18 and, after a visual inspection, note that the motion estimate and the choice between the white and black arrows is satisfactory.

6 Discussion

The key idea in the outlined approach is known as the principal axis analysis in the mechanical engineering and statistical literature. The task is to minimize an error function as given in (2). The immediate question is whether this error function is a suitable one since the high frequencies contribute to the error much more than the low frequencies,

due to the quadratic term. Consequently the noise, which usually has high frequency components, would make the error noise sensitive. The discrete image data is assumed to be obtained from a bandlimited continuous image. The experimental results show that this is not a serious problem since a smooth suppression of high frequency terms is assumed by utilizing spatially spread interpolation functions. It can in fact quite easily be shown that the noise amplification in the presented algorithms, due to the choice of $\mathcal{L}^2$ norm, is equivalent to the noise amplification in conventional zero crossings of Laplacian algorithm for edge detection, since Laplacian filtering corresponds to multiplication by $\|\bar{\omega}\|^2$ in the frequency domain. An alternative to the proposed error function would be to use the absolute value of the difference instead of squared difference in (2). But this causes many computational difficulties since minimizing such an error function in the frequency domain would be too complicated, let alone computing it in the spatial domain.

Not only its solutions but also the frequency domain minimization problem itself is possible to interpret in the spatial domain. It can be shown that the obtained estimate of the optimal axis is also the normal of an optimal plane parallel to which a translation of the entire neighbourhood can be performed with minimal error in the $\mathcal{L}^2$ sense.

In Section 4 properties of the complex number field are utilized to show that the eigenvector analysis of a 2×2 matrix is not necessary. The significant result is that averaging upon the complex valued partial derivative image, $(f_x + if_y)^2$, acts as the least square optimization process which is intuitively appealing. In Section 5 one is tempted to use the field of quaternions, [15], to obtain a similar simplicity for the 3–D case. The analogy to equation (24) provides us:

$$\bar{x}_1^t \bar{x}_2 = -\frac{1}{2}(\mathbf{x_1 x_2} + \mathbf{x_2 x_1}) \tag{45}$$

where $\mathbf{x_1}$ is the quaternion with the scalar part (principal unit) having the value zero and the other units corresponding to i, j, k having the values obtained from the first, second and third components of the three–dimensional vector $\bar{x}_1$. But since the field of quaternions, in contrast to the field of complex numbers, is not commutative the analogy resulting in simplifications cannot be followed, (25).

The results from Galois theory, [24], tell us that the polynomials with degrees less than 5 can be solved for their roots algebraically. That is, the roots are possible to obtain by utilizing the arithmetic operations of squares, square roots, cubics, cubics roots ..etc., a finite number of times. Thus fast, optimal orientation detection in the least squares sense with corresponding certainties can be achieved for such data intensive domains as volume images, image sequences and volume sequences. The time complexity of the presented algorithms depends on the available hardware architecture and the implementation technique of convolutions, (FIR, IIR, separable).

7 Conclusion

The presented experiments show that frequency domain based minimization of the error function or the moment of inertia gives good estimates of the local orientation in n–D together with corresponding confidence measures. The presented algorithms, which

are entirely computed in the spatial domain, implicitly perform the least square fitting through averaging.

In the 2–D case it is shown that the eigenvector computation is equivalent to averaging over the square of the complex gradient. The optical flow problem is formulated as a pattern recognition problem (recognition of a line or a plane in the frequency domain) and for the detection, the linear symmetry method is proposed. The closed form solutions of the resulting eigenvalue problem and the separability of the filters contribute to a decrease in the execution time compared to the iterative methods and methods using non-separable and/or numerous filters.

Both in 2-D and 3-D problems the algorithms were implemented on a machine using integer arithmetic and the intermediate results were stored as integers. Despite the known difficulties attached to it, the integer aritmetic is widely used in parallel machine architectures because of technical and economical reasons. We have shown that the approach taken here to solve the orientation problem, is robust enough to be implemented on machines using integer arithmetics.

Acknowledgements

Credit should be given to to the Computer Vision Laboratory staff of Linköping university, especially to Dr. Hans Knutsson, for constructive discussions; to the Swedish National Board for Technical Development and Thomson CSF for their financial support and finally to the referees who had valuable suggestions on the organisation and contents of the experimental sections.

References

[1] E. H. Adelson, J. R. Bergen *Spatiotemporal energy models for the perception of motion* Jour. of the opt. soc. of America A, 1(2) 1985 pp 284-299

[2] E. H. Adelson and J. A. Movshon *Phenomenal coherence of moving gratings* Nature 200, 523-525 (1982)

[3] P. Anandan *Computing dense displacement fields with confidence measures in scenes containing occlusion* SPIE vol. 521 Intelligent robots and computer vision (1984).

[4] O. Axelsson and V. A. Barker *Finite element solution of boundary value problems: theory and computation* Orlando, Academic., (1984)

[5] R. Bajcsy and L.I. Lieberman *Texture gradient as a depth cue.*Computer graphics and image processing, vol. 5, pp.52-67 1976

[6] J. Bigün C. Horne *Feature space dimensionality reduction by using local representability* Int. conf. on signal processing, Beijing, Oct. 22-26 1990.

[7] J. Bigün, G.H. Granlund *Optimal orientation detection of linear symmetry*, First international conf. on computer vision, London, June 1987. pp. 433-43

[8] J Bigün. *Recognition of local symmetries in gray value images by harmonic functions*, IAPR, 9'th international conference on Pattern Recognition, Rome, November 1988.

[9] J. Bigün *A structure feature for image processing applications based on spiral functions* Computer vision, graphics and image processing. No. 51, pp. 166-194 (1990)

[10] C. Blakemore and F. W. Campbell *On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images* J. Physiology (1969), 203 pp. 237-260

[11] P. Brodatz *Textures* Dover publications, Inc. New York, 1966

[12] P. Burt: *Fast filter transforms for image processing* Computer graphics and image processing 16, pp. 20-51 (1981)

[13] P. Burt: *Smart sensing within a pyramid vision machine* IEEE proceedings vol. 76, No. 8 1988 pp. 1006-1015

[14] F. W. C. Campbell and J. Robson *Application of Fourier analysis to the visibility of gratings* J. physiol., London, 197, pp. 551-566 1968.

[15] R. Courant, D. Hilbert, *Methods of Mathematical Physics* Interscience Publishers, Inc., New York, 1953.

[16] J.L. Crowley and R.M. Stern *Fast computation of the difference of low-pass transform* PAMI-6, 212-222 (1984)

[17] P.E. Danielsson: *Rotation invariant linear operators with directional response.* Proceedings of 5'th international conference on pattern recognition, December 1980.

[18] J. Daugman *Six Formal Properties of Two-Dimensional Anisotropic Visual Filters: Structural Principles and Frequency / Orientation Selectivity* IEEE trans. on Systems, Man and Cybernetics vol. 13 1983 pp. 882–887

[19] D. E. Dudgeon, R. M. Mersereau *Multi dimensional digital signal processing* Englewood Cliffs, N.J. Prentice Hall, (1981)

[20] O. D. Faugeras, N. Ayache and Z. Zhang *A preliminary investigation of the problem of determining ego and object motions from stereo* 9th ICPR, pp. 242-246, 1988, Rome.

[21] D. J. Fleet and A. D. Jepson *Hierarchical construction of orientation and velocity selective filters* PAMI vol. 11 No. 3 1989 pp. 315-325

[22] D. Gabor *Theory of communication* J. Inst., Elec. Eng. vol. 93, pp. 429-459, 1946.

[23] S. Geman and D. Geman *Stochastic relaxation, Gibbs distributions and the Bayesian restauration of images* IEEE trans. Pattern analysis and machine intelligence, no. 6, pp. 721-741, nov. 1984.

[24] L. J. Goldstein *Abstract Algebra* Prentice–Hall, Inc., Englewood Cliffs, New Jersey, 1973.

[25] G.H. Granlund *In Search of a General Picture Processing Operator*, Computer Graphics and Image Processing 8, 155–173 (1978).

[26] A. Grosssman and J. Morlet Decomposition of Hardy functions into square integrable wavelets of constant shape SIAM j. math., vol. 15, pp. 723-736, (1984)

[27] O. Hansen and J. Bigün *Detection of local symmetries in multi-dimensional images* 5th Int. conf. on image analysis and processing. 20-22 september 1989, positano, Italy

[28] R. M. Haralick and I. Dinstein *Textural features for image classification* IEEE trans. syst. man and cybern. vol 3 pp. 610-621 nov. 1973

[29] D. J. Heeger *Optical Flow from Spatiotemporal Filters*, First international conf. on computer vision, London, June 1987. pp. 181-190

[30] B. K. P. Horn and B. G. Schunck *Determining optical flow* Artificial intelligence 17 pp. 185-203 (1981)

[31] M. K. Hu, *Visual pattern recognition by moment invariants* IRE Tr. on information Theory, pp. 179- 187 1962

[32] D.H. Hubel and T.N. Wiesel *Receptive fields of single neurones in the cat's striate cortex* J. physiology (London), No. 148, pp. 574-591 (1959)

[33] D. H. Hubel, T. N. Wiesel *Brain Mechanisms of Vision* Scientific American, sep. 79 Vol.241 No.3 pp. 150-162

[34] R. A. Hummel: *Feature Detection Using Basis Functions.* Computer Graphics and Image Processing 9, 1979.

[35] B. Jähne *Motion determination in space-time images* SPIE VOL. 1135 image processing III PP. 147-152 (1989)

[36] H. Knutsson *Filtering and reconstruction in image processing*, Linköping studies in science and technology Dissertations No:88, 1982.

[37] K. I. Laws *Textured image segmentation* Ph. D. dissertation, Dept. of Elec. Eng., Univ. Southern California, 1980

[38] S. Marcelja *Mathematical description of the responses of simple cortical cells* J. Opt. soc. am. 70, pp. 1297-1300 (1980)

[39] D. Marr *Vision* W. H. Freeman Inc., San Fransisco, 1982.

[40] I.L. Meriam: *Statics.* John Wiley & Sons, New York, 1980.

[41] J. A. Movshon, I. D. Thompson and D. J. Tolhurst *Spatial and temporal contrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex* J. Physiol. (London) 283, pp 101-120 (1978).

