A FRAMEWORK FOR ESTIMATION OF ORIENTATION AND VELOCITY

Klas Nordberg and Gunnar Farnebäck

Computer Vision Laboratory, Department of Electrical Engineering, Linköping University

ABSTRACT

The paper makes a short presentation of three existing methods for estimation of orientation tensors, the so-called structure tensor, quadrature filter based techniques, and techniques based on approximating a local polynomial model. All three methods can be used for estimating an orientation tensor which in the 3D case can be used for motion estimation. The methods are based on rather different approaches in terms of the underlying signal models. However, they produce more or less similar results which indicates that there should be a common framework for estimation of the tensors. Such a framework is proposed, in terms of a second order mapping from signal to tensor with additional conditions on the mapping. It it also shown that the three methods in principle fall into this framework.

1. INTRODUCTION

The optic-flow equation

$$(\nabla g)^T \tilde{\mathbf{v}} = 0, \tag{1}$$

implies that the motion vector $\tilde{\mathbf{v}} = (v_1, v_2, 1)^T$ must be orthogonal to the local spatio-temporal image gradient ∇g. However, eq. (1) does not provide a unique solution for $\tilde{\mathbf{v}}$, and consequently additional constraints need to be introduced for obtaining a unique solution $\tilde{\mathbf{v}}$, see e.g., [6]. As a alternative [2], eq. (1) can be reformulated as

$$(\nabla g)(\nabla g)^T \tilde{\mathbf{v}} = \mathbf{0}, \tag{2}$$

which implies that $\tilde{\mathbf{v}}$ must be an eigenvector to $(\nabla g)(\nabla g)^T$, with zero eigenvalue. Again, this equation does not provide a unique $\tilde{\mathbf{v}}$, but if we compute a local mean of $(\nabla g)(\nabla g)^T$ over a region Ω in which $\tilde{\mathbf{v}}$ can be assumed to be constant, the corresponding relation becomes

$$\left[\int_\Omega p(\mathbf{x})\ (\nabla g_{\mathbf{x}})(\nabla g_{\mathbf{x}})^T \, d\mathbf{x} \right] \tilde{\mathbf{v}} = \mathbf{0}, \tag{3}$$

which under certain conditions provides unique solutions $\tilde{\mathbf{v}}$.

The expression within the brackets of eq. (3) is often called the *structure tensor*. It provides a compact representation of the local structure (moving point/line) by means of its eigenvalues, and the velocity in terms of its eigenvectors. In general, it carries information about the local 3D structure of the corresponding spatio-temporal volume, in particular its orientation, which implies that estimation of 3D orientation and structure can be used for computing of local motion. This means that other techniques for estimation of local orientation, not based on the mean of outer products of gradients, can be used also for motion estimation.

We will here review three techniques for estimation of local orientation, all resulting in a tensor based representation. These techniques are rather different in terms of the underlying signal models, but they provide more or less the similar result. The main result of the paper is a formulation of a common computational framework, in which the three techniques are just different ways of choosing parameters.

2. EARLY WORK

In [4], the double-angle representation for local 2D orientation is defined as the complex number $\mathbf{z} = A\, e^{2i\theta}$, where θ is the directional angle of a line or edge, and $A > 0$ is a measure of confidence for the orientation statement. Since $\arg(\mathbf{z}) = 2\theta$, it provides a continuous and averageable representation of local orientation. In [4], $\mathbf{z}$ is estimated locally by a process which measures “energy” in small regions of the Fourier domain (F.D.) and sets $\mathbf{z}$ according to where the maximum energy is found and how much it is.

In [8] it is shown that $\mathbf{z}$ can be computed by convolving the image with a set of so-called quadrature filters. The filters F_k are designed in the F.D. according to

$$F_k(\mathbf{u}) = R(u)\ \mathrm{ramp}^2(\hat{\mathbf{u}}^T \hat{\mathbf{m}}_k) \tag{4}$$

$$\mathrm{ramp}(x) = \frac{1}{2}(x + |x|), \quad \mathbf{u} = u\, \hat{\mathbf{u}}, \quad u = \|\mathbf{u}\| \tag{5}$$

where R is a radial weighting function, and $\hat{\mathbf{m}}_k$ are filter direction vectors. Notice that $F_k = 0$ when $\mathbf{u}^T \hat{\mathbf{m}}_k < 0$, which means that the filters are complex valued in the spatial domain (S.D.). Given that there are 3 or more filters in the set and that the filter directions $\hat{\mathbf{m}}_k$ are evenly distributed in a half-plane at angles θ_k, and q_k denotes the re-

