On the Mathematical Foundations of Smoothness Constraints for the Determination of Optical Flow and for Surface Reconstruction

M. A. Snyder

Computer and Information Science Department
University of Massachusetts

COINS Technical Report 89-5

January 31, 1989

On the Mathematical Foundations of Smoothness Constraints for the Determination of Optical Flow and for Surface Reconstruction *

M. A. Snyder
Computer and Information Science
University of Massachusetts
Amherst, Mass. 01003

January 31, 1989

*supported by DARPA/Army ETL under grant DACA76-85-C-0008.

Abstract

Gradient–based approaches to the computation of optical flow often use a minimization technique incorporating a smoothness constraint on the optical flow field. In this paper, we derive the most general form of such a smoothness constraint which is quadratic in first derivatives of the flow field, and quadratic in first or second derivatives of the grey–level image intensity function, based on three simple assumptions about the smoothness constraint: (1) that it be expressed in a form which is independent of the choice of Cartesian coordinate system in the image; (2) that it be positive definite; and (3) that it not couple different components of the optical flow. We show that there are essentially only four such constraints; any smoothness constraint satisfying (1,2,3) must be a linear combination of these four, possibly multiplied by certain quantities invariant under a change in the Cartesian coordinate system. Beginning with the three assumptions mentioned above, we mathematically demonstrate that all the best–known smoothness constraints appearing in the literature are special cases of this general form, and, in particular, that the "weight matrix" introduced by Nagel is essentially (modulo invariant quantities) the only physically plausible such constraint. We also show that the results of Brady and Horn on "rotationally symmetric" performance measures for surface reconstruction are simple corollaries of our main results, and in fact that such performance measures are invariant under the larger group of transformations consisting of rigid motions of the plane.

1 Introduction

The computation of optical flow from a pair of frames in a dynamic image sequence is an important problem in computer vision. One of the major techniques that has been developed to address this problem is to minimize the sum of two functionals, one based on the (local) intensity constancy constraint, and the other on a more global feature of the optical flow field, usually called a smoothness constraint.

It is generally known that the local intensity constraint does not uniquely determine the optical flow vector **U** at a point in the image. Along linear structures in the image only the component normal to the local edge direction may be determined, and at points in homogeneous areas even that information may not be available [Anan87].

The simplest example of this gradient-based approach is the work of Horn and Schunck [Horn81]. Here, the temporal variation $\partial I/\partial t$ of the grey-level image intensity function $I(x,y,t)$ at a fixed point in the image, and the spatial variation $\vec{\nabla} I$ of I at a fixed time are measured. These two quantities are related (under various assumptions–see [Horn81,Schu84a,Schu84b,Horn87]) to the optical flow $\vec{U}(x,y,t)$ via an image intensity constancy constraint:

$$\frac{\partial I}{\partial t} + \vec{U} \cdot \vec{\nabla} I = 0, \tag{1}$$

or in matrix notation,

$$I_t + \mathbf{U}^T \boldsymbol{\nabla} I = 0. \tag{2}$$

Here we have defined the matrix $\mathbf{U}^T = (U_x, U_y) \equiv (u, v)$, and $\boldsymbol{\nabla}^T I \equiv (I_x, I_y)$; we denote the fact that some object is a matrix by using the corresponding bold face symbol. We represent the derivative of I with respect to the quantity ξ by $I_\xi \equiv \partial I / \partial \xi$.

The single equation (2) is not, of course, sufficient to determine the quantity **U**, since **U** has two components. Hence, it is clear that some additional constraint must be used to determine the optical flow field. Such a constraint typically demands some sort of consistency between neighboring flow vectors, i.e., it involves derivatives of **U**. Such constraints are usually called "smoothness constraints."

A significant problem for such an approach, however, is that there are situations, such as at motion boundaries, where neighboring flow vectors need not be consistent in this sense. Smoothness constraints will therefore be expected to encounter difficulties near such boundaries. Indeed, the physical motivation for the work of Nagel and Enkelmann on "oriented smoothness constraints" [Nage86] was to try to supress such a constraint in the direction perpendicular to such boundaries.