[42] H. H. Nagel *Image sequences- (Ten (octal) years - From phenomenology towards a theoretical foundation* In proc. of Int conf. on patt. recognition (IAPR), oct. 27-31, 1986, paris, pp. 1174-1185.

[43] H. H. Nagel *On the estimation of optical flow: relations between different and some new results* Artificial intelligence 33 (1987) pp. 299-324

[44] A. Papoulis *Signal Analysis* Mc. Gra., 1977.

[45] M. Porat and Y. Y. Zeevi *Localized texture processing in vision: analysis and synthesis in the Gaborian space* IEEE Bio Medical Eng. BME-36 pp. 115-129 (1989)

[46] W. Reichardt *Autocorrelation, a principle for the evaluation of sensory information by the central nervous system* In Sensory communication, W. A. Rosenblith, ed. MIT press, Cambridge, Mass., pp. 303-317 (1961).

[47] B. Sakitt and H. B. Barlow *A model for the economical encoding of the visual image in cerebral cortex* Biol. Cybern. 43, pp. 97-108 (1982).

[48] S. K. Sethi and R. Jain *Finding trajectories of feature points in a monocular image sequence* PAMI-9 1, PP. 56-73 (1987)

[49] M. Spann, R. Wilson *A quad-tree approach to image segmentation which combines statistical and spatial information* Pattern Recognition Vol. 18 Nos 3/4, pp 257-269 1989

[50] R. Szelski *Fast surface interpolation using hierarchical basis functions* PAMI 6 pp. 513-528 (1990).

[51] S. Tanimoto and T. Pavlidis *A hierarchical data structure for picture processing* Computer graphics image processing. 4, 104-119 (1975)

[52] D. H. Tolhurst and J. A. Movshon *Spatial and temporal contrast sensitivity of striate cortical neurons* Nature 257, pp. 674-675 (1975)

[53] S. Ullman *The interpretation of visual motion* MIT U. Press, Cambridge, Mass., (1979)

[54] M. Unser *Local linear transforms for texture measurements* Signal Processing 11 pp. 61-79 (1986)

[55] M. Unser and M. Eden *Multiresolution Feature Extraction and Selection for Texture Feature Segmentation* IEEE Trans on Pattern Analysis and Machine Intelligence, 11: pp.717-727, July 1989.

[56] A. B. Watson and A. J. Ahumada, Jr. *Model of human visual-motion sensing* Jour. of the opt. soc. of America A, 1(2) 1985 pp 322-342

[57] A. B. Watson *Detection and recognition of simple spatial forms* in Physical and Biological Processing of images, A.C. Slade, ed. Springer-Verlag, Berlin (1983)

[58] J. Weng, T.S. Huang and N. Ahuja *3-D motion estimation, understanding and prediction from noisy image sequences* IEEE PAMI-9, No. 3, pp. 370-389 (1987)

[59] E. Wigner *On the quantum correction for thermodynamic equilibrium* Phys. Rev. 40 pp. 749-759, (1932)

[60] A. Wouk: *A course of applied functional analysis.* Wiley, New York, (1979).

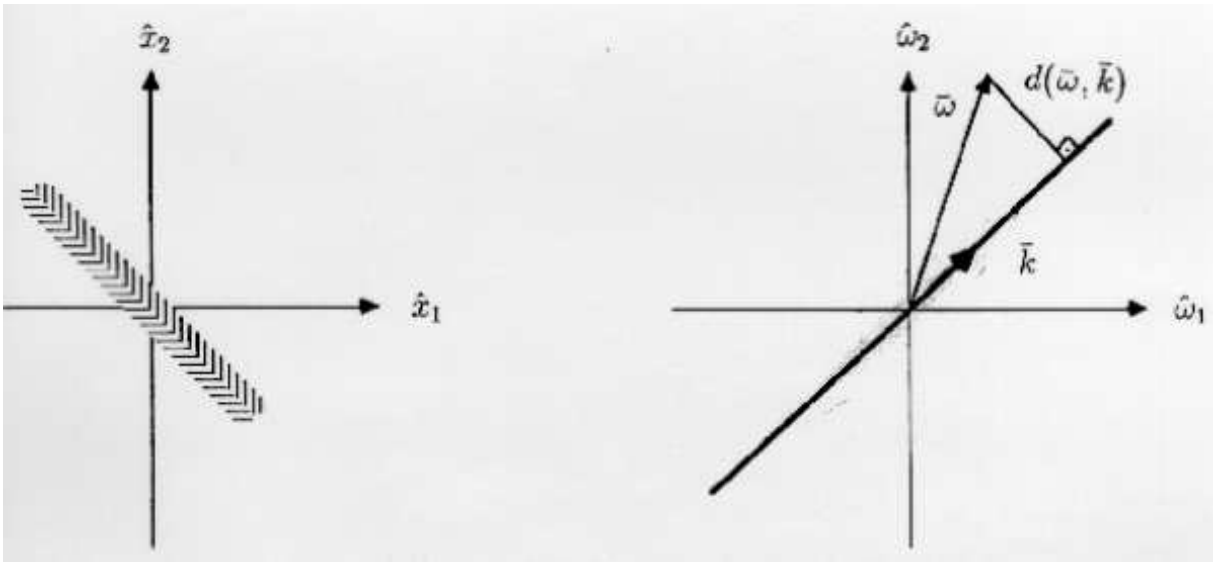

Figure 1: The figure on the left is a local image with a dominant orientation. The figure on the right illustrates the magnitude of the frequency domain which is concentrated to a line denoted by $\bar{k}$. $d(\bar{\omega}, \bar{k})$ is the distance between a point $\bar{\omega}$ and the line $\bar{k}$.

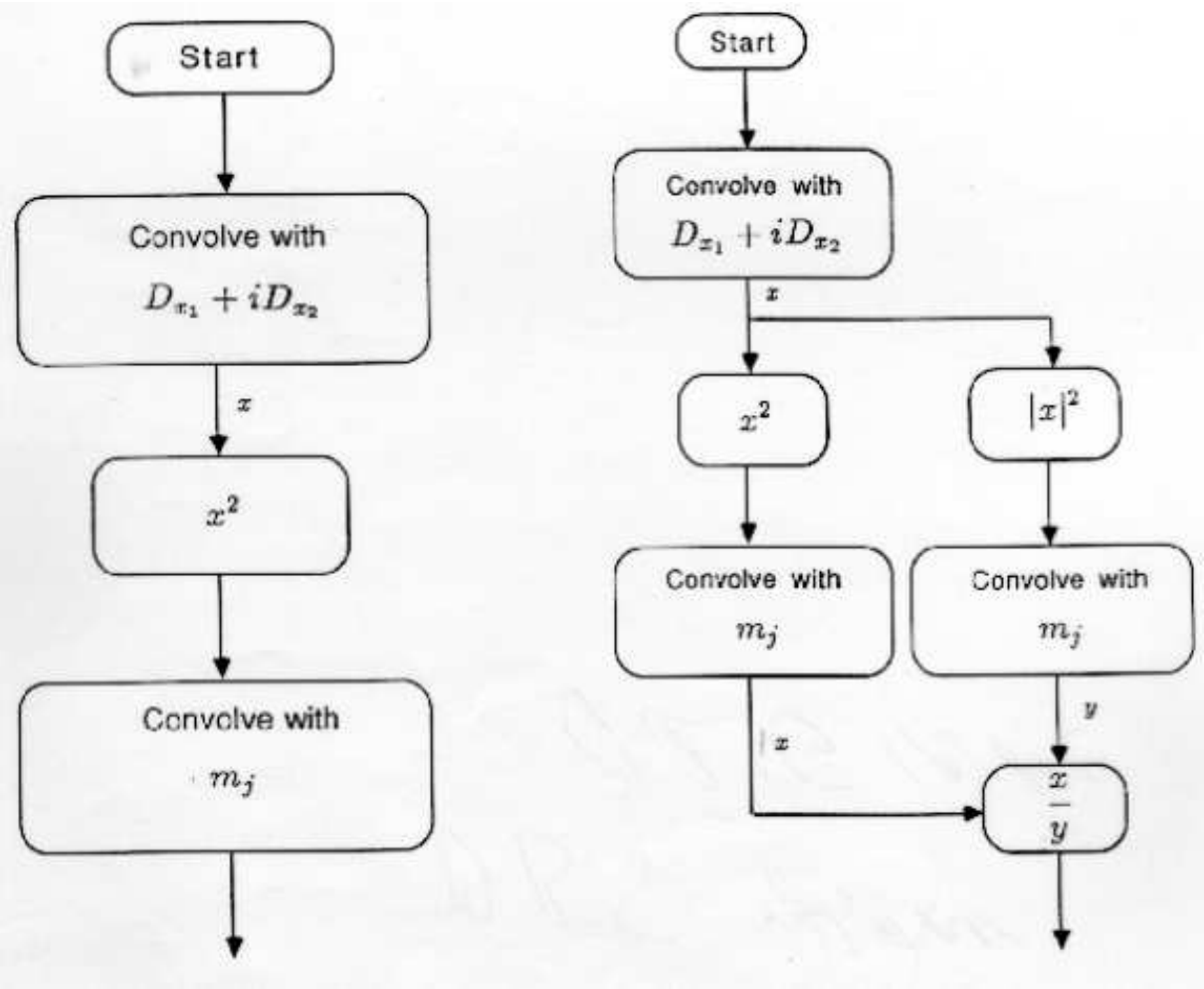

Figure 2: The flowcharts of the algorithms computing energy dependent (left) and energy independent (right) certainties together with optimal orientation estimate of linear symmetry parameters. The resulting images are complex valued with magnitudes of the pixels being certainties and arguments being orientations.