---

This work has been made within the WITAS project, funded by the Knut and Alice Wallenberg foundation.

sponse from filter k, then $\mathbf{z}$ can be computed as

$$\mathbf{z} = \sum_k |q_k| \, e^{2i\theta_k} \tag{6}$$

Extensions of the 2D methods and representations for local orientation are highly relevant for motion estimation, but were not immediately discovered. The problem is addressed in [9] which proposes a 5D vector as a descriptor of 3D orientation. The vector is estimated using the same type of quadrature filter as in eq. (4), extended to the 3D case.

3. MATRICES AND TENSORS

A somewhat different path is explored in [1] by addressing the problem of finding the dominant orientation $\hat{\mathbf{n}}$ for the n-dimensional case. A solution is given by computing a matrix $\mathbf{J}$ which in the F.D. can be formulated as

$$\mathbf{J} = \int_{F.D.} \mathbf{u}\,\mathbf{u}^T \, |G(\mathbf{u})|^2 d\mathbf{u} \tag{7}$$

and has an approximation in the S.D. according to

$$\mathbf{J} = \int_{S.D.} p(\mathbf{x}) \, (\nabla g)(\nabla g)^T \, d\mathbf{x} \tag{8}$$

where p is a weight function (typically a Gaussian) and ∇g is the local gradient of g. Notice that $\mathbf{J}$ amounts to the expression inside the bracket of 3 and that in the ideal case, $\mathbf{J}$ can be expressed as

$$\mathbf{J} = \lambda \, \hat{\mathbf{n}} \, \hat{\mathbf{n}}^T. \tag{9}$$

where $\hat{\mathbf{n}}$ is the wanted orientation vector.

An algorithm for computing the elements of the matrix $\mathbf{J}$ by means of filter responses, q_k, from the same filters which were used to compute 5D vector is presented in [10]. The corresponding descriptor $\mathbf{T}$ is computed as

$$\mathbf{T} = \sum_k |q_k| \, \tilde{\mathbf{N}}_k \tag{10}$$

$$\tilde{\mathbf{N}}_k = \hat{\mathbf{m}}_k \, \hat{\mathbf{m}}_k^T - (N+2)^{-1} \, \mathbf{I} \tag{11}$$

where N is the dimensionality of the signal.

In [10], $\mathbf{T}$ is referred to as a *tensor*. The appropriateness of this concept is outside the scope of this presentation, but it should be noted that even though the estimation methods differ in eqs. (8) and (10), the result is the same for the ideal case. In the literature, *structure tensor* normally refers to constructing $\mathbf{J}$ according to eq. (8), which is a bit unfortunate if we want to distinguish the estimation procedure from the resulting representation. In the following, we will simply see $\mathbf{J}$ and $\mathbf{T}$ as similar types of descriptors, both given by eq. (9), and refer to them as *orientation tensors*.

4. POLYNOMIAL APPROXIMATION

Yet another method of estimating an orientation tensor is presented in [3]. It is based on a local polynomial model of the signal, using polynomials up to second order:

$$g_{\text{model}}(\mathbf{x}) = \mathbf{x}^T \, \mathbf{A} \, \mathbf{x} + \mathbf{x}^T \mathbf{b} + c \tag{12}$$

The approach is based on making a weighted least squares approximation of g_{model} to the local signal, giving $\mathbf{A}, \mathbf{b}, c$. Normally the weight function is a stationary function, referred to as the *applicability function* $a(\mathbf{x})$. Through the theory of normalized convolution [11], $\mathbf{A}, \mathbf{b}, c$ can be computed as filter responses where the filters are given by dual basis functions relative to the polynomial basis.

Furthermore, an orientation tensor can be computed as

$$\mathbf{T} = \mathbf{A}\mathbf{A}^T + \gamma \, \mathbf{b} \, \mathbf{b}^T \tag{13}$$

which for the ideal case of a single line or edge gives a tensor as in eq. (9). Notice that this $\mathbf{T}$ is obtained as a second order function of the local signal.