Because of the corrupting effects of noise, aliasing, and other artifacts of the measurement process (as well as for the cases where the intensity–constancy constraint (2) is not valid even in the ideal case), it is not ordinarily appropriate to express this smoothness constraint as an additional equation, since these corrupting influences will customarily prevent (2) from being satisfied exactly anyway. This suggests an approach which looks for a flow $\overline{\mathbf{U}}$ which minimizes some combination of the degree to which **U** fails to satisfy the intensity constancy constraint (2), and the variation of **U** from "smoothness." There is no physical reason why such a **U** should be the "correct" flow which gave rise to the measured spatiotemporal image gradients, but we would expect that except in pathological cases, $\overline{\mathbf{U}}$ should approximate the "correct" **U**.

The approach that has usually been taken is to minimize, over the space of all possible optical flow fields $\mathbf{U}$, a functional $\Pi[\mathbf{U}]$ given by

$$\Pi[\mathbf{U}] = \int \Sigma_{\text{int}}\, dx\, dy + \int \Sigma_{\text{sm}}\, dx\, dy, \tag{3}$$

where the integrals are over the image. Here

$$\Sigma_{\text{int}} = \left(I_t + \mathbf{U}^T \nabla I\right)^2, \tag{4}$$

and Σ_{sm} is some quantity related to "smoothness." Since upon integration the quantity Σ_{sm} yields a number related to the deviation of the flow field from "smoothness," we will call it the *smoothness density* or, for simplicity, simply the *density*. Brady and Horn [Brad83] discuss this approach in some detail, and cite psychophysical evidence that something very like this is performed by some parts of the human visual system.

We note that the reason for choosing Σ_{int} as a quadratic functional (rather than, say, a quartic) is for mathematical simplicity only—a quadratic functional is convex and hence unimodal. In other words, such functionals are manifestly positive definite (for real functions), a desirable feature of any minimization approach.

The choice for Σ_{sm} is less obvious, and has less physical justification, since "smooth" is a vague concept. But, as we have stated, smoothness densities usually involve derivatives of $\mathbf{U}$. The density should also be positive definite.[1] Clearly, it is simplest to choose Σ_{sm} to involve only first derivatives of $\mathbf{U}$ (although other choices could be—and have been—made). We will call such constraints *first degree smoothness constraints*.

We will confine ourselves to first degree constraints in this work, leaving the question of second or higher degree constraints (such as the second degree constraint considered by

[1]The requirement is actually that the quantity be bounded from below, so as to guarantee the existence of a minimum. But any such quantity can be made into a positive definite function by adding an appropriate constant. Since the constant does not depend on $\mathbf{U}$ or its derivatives, both quantities give the same Euler–Lagrange equations. Hence we may, without loss of generality, assume the functional to be positive definite.

Anandan and Weiss [Anan85]) to the sequel to this work [Snyd89]. Smoothness densities have usually been chosen on the basis of either simplicity, or of heuristic arguments. In this paper, we proceed in the opposite way by defining the smoothness density mathematically, and then deriving all possible such smoothness densities.

We therefore consider the most general form of such a quadratic first degree density:

$$\Sigma_{sm} = \sum_{ijk\ell} f_{ij}^{k\ell} \partial_i U_k \partial_j U_\ell \tag{5}$$

where $i, j, k, \ell = 1, 2$, with $\partial_i = \partial / \partial x_i$, $x_1 = x$, $x_2 = y$, $\vec{U} = (U_1, U_2) = (u, v)$ is the optical flow vector, and $f_{ij}^{k\ell}$ does not depend on $\mathbf{U}$ or its derivatives.