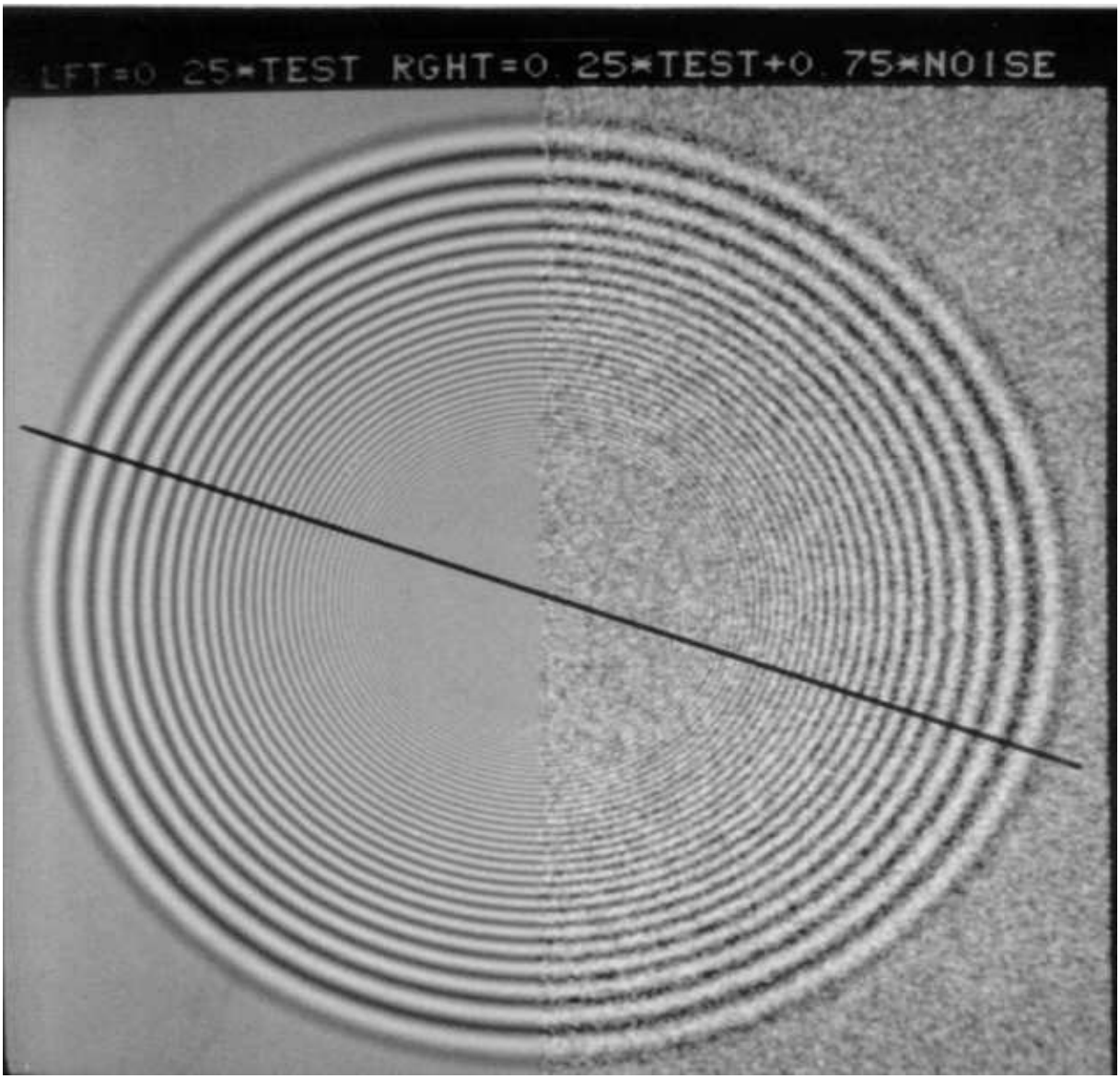

Figure 3: The test image used in the experiments. The straight line is the line along which the signals of results are presented in the subsequent figures.

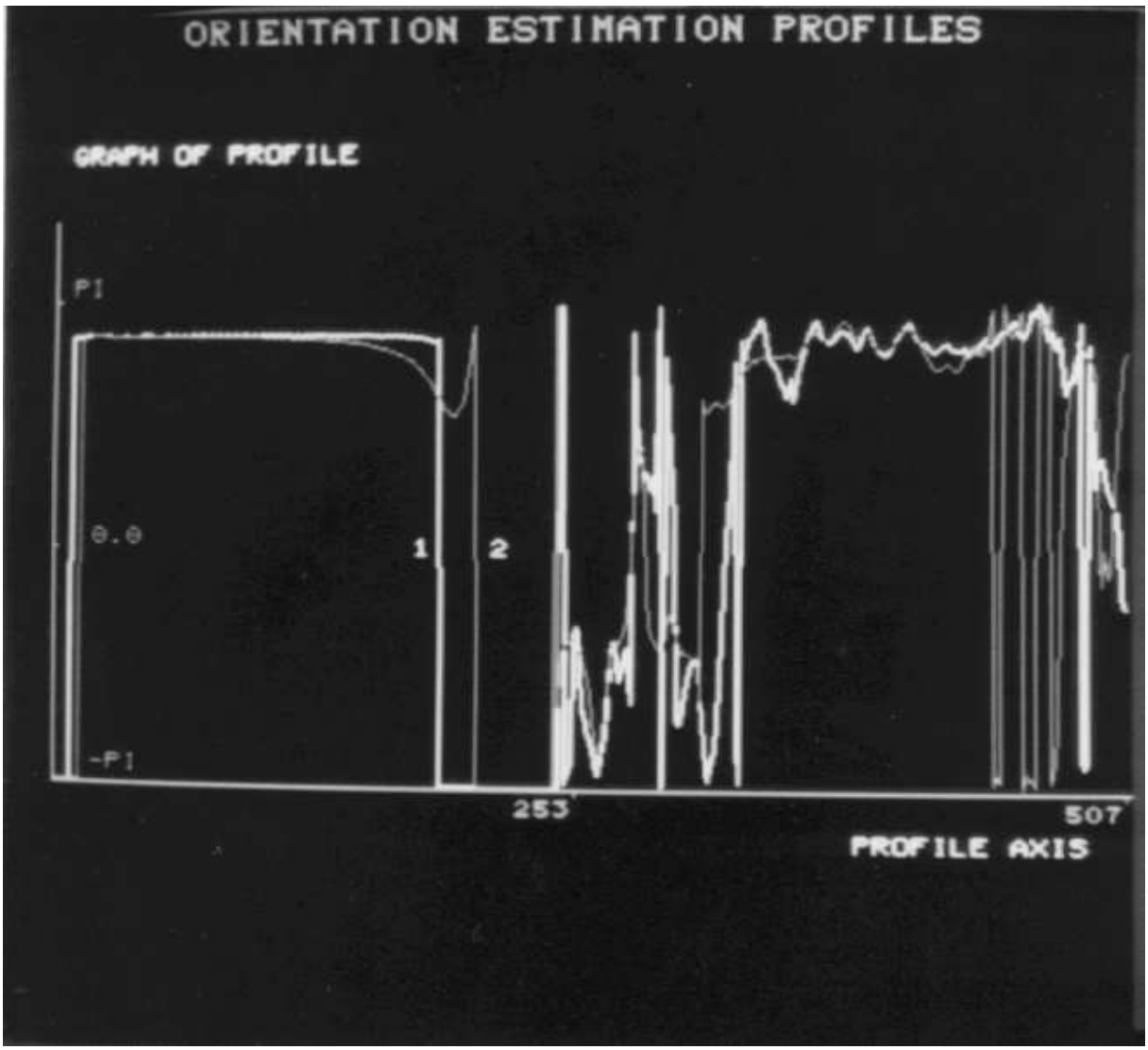

Figure 4: Orientation estimation with two different filter configurations. Graph 1 illustrates 9×9 and 15×15 configuration at the two steps of the algorithm, while 2 is due to 5×5 and 21×21 configuration.

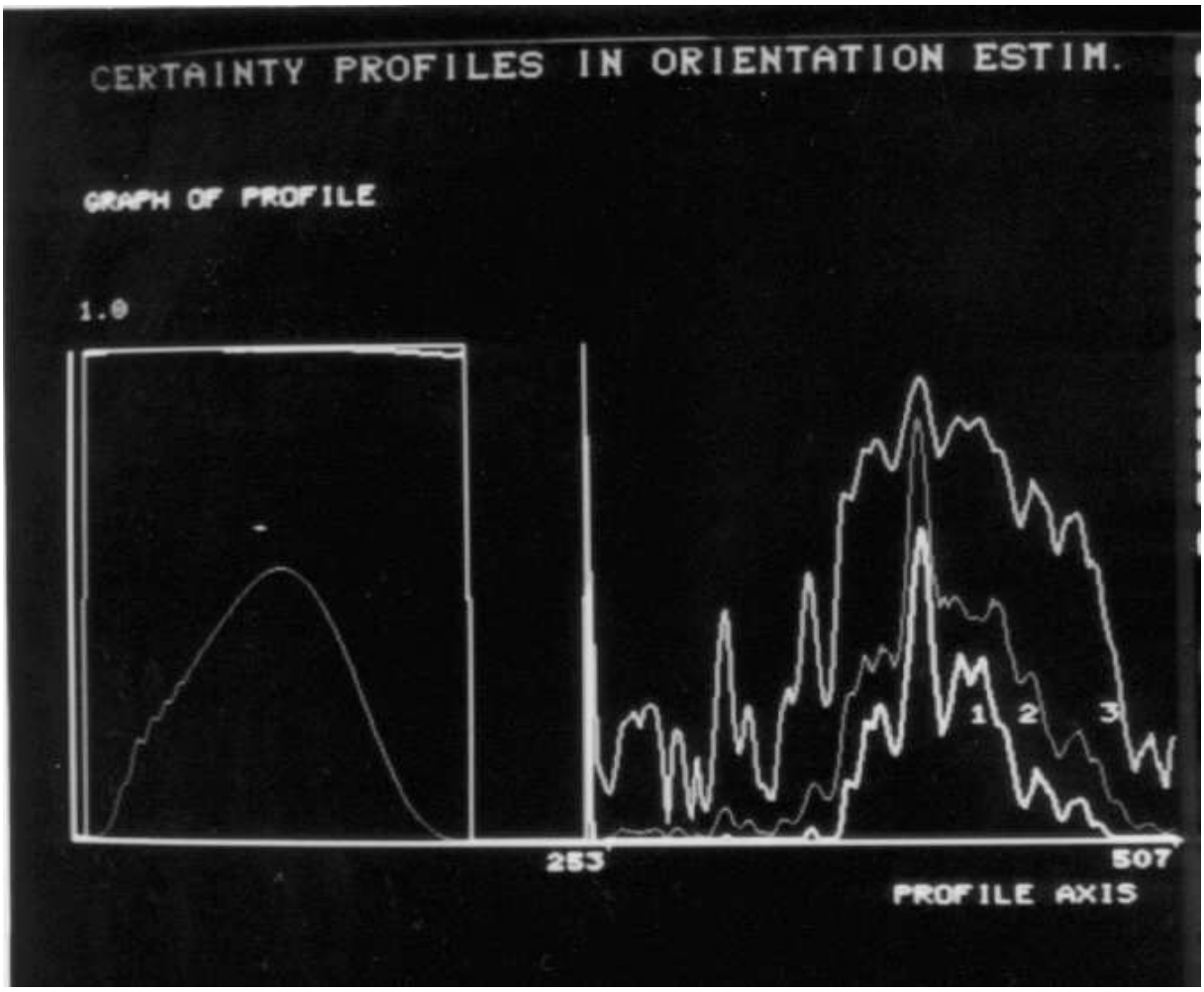

Figure 5: Certainty measures for the signal 1 in Figure 4. Graphs 1 and 3 represent C_{f2} with $c = 6$ and $c = 1$ respectively, while 2 represents C_{f1}.