It is also shown that very efficient implementations for the estimation of $\mathbf{A}, \mathbf{b}, c$, and $\mathbf{T}$ can be made since the resulting filters are Cartesian separable.

5. COMMON FRAMEWORK

The issue of this presentation is whether there is a common framework in which the different methods for estimation of an orientation tensor can be placed. Steps in this direction has already been taken, see [7]. One answer is also given in [12], where it is shown that if the tensor should transform in a certain way relative to transformations of the signal, e.g., if a rotation of the signal rotates the eigenvectors of the tensor in the corresponding way, then sufficient and necessary conditions for polynomial functions from the signal to the elements of the tensor can be derived.

The method in eq. (10) cannot be characterized as a second order mapping, and cannot even be approximated as a power series since the magnitude function is not analytic. It should also be noted that the resulting tensor is not subject to equivariant transformations relative to rotations of the signal unless the signal is strictly simple.

On the other hand, the basic method which results in eq. (10) can be modified to a formulation of the tensor as a second order mapping

$$\mathbf{T} = \sum_k |q_k|^2 \, \tilde{\mathbf{N}}_k \tag{14}$$

if the corresponding set of quadrature filters and tensors $\tilde{\mathbf{N}}_k$ are chosen appropriately. This can, e.g., be done by keeping the same number of filters and filter directions $\hat{\mathbf{m}}_k$ but changing the filter functions to

$$F_k(\mathbf{u}) = R(u)\,\text{ramp}(\hat{\mathbf{u}}^T \hat{\mathbf{m}}_k) \tag{15}$$

The fact that all of the above methods (with the modified quadrature filter approach) can be formulated as second order mappings from signal to tensor indicates that such mappings can be used as a common ground for computing the tensor. See [13] for an overview of signal processing based on second order filters. If g is the local (real valued) signal, a general second order convolution on g can be defined as

$$q(\mathbf{y}) = \iint f(\mathbf{y}_1, \mathbf{y}_2)\, g(\mathbf{x} - \mathbf{y}_1)\, g(\mathbf{x} - \mathbf{y}_2)\, d\mathbf{y}_1\, d\mathbf{y}_2 \tag{16}$$

where $f(\mathbf{y}_1, \mathbf{y}_2) = f(\mathbf{y}_2, \mathbf{y}_1)$. Assuming a local view we can set $\mathbf{x} = 0, q = q(0)$, and get

$$q = \iint f(\mathbf{y}_1, \mathbf{y}_2)\, g(-\mathbf{y}_1)\, g(-\mathbf{y}_2)\, d\mathbf{y}_1\, d\mathbf{y}_2 \tag{17}$$

In the F.D. this corresponds to

$$q = \iint F(\mathbf{u}, \mathbf{v})\, G(\mathbf{u})\, G(\mathbf{v})\, d\mathbf{u}\, d\mathbf{v} \tag{18}$$

where F is the $2n$-dim. Fourier transform (F.T.) of the filter function f, and G is the n-dim. F.T. of g. Notice that if $q \in \mathbb{R}$ for all real g then

$$F(-\mathbf{u}, -\mathbf{v}) = \overline{F(\mathbf{u}, \mathbf{v})} \tag{19}$$

We may define a set of second order filters, one for each element of the tensor, T_{ij}, which we denote F_{ij}. Given that the tensor can be computed as a second order function of the signal, we get

$$\mathbf{T}_{ij} = \iint F_{ij}(\mathbf{u}, \mathbf{v})\, G(\mathbf{u})\, G(\mathbf{v})\, d\mathbf{u}\, d\mathbf{v} \tag{20}$$

The issue is now how should we choose the functions F_{ij} to get a tensor $\mathbf{T}$ which is useful for orientation representation? One formulation is already given in [1], eq. (7), where $\mathbf{T}$ is formed by integrating "energy" times the corresponding orientation tensor over all points in the F.D. To avoid the influence of high frequency components, it seems reasonable to also include a weight function

$$\mathbf{T} = \int_{F.D.} \mathbf{u}\,\mathbf{u}^T\, |G(\mathbf{u})|^2\, W^2(\mathbf{u}) d\mathbf{u} \tag{21}$$

which we assume to be rotational symmetric. This means that F_{ij} is given by

$$F_{ij}(\mathbf{u}, \mathbf{v}) = -\delta(\mathbf{u} + \mathbf{v})\, W(\mathbf{u})\, W(\mathbf{v})\, \mathbf{u}\,\mathbf{v}^T \tag{22}$$