In order to make life bearable, we also introduce the *Einstein Summation Convention*: if in any product an index is repeated, it is to be understood that the index is to be summed over from 1 to 2. We therefore rewrite (5) as

$$\Sigma_{sm} = f_{ij}^{k\ell} \partial_i U_k \partial_j U_\ell. \tag{6}$$

Since i, j, k, and ℓ are repeated indices, it is understood that in (6) they are to be summed over. In the event that repeated indices in a product are not to be summed over, we will denote that by the phrase "(no sum)" next to the expression.

We will see in the next section that all of the smoothness densities so far proposed (see, e.g., [Horn81,Nage86]) satisfy the following three conditions:

1. They are invariant under a change of the Cartesian coordinate system in the image plane.

2. They are positive definite

3. They do not mix different components of **U**, i.e., the components of **U** are decoupled in (6).

We discuss the significance of each of these conditions in turn:

1. The condition that the smoothness density be invariant under a change of the Cartesian coordinate system of the image plane is equivalent to stating that the value obtained for the integral of Σ_{sm} over the image plane is independent of the coordinate system chosen for its evaluation. This seems eminently reasonable: the image itself has no preferred Cartesian coordinate system, so why impose one on it? This is equivalent to the condition that the Euler–Lagrange equations which follow from using this smoothness density are covariant under a change in coordinate system (i.e., that they have the same form in all Cartesian coordinate systems). We show in Appendix A that our condition is equivalent to the requirement that the density Σ_{sm} transform as a scalar under the action of the semi–direct product group ISO(2) of rigid transformations of the plane—the Euclidean group of the plane (see App. A). The situation in this respect is like the requirement of Galilean or Lorentz invariance in physics. Requiring that fundamental objects (like the Lagrangian) be invariant under some group of transformations results in the covariance of the resultant equations of motion under that same group of transformations.

2. The requirement of positive definiteness for Σ_{sm} is necessary to guarantee that $\Pi[\mathbf{U}]$ has a minimum, as discussed previously.

3. The requirement that the components of **U** be decoupled in Σ_{sm} has no physical basis I am aware of. We consider the effect of such a coupling in [Snyd89]. This

requirement is equivalent to demanding that $\Sigma_{sm}[\mathbf{U}] = \Sigma_{sm}[u] + \Sigma_{sm}[v]$.

In the next section, we elevate the properties (1,2,3) to the status of requirements for any smoothness constraint, and discuss the implications of this for the possible smoothness densities.

We believe that the demand of invariance under a change in the Cartesian coordinate system should be made not only of quantities like the smoothness density, but for any functional. We state this here as the "Zeroth Law of Computer Vision":

The 0^{th} Law of Computer Vision

Any functional having as domain functions defined over the image plane must be invariant under ISO(2)

2 The Definition of a Smoothness Density

2.1 General expression for the smoothness density

We will consider only smoothness densities of the form (6). We will require Σ_{sm} to satisfy the following Requirements:

I. Σ_{sm} is invariant under ISO(2).

II. Σ_{sm} is positive definite.

III. The structure of Σ_{sm} is such that the different components (u, v) of the optical flow field $\mathbf{U}$ are decoupled in all Cartesian coordinate systems. That is, Σ_{sm} can be written as the sum of two integrands, one which depends only on u, and the other only on v.

3. They do not mix different components of **U**, i.e., the components of **U** are decoupled in (6).

We discuss the significance of each of these conditions in turn:

1. The condition that the smoothness density be invariant under a change of the Cartesian coordinate system of the image plane is equivalent to stating that the value obtained for the integral of Σ_{sm} over the image plane is independent of the coordinate system chosen for its evaluation. This seems eminently reasonable: the image itself has no preferred Cartesian coordinate system, so why impose one on it? This is equivalent to the condition that the Euler–Lagrange equations which follow from using this smoothness density are covariant under a change in coordinate system (i.e., that they have the same form in all Cartesian coordinate systems). We show in Appendix A that our condition is equivalent to the requirement that the density Σ_{sm} transform as a scalar under the action of the semi–direct product group ISO(2) of rigid transformations of the plane—the Euclidean group of the plane (see App. A). The situation in this respect is like the requirement of Galilean or Lorentz invariance in physics. Requiring that fundamental objects (like the Lagrangian) be invariant under some group of transformations results in the covariance of the resultant equations of motion under that same group of transformations.