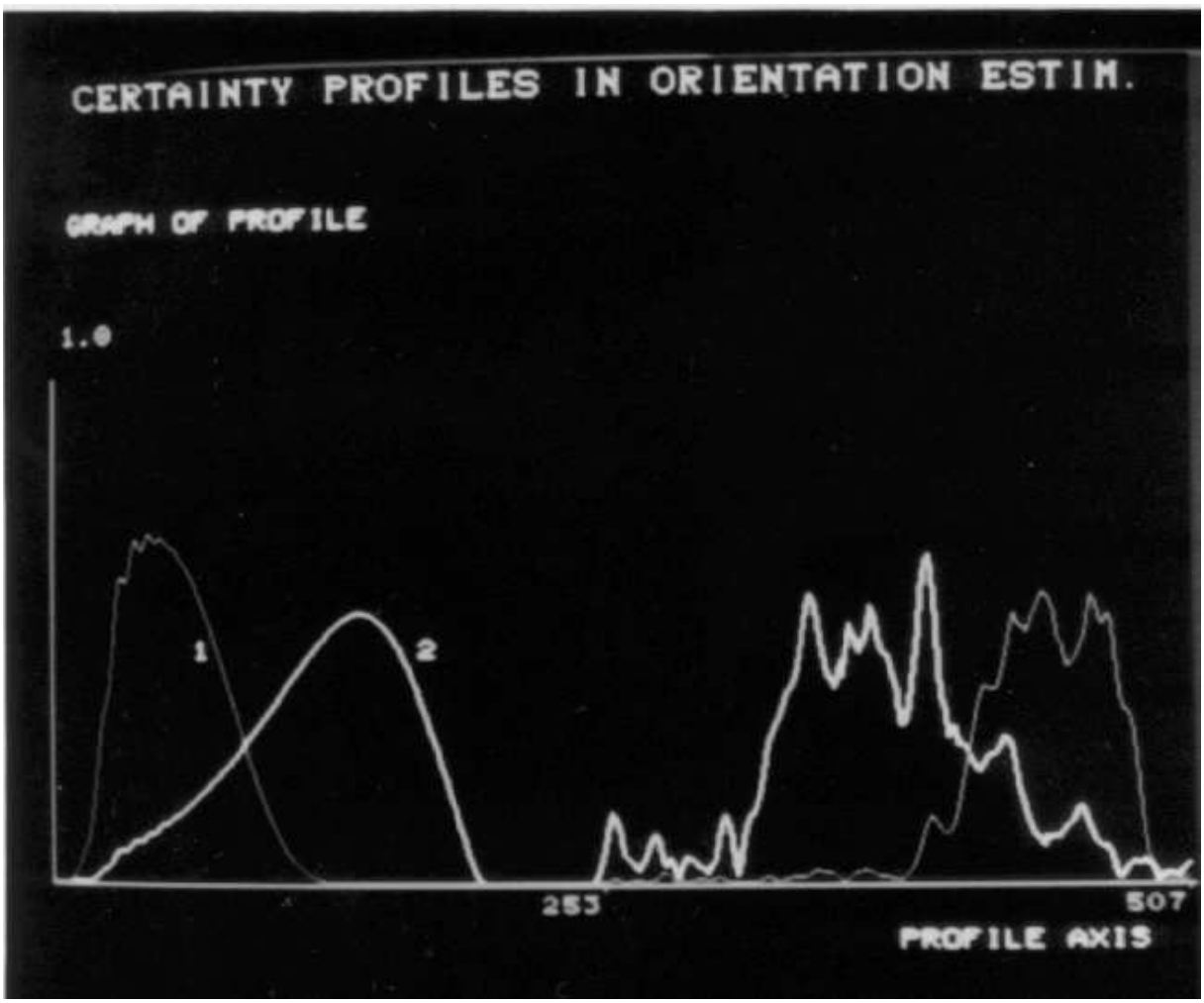

Figure 6: Frequency dependence of C_{f1}. Graph 1 corresponds to 15×15 and 21×21 configuration at the two steps of the algorithm. 2 corresponds to 5×5 and 19×19 configuration.

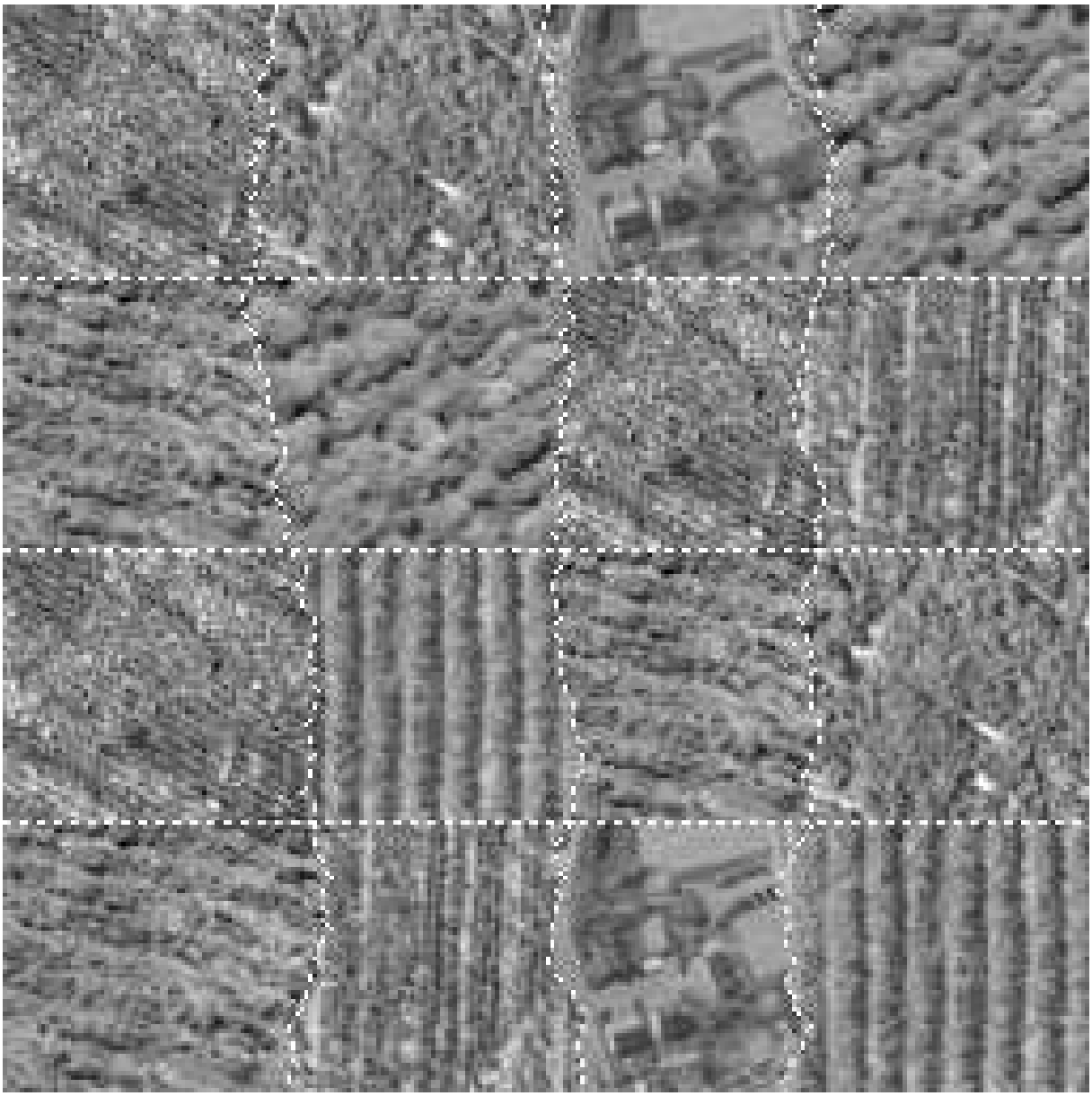

Figure 7: The used aerial texture image containing 7, different textures corresponding to fields, forests and a residential area.