This can be thought of as an ideal construction of $\mathbf{T}$ since it represents a superposition of energy contributions, so that if the signal contains two or more lines or edges, $\mathbf{T}$ is the sum of the corresponding orientation tensors. Unfortunately, this type of second order filter function cannot be realized since it has infinite support in the S.D. However, it can be approximated, e.g., as

$$F_{ij}(\mathbf{u}, \mathbf{v}) = -P(\mathbf{u} + \mathbf{v})\, \mathbf{u}_i\, \mathbf{v}_j\, W(\mathbf{u})\, W(\mathbf{v}) \tag{23}$$

where P is an approximation of the impulse function δ. For example, we can choose $P(\mathbf{u}) = e^{-(|\mathbf{u}|/\sigma)^2}/\sigma$, where σ is reasonably small. This corresponds to the formulation given in eq. (8), where the gradient is computed from $w * g$ rather than from g itself.

An alternative approach is to consider the case of ideal representation of a single line or edge. In this case $G(\mathbf{u})$ vanishes for all $\mathbf{u} \neq t\,\hat{\mathbf{n}}$, and eq. (20) can then be written

$$\mathbf{T}_{ij} = \iint_{-\infty}^{\infty} F_{ij}(t\,\hat{\mathbf{n}}, \tau\,\hat{\mathbf{n}})\, G'(t)\, G'(\tau)\, dt\, d\tau \tag{24}$$

where $G'(t) = G(t\,\hat{\mathbf{n}})$. A sufficient condition for F_{ij} to produce $\mathbf{T} = \lambda\,\hat{\mathbf{n}}\,\hat{\mathbf{n}}^T$ is then given by

$$F_{ij}(t\,\hat{\mathbf{n}}, \tau\,\hat{\mathbf{n}}) = \hat{\mathbf{n}}_i\,\hat{\mathbf{n}}_j\, H(t, \tau) \tag{25}$$

where H for the moment is arbitrary except for

$$H(t, \tau) = H(\tau, t), \quad H(-t, -\tau) = \overline{H(t, \tau)} \tag{26}$$

to make $\mathbf{T}$ real. This gives

$$\mathbf{T} = \hat{\mathbf{n}}\,\hat{\mathbf{n}}^T \iint H(t, \tau)\, G'(t)\, G'(\tau)\, dt\, d\tau \tag{27}$$

It is then natural to require that H is such that

$$\lambda = \iint H(t, \tau)\, G'(t)\, G'(\tau)\, dt\, d\tau \geq 0 \text{ for all } G' \tag{28}$$

From this follows that any set of functions F_{ij} which meet conditions in eqs. (25) and (28) will produce $\mathbf{T} = \lambda\,\hat{\mathbf{n}}\,\hat{\mathbf{n}}^T$, $\lambda > 0$, for the single orientation case.

6. TEST

For the case that $\mathbf{T}$ is estimated as in eq. (8), we get

$$H_1(t, \tau) = -P'(t + \tau)\, t\, \tau\, W'(t)\, W'(\tau) \tag{29}$$

where P' and W' are the 1-variable versions of P and W.

In the case that $\mathbf{T}$ is estimated as in eq. (14), for the case that the filter functions are given by eq. (15), we get

$$F_{ij}(t\,\hat{\mathbf{n}}, \tau\,\hat{\mathbf{n}}) = \sum_k (\hat{\mathbf{n}}^T \hat{\mathbf{m}}_k)^2\, \tilde{\mathbf{N}}_{ij,k} S(t, \tau) R(t) R(\tau) \tag{30}$$

$$S(t, \tau) = \frac{\text{step}(t)\text{step}(-\tau) + \text{step}(-t)\text{step}(\tau)}{2} \tag{31}$$