2. The requirement of positive definiteness for Σ_{sm} is necessary to guarantee that $\Pi[\mathbf{U}]$ has a minimum, as discussed previously.

3. The requirement that the components of **U** be decoupled in Σ_{sm} has no physical basis I am aware of. We consider the effect of such a coupling in [Snyd89]. This

requirement is equivalent to demanding that $\Sigma_{sm}[\mathbf{U}] = \Sigma_{sm}[u] + \Sigma_{sm}[v]$.

In the next section, we elevate the properties (1,2,3) to the status of requirements for any smoothness constraint, and discuss the implications of this for the possible smoothness densities.

We believe that the demand of invariance under a change in the Cartesian coordinate system should be made not only of quantities like the smoothness density, but for any functional. We state this here as the "Zeroth Law of Computer Vision":

The 0^{th} Law of Computer Vision

Any functional having as domain functions defined over the image plane must be invariant under ISO(2)

2 The Definition of a Smoothness Density

2.1 General expression for the smoothness density

We will consider only smoothness densities of the form (6). We will require Σ_{sm} to satisfy the following Requirements:

I. Σ_{sm} is invariant under ISO(2).

II. Σ_{sm} is positive definite.

III. The structure of Σ_{sm} is such that the different components (u, v) of the optical flow field $\mathbf{U}$ are decoupled in all Cartesian coordinate systems. That is, Σ_{sm} can be written as the sum of two integrands, one which depends only on u, and the other only on v.

We have seen that requirement **I** is necessary in order that the integral have a unique value, independent of the coordinate system, for a given optical flow and image intensity. Requirement **II** is necessary in order to ensure that the smoothness integral have a minimum. Requirement **III** is not in any sense necessary, but it is characteristic of all the smoothness densities so far proposed; it is equivalent to assuming that the two components of the optical flow are "smoothed" independently.

Requirement **I** has as an immediate consequence that the integrand $\Sigma_{sm} \equiv f$ must be an ISO(2) scalar, i.e., that $f'(\mathbf{r}') = f(\mathbf{r})$, where $\mathbf{r}' = \mathbf{R}\mathbf{r} + \mathbf{t}$; here $\mathbf{R} = \mathbf{R}(\theta) \in SO(2)$ is the 2×2 rotation matrix, and $\mathbf{t}$ is the translational vector (see Appendix A).

We now use two of the three requirements to write the smoothness density (6) in a more convenient form. We first use requirement **III** to see that f must be of the form:

$$
\begin{aligned}
f &= f^{11}_{ij}\partial_i U_1 \partial_j U_1 + f^{22}_{ij}\partial_i U_2 \partial_j U_2 \\
&\equiv A_{ij}\ \partial_i U_1 \partial_j U_1 + B_{ij}\ \partial_i U_2 \partial_j U_2,
\end{aligned} \tag{7}
$$

where $\mathbf{U}^T = (U_1, U_2)$. In order that this structure be valid in any Cartesian coordinate system, it follows that

$$
A_{ij} = B_{ij} \equiv F_{ij}. \tag{8}
$$

This can easily be seen as follows.