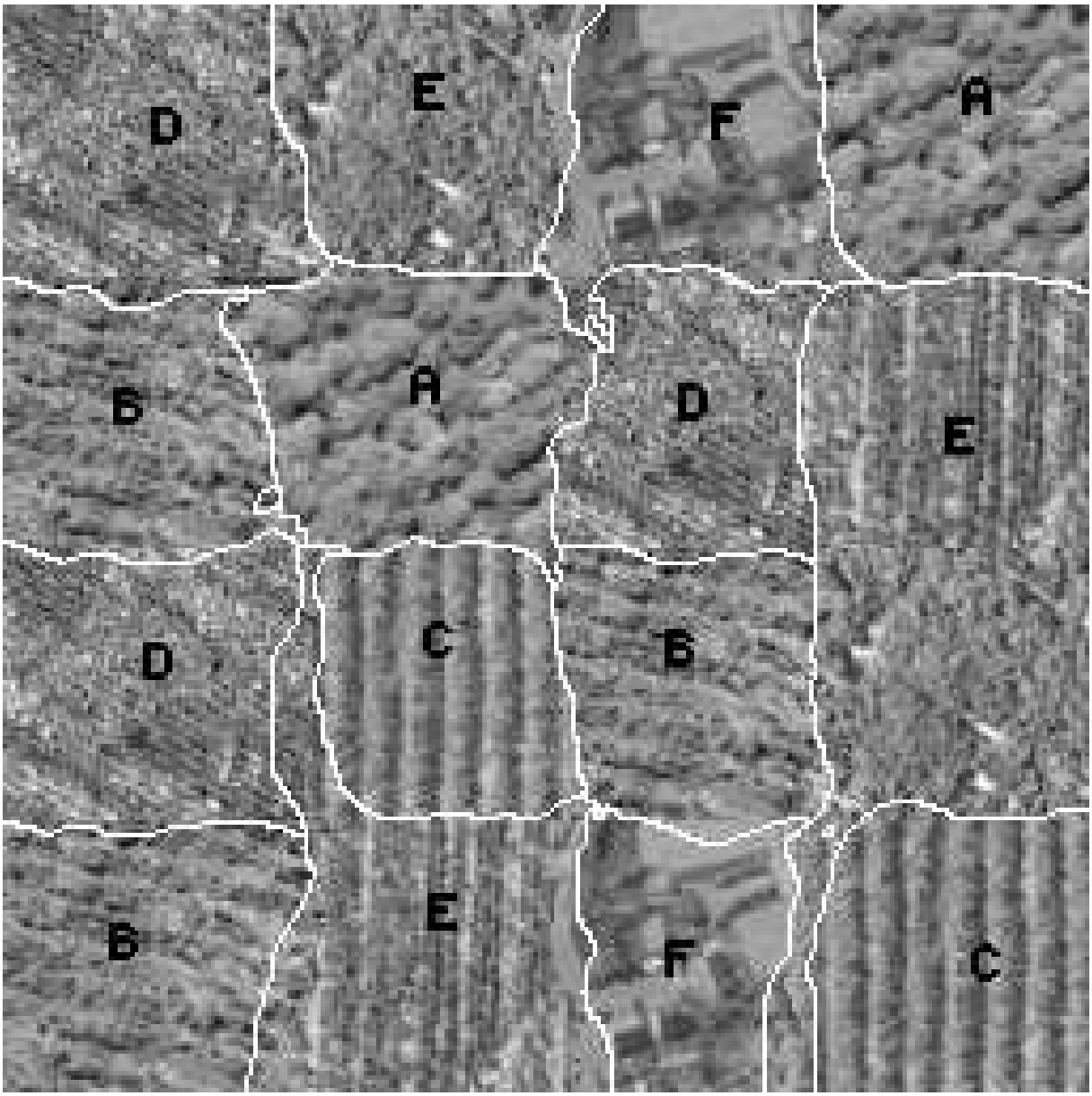

Figure 8: The result of unsupervised feature selection and unsupervised segmentation using the linear symmetry approach.

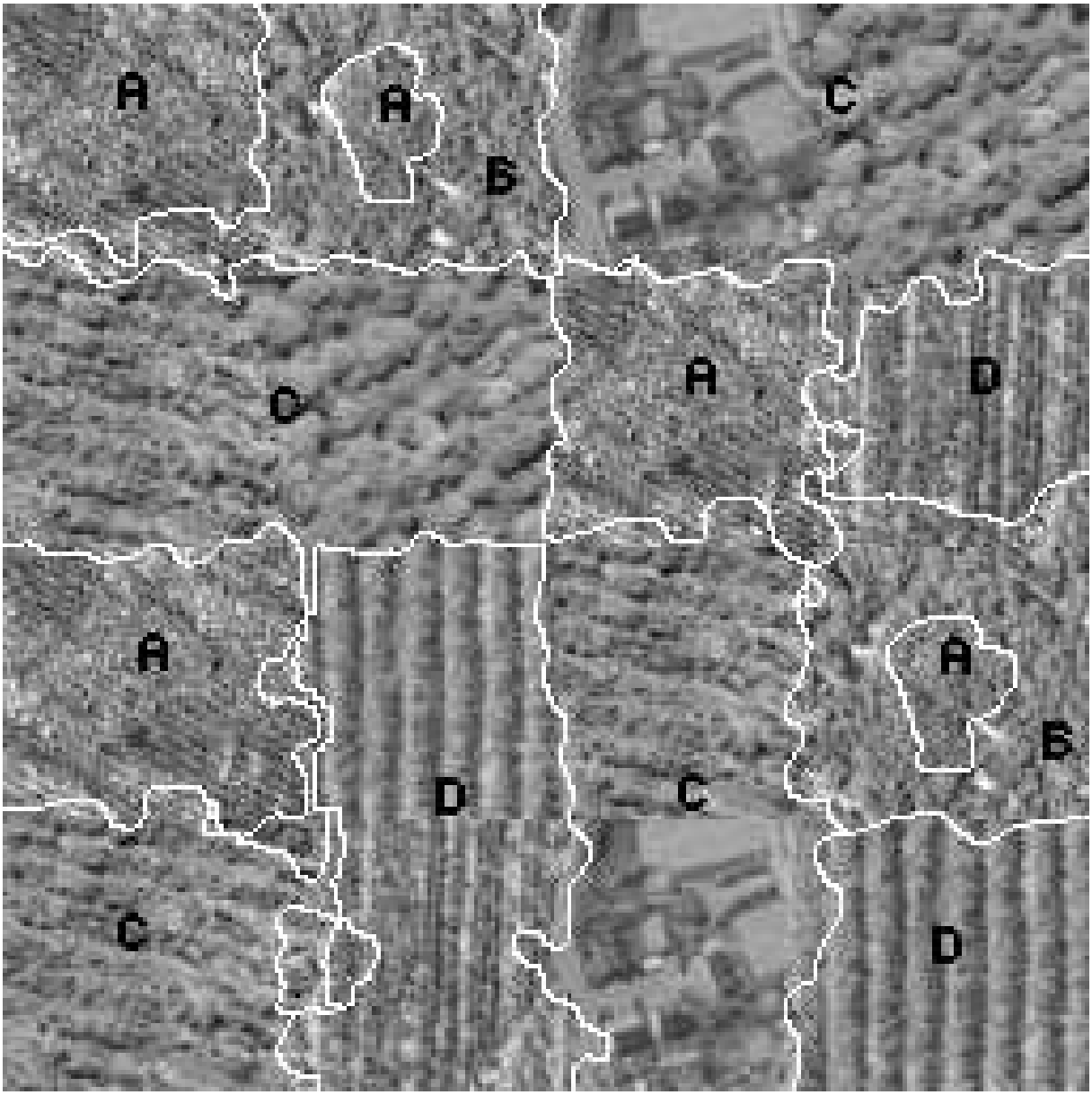

Figure 9: The segmentation result using Laws features.

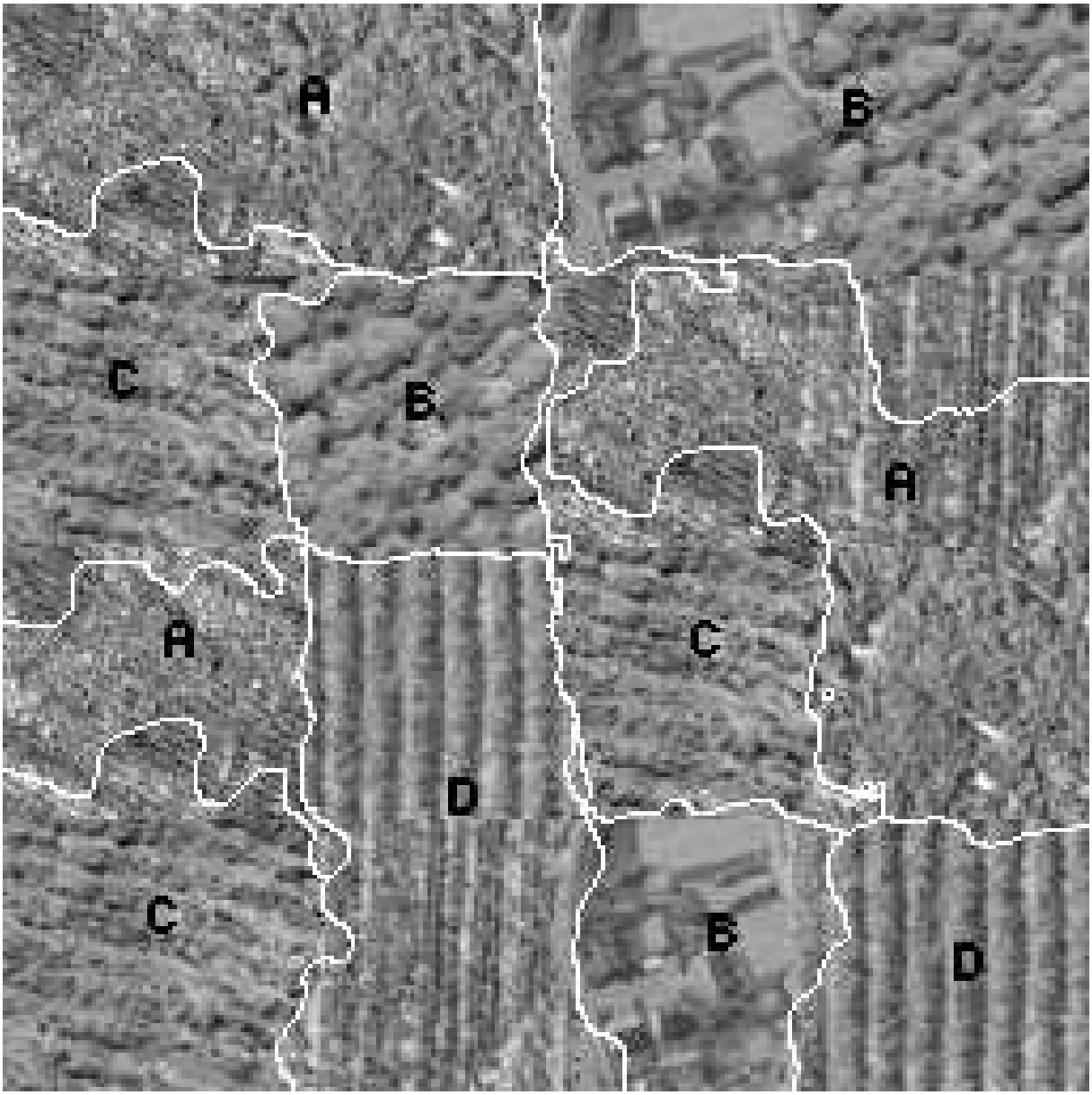

Figure 10: The segmentation result using Unser features.