In [10], $\hat{\mathbf{m}}_k$ and $\tilde{\mathbf{N}}_k$ are always chosen so that

$$\sum_k (\hat{\mathbf{n}}^T \hat{\mathbf{m}}_k)^2 \, \tilde{\mathbf{N}}_{ij,k} = \hat{\mathbf{n}}_i \, \hat{\mathbf{n}}_j \tag{32}$$

from which follows that this F_{ij} satisfies eq. (25) with

$$H_2(t, \tau) = S(t, \tau) \, R(t) \, R(\tau) \tag{33}$$

Finally, if $\mathbf{T}$ is estimated according to eq. (13), using a symmetric Gaussian applicability function $a(\mathbf{x})$, it is easy to show that

$$F_{ij}(\mathbf{u}, \mathbf{v}) = \mathbf{u}_i \, \mathbf{v}_j \, [\frac{1}{4} \, \mathbf{u}^T \mathbf{v} - \gamma] \, A(\mathbf{u}) \, A(\mathbf{v}) \tag{34}$$

where A is the F.T. of a. From this follows

$$\begin{aligned} F_{ij}(t \, \hat{\mathbf{n}}, \tau \, \hat{\mathbf{n}}) &= \\ &= \hat{\mathbf{n}}_i \, \hat{\mathbf{n}}_j \, [\frac{1}{4} \, t \, \tau - \gamma] \, t \, \tau \, A'(t) \, A'(\tau) \end{aligned} \tag{35}$$

and

$$H_3(t, \tau) = [\frac{1}{4} \, t \, \tau - \gamma] \, t \, \tau \, A'(t) \, A'(\tau) \tag{36}$$

We have thus concluded that all three estimation techniques fall within the proposed framework, i.e., we can find second order filters which satisfy eq. (25), and the corresponding functions H_k all satisfy eq. (28) since the resulting tensor has $\lambda \geq 0$.

7. SUMMARY AND CONCLUSIONS

A short review of three methods for estimation of local orientation tensors has been given, together with relations to motion estimation. The signal models being used are rather different, but they produce similar result which motivates the existence of common framework for estimation of orientation tensors. Such a framework has been proposed; a second order mapping from signal to tensor, eq. (20), with additional conditions, eqs. (25) and (28). It has also been shown that the three methods in principle fall into the framework.

8. REFERENCES

[1] J. Bigün and G. H. Granlund. Optimal Orientation Detection of Linear Symmetry. In *Proceedings of the IEEE First International Conference on Computer Vision*, pages 433–438, London, Great Britain, June 1987.

[2] J. Bigün, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):775–790, August 1991. Report LiTH-ISY-I-1148, Linköping University, Sweden, 1990.

[3] Gunnar Farnebäck. *Polynomial Expansion for Orientation and Motion Estimation*. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, 2002. Dissertation No 790, ISBN 91-7373-475-6.

[4] G. H. Granlund. In Search of a General Picture Processing Operator. *Computer Graphics and Image Processing*, 8(2):155–173, 1978.

[5] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer Academic Publishers, 1995. ISBN 0-7923-9530-1.

[6] B. Jähne, H. Haussecker, and P. Geissler, editors. *Handbook of Computer Vision and Applications*. Academic Press, 1999. ISBN 0-12-379770-5.

[7] Björn Johansson and Gunnar Farnebäck. A Theoretical Comparison of Different Orientation Tensors. In *Proceedings SSAB02 Symposium on Image Analysis*, pages 69–73, Lund, March 2002. SSAB.

[8] H. Knutsson. *Filtering and Reconstruction in Image Processing*. PhD thesis, Linköping University, Sweden, 1982. Diss. No. 88.

[9] H. Knutsson. Producing a Continuous and Distance Preserving 5-D Vector Representation of 3-D Orientation. In *IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management - CAPAIDM*, pages 175–182, Miami Beach, Florida, November 1985. IEEE.

[10] H. Knutsson. Representing Local Structure Using Tensors. In *The 6th Scandinavian Conference on Image Analysis*, pages 244–251, Oulu, Finland, June 1989. Report LiTH-ISY-I-1019, Computer Vision Laboratory, Linköping University, Sweden, 1989.

[11] H. Knutsson and C-F. Westin. Normalized and Differential Convolution: Methods for Interpolation and Filtering of Incomplete and Uncertain Data. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 515–523, New York City, USA, June 1993. IEEE.

[12] K. Nordberg. *Signal Representation and Processing using Operator Groups*. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, 1995. Dissertation No 366, ISBN 91-7871-476-1.

[13] G. L. Sicuranza. Quadratic filters for signal processing. *Proceedings of the IEEE*, 80(1):1263–1285, August 1992.