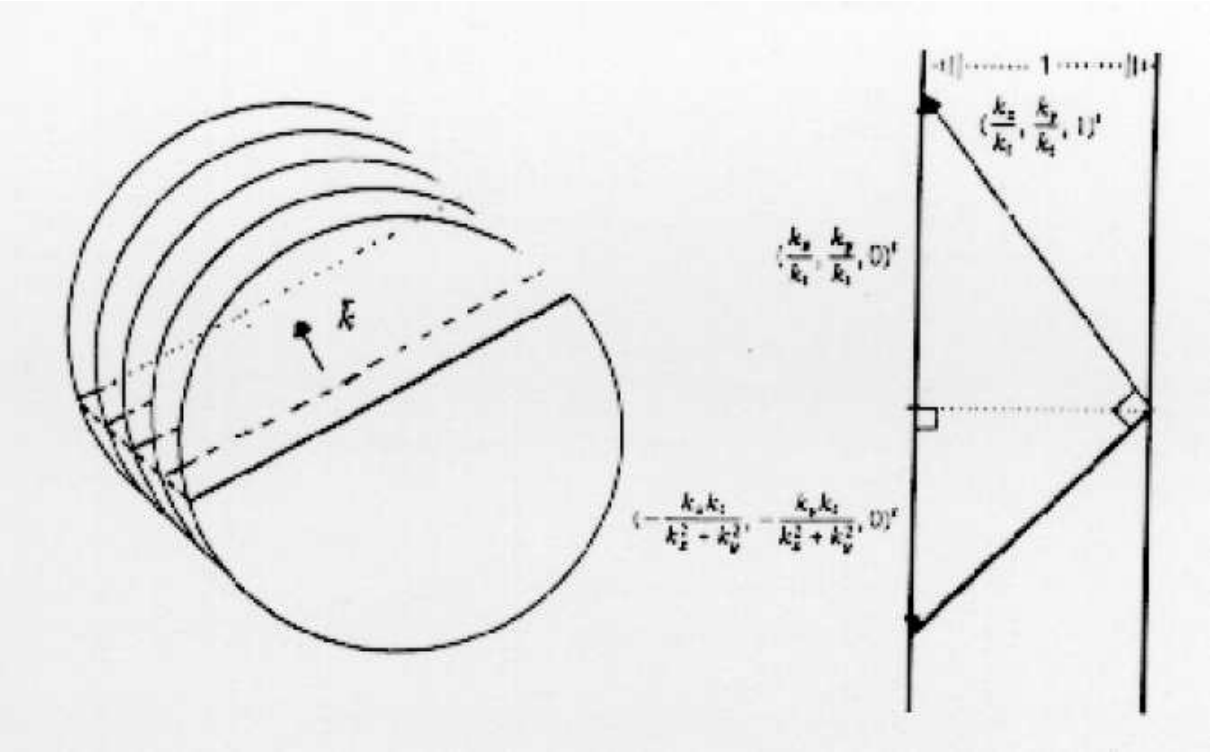

Figure 11: The figure a) illustrates a small region of the visual field in which a translation of a line occurs b) illustrates the geometry used to derive the 2D velocity vector from the 3D normal vector associated with the translation.

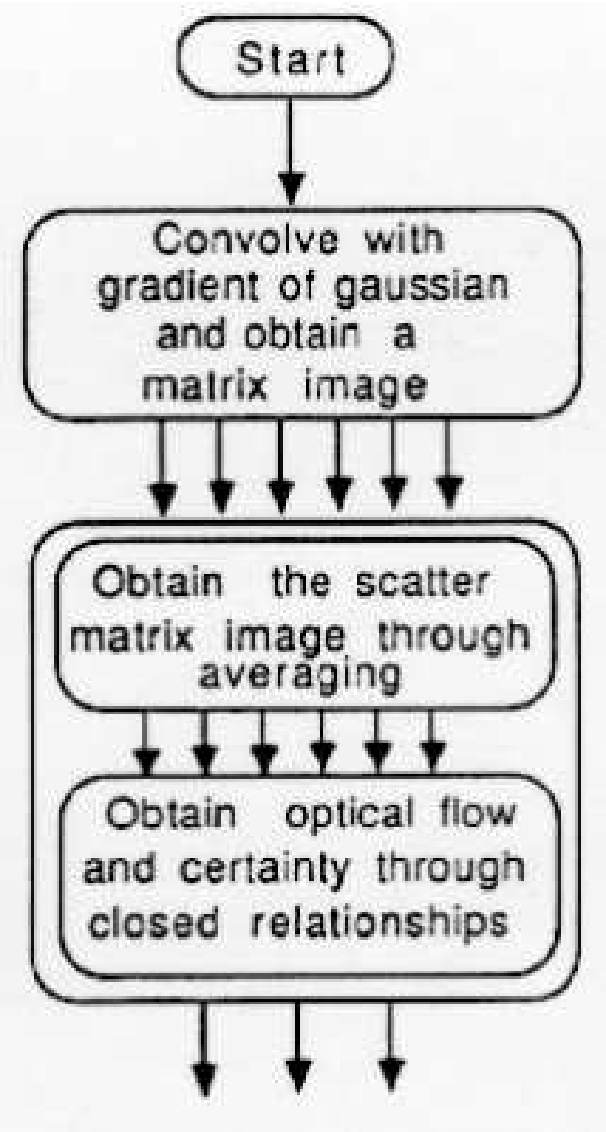

Figure 12: The flow chart illustrates the implemented algorithm.

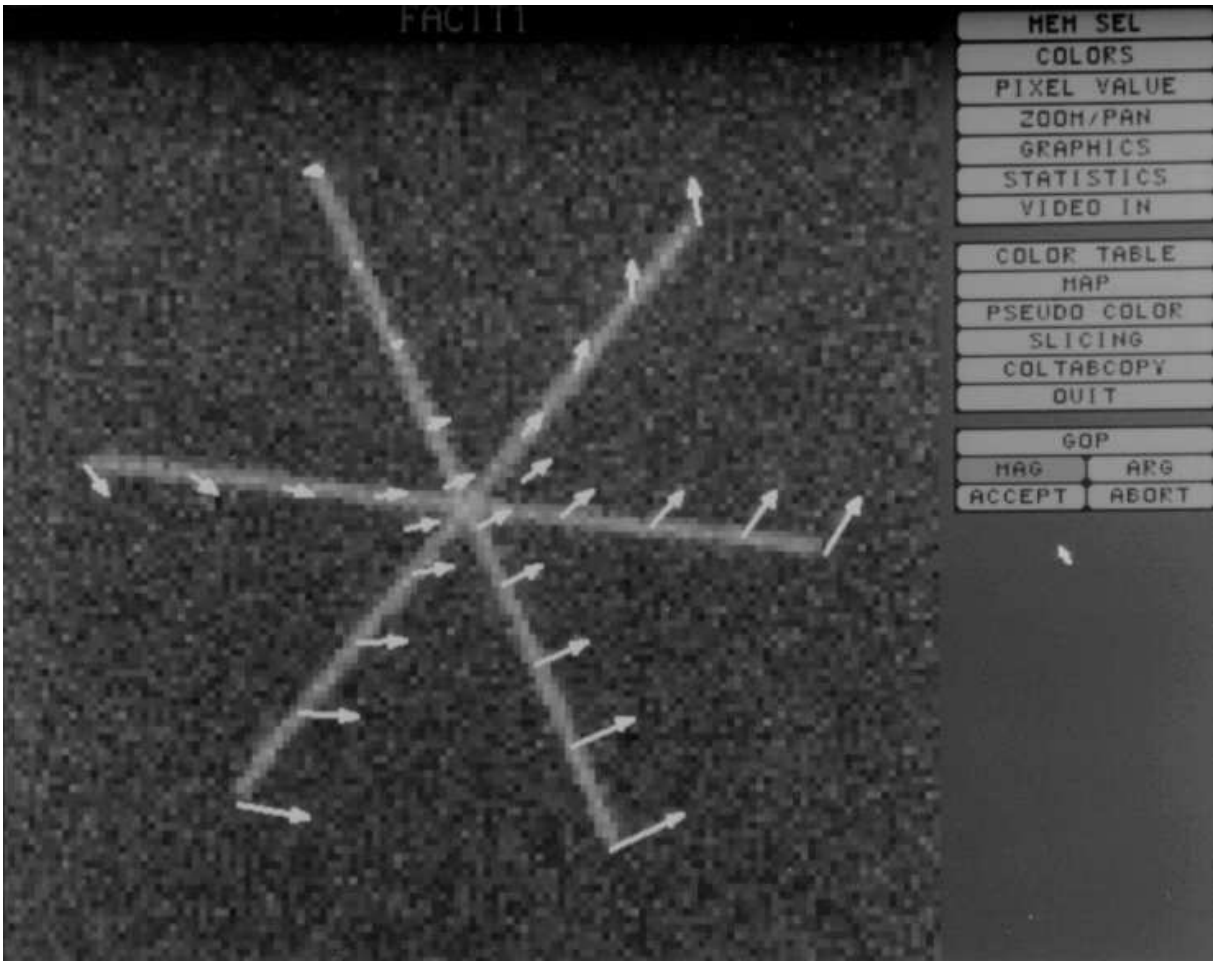

Figure 13: One frame from a synthetic test sequence. The star shaped object rotates 1.8^o counter clockwise round its center, and translates 0.5 pixels up, 1 pixel to the right between each frame. The vectors illustrate the true velocities. The frame size is 128×128.

Figure 14: One frame from a video sequence shot from a car driving on a road.

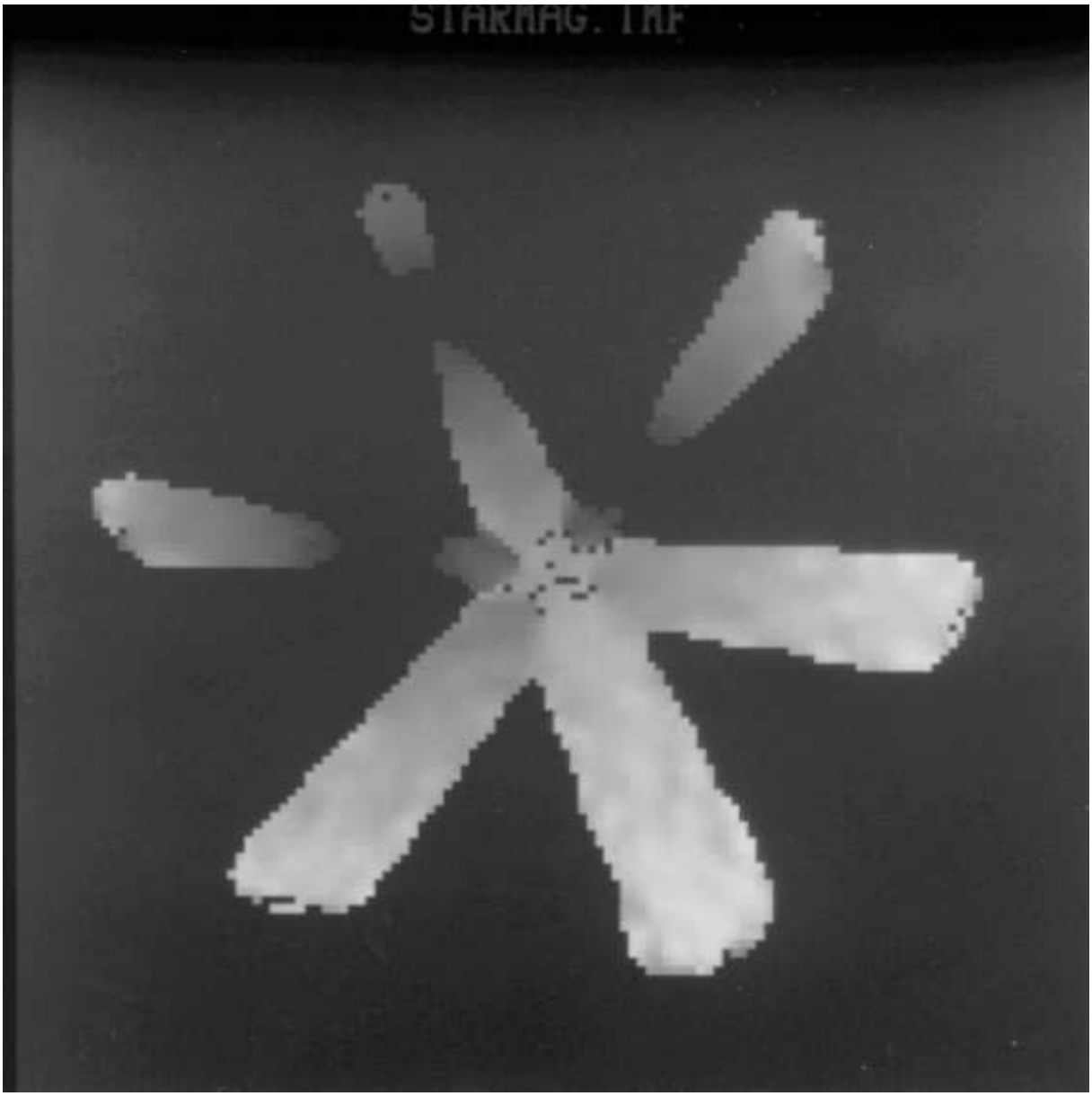

Figure 15: The graylevel corresponds to the magnitude of the estimated optical flow vectors resulting from the proposed algorithm.

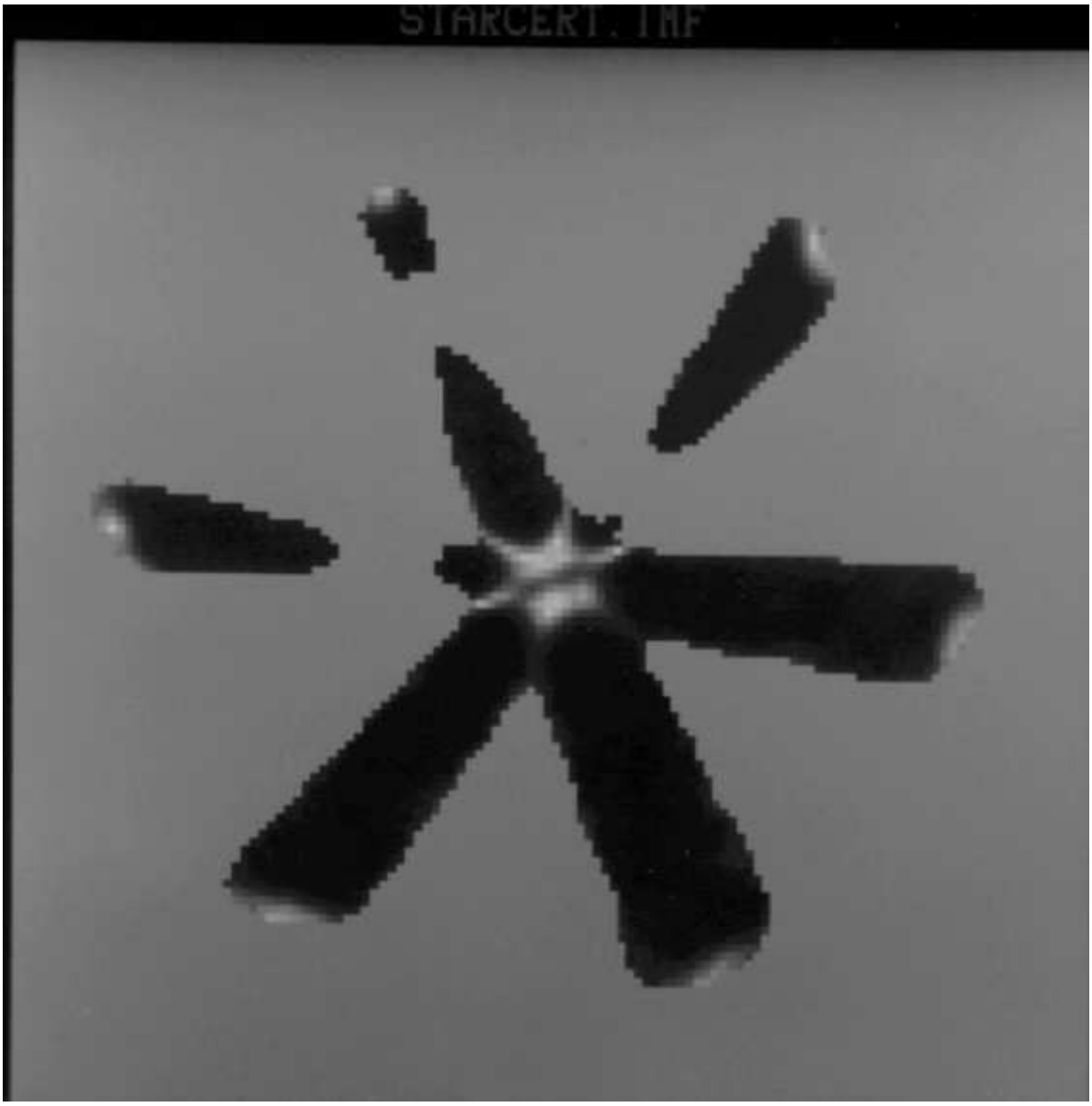

Figure 16: The certainty measures obtained from the proposed algorithm. Gray means total uncertainty, whiteness indicates the degree of certainty for the moving point case, while blackness is proportional to the degree of certainty for the moving line case.

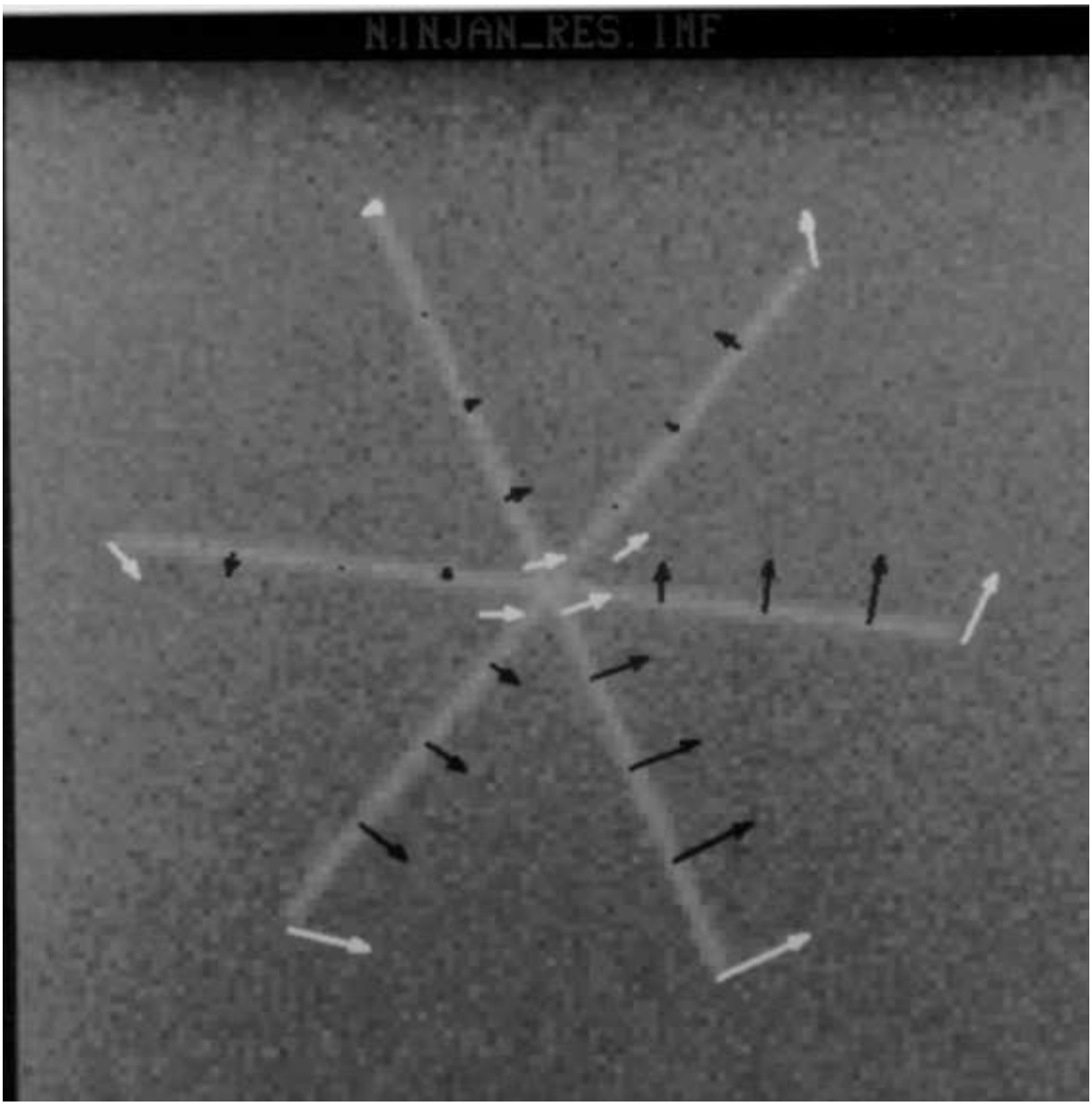

Figure 17: In this figure the original frame is overlayed with arrows indicating the optical flow vectors at some sampled positions. White arrows means that there is no aperture problem, while black arrows means that the optical flow vector is perpendicular to the moving edge.

Figure 18: Optical flow vectors overlayed on a frame from the natural sequence.

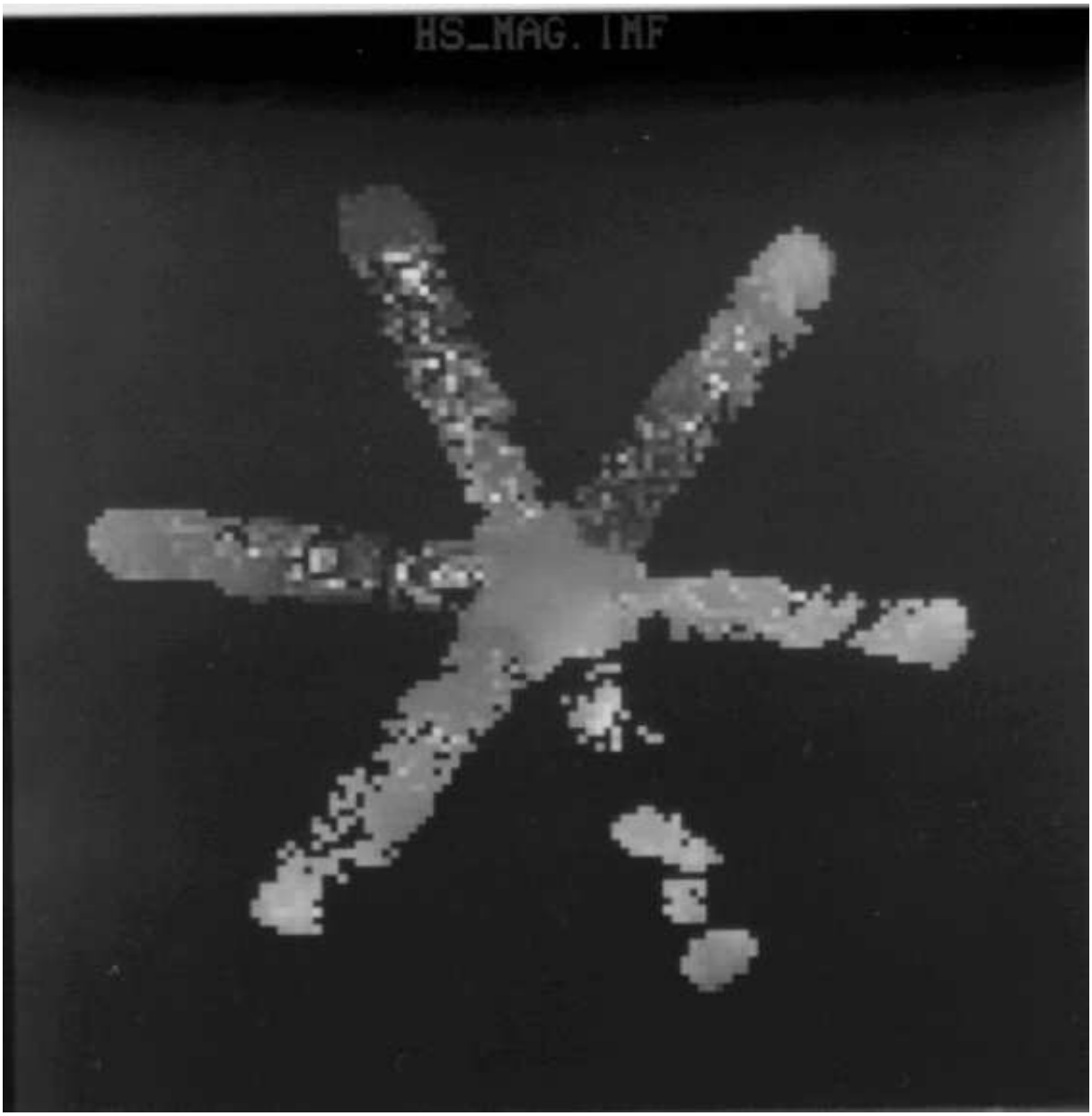

Figure 19: The magnitude of the optical flow vectors resulting from the implemented Horn Schunk algorithm.

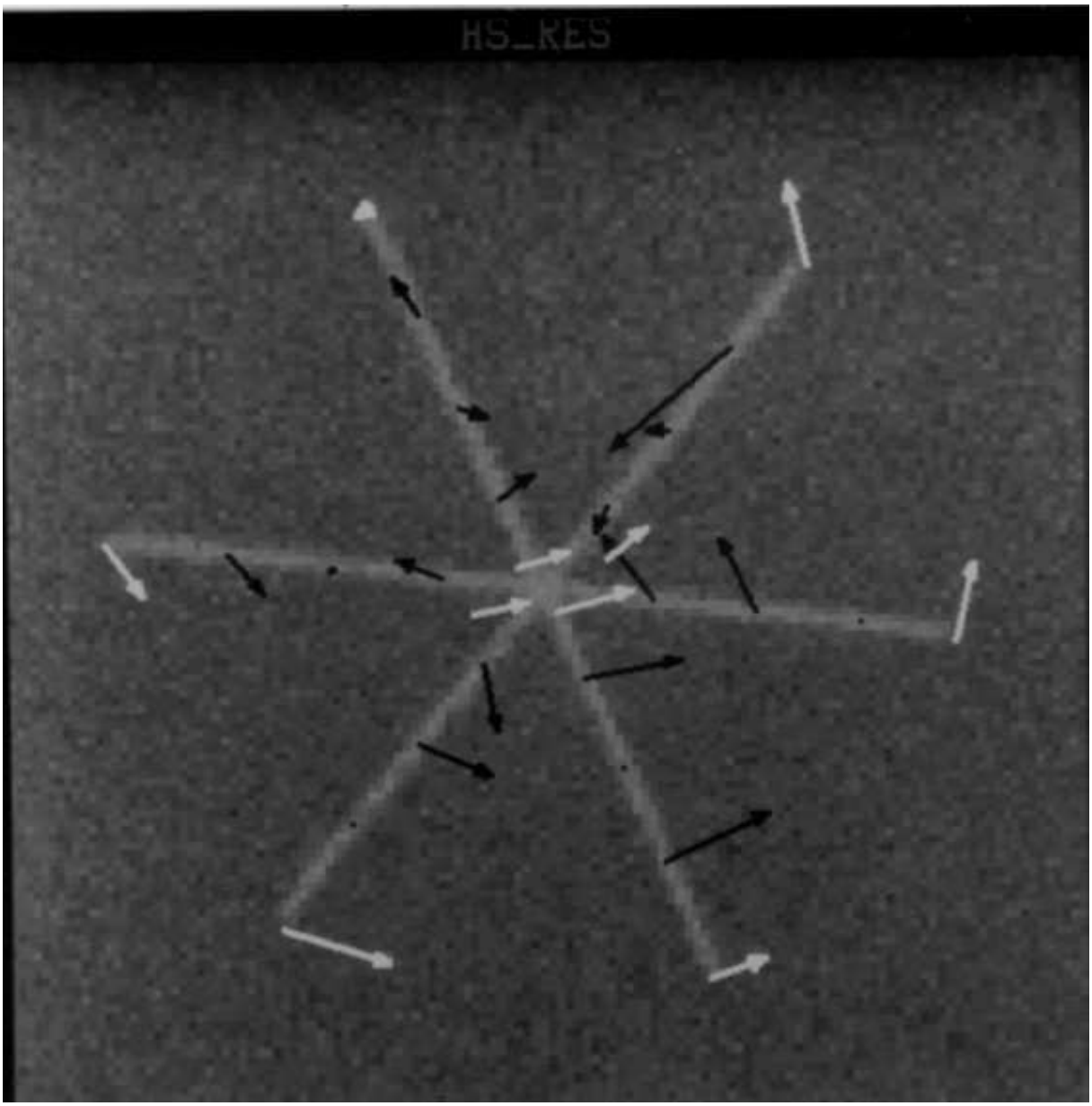

Figure 20: The arrows indicate the optical flow vectors for the Horn Schunk algorithm at the same positions as in the corresponding figure for the proposed algorithm.