We note that under an ISO(2) transformation parametrized by θ and $\mathbf{T}$, $\mathbf{U}$ and $\boldsymbol{\nabla}$ transform like

$$
\mathbf{U} \longrightarrow \mathbf{U}' = \mathbf{R}\mathbf{U}, \quad \boldsymbol{\nabla} \longrightarrow \boldsymbol{\nabla}' = \mathbf{R}\boldsymbol{\nabla}, \tag{9}
$$

which means that

$$
\boldsymbol{\nabla}' u' = \cos\theta\, \mathbf{R}\boldsymbol{\nabla} u + \sin\theta\, \mathbf{R}\boldsymbol{\nabla} v,
$$

$$\nabla' v' = -\sin\theta \, \mathbf{R} \nabla u + \cos\theta \, \mathbf{R} \nabla v.$$

Therefore,

$$\begin{aligned}
\nabla^T u \mathbf{A} \nabla u + \nabla^T v \mathbf{B} \nabla v &\equiv (\nabla' u')^T \mathbf{A}' (\nabla' u') + (\nabla' v')^T \mathbf{B}' (\nabla' v') \\
&= \nabla^T u \left[\mathbf{R}^T (\mathbf{A}' \cos^2\theta + \mathbf{B}' \sin^2\theta) \mathbf{R} \right] \nabla u \\
&+ \nabla^T v \left[\mathbf{R}^T (\mathbf{A}' \sin^2\theta + \mathbf{B}' \cos^2\theta) \mathbf{R} \right] \nabla v \\
&+ 2 \sin\theta \cos\theta \, \nabla^T u \left[\mathbf{R}^T (\mathbf{A}' - \mathbf{B}') \mathbf{R} \right] \nabla v \qquad (10)
\end{aligned}$$

The absence of the last term for all θ implies that $\mathbf{A}' = \mathbf{B}'$. By taking $\theta = 0$, the claim (8) is established.

We can use (10) to show more, however. Using (8) in (10), we see that in terms of $\mathbf{F}$:

$$\mathbf{F}' = \mathbf{R} \mathbf{F} \mathbf{R}^T. \qquad (11)$$

That is, $\mathbf{F}$ transforms as a second-rank tensor under ISO(2).

We see that f may be written in the form

$$\begin{aligned}
f &= F_{ij} (\partial_i U_1 \partial_j U_1 + \partial_i U_2 \partial_j U_2) \\
&= F_{ij} (\partial_i U_k \partial_j U_k) = \partial_i U_k F_{ij} \partial_j U_k.
\end{aligned}$$

Defining the matrix $\mathbf{\Omega} \equiv \nabla \mathbf{U}^T$, having matrix elements $(\mathbf{\Omega})_{mn} = \partial_m U_n$, we see that

$$\begin{aligned}
f &= (\mathbf{\Omega})_{ki} (\mathbf{F})_{ij} (\mathbf{\Omega})_{jk} = \left(\mathbf{\Omega}^T \mathbf{F} \mathbf{\Omega} \right)_{kk} \qquad (12) \\
&\equiv \mathrm{tr} \left(\mathbf{\Omega}^T \mathbf{F} \mathbf{\Omega} \right), \qquad (13)
\end{aligned}$$

where $\mathrm{tr}(\mathbf{A}) = \mathbf{A}_{ii}$ denotes the trace of the matrix $\mathbf{A}$.

We will call the matrix **F** the *interaction* for the smoothness density f. This is appropriate, since the structure of **F** determines how the various derivatives of **U** combine (i.e., interact) in the smoothness density. Nagel [Nage86] was the first to write the smoothness density in this form. We see here that this is a simple consequence of Requirements I and III.

2.2 Properties of F; Scalar– and Tensor–based Interactions

Since $\mathbf{\Omega}$ is the outer product of two objects that transform as a vector, we see (cf. Appendix A) that $\mathbf{\Omega}$ (and hence $\mathbf{\Omega}^T$) behaves, under ISO(2), as a second rank tensor:

$$\mathbf{\Omega} \longrightarrow \mathbf{R}\mathbf{\Omega}\mathbf{R}^T, \tag{14}$$

$$\mathbf{\Omega}^T \longrightarrow \mathbf{R}\mathbf{\Omega}^T\mathbf{R}^T. \tag{15}$$

It follows from this transformation of $\mathbf{\Omega}$ and $\mathbf{\Omega}^T$, and the requirement that f be an ISO(2) scalar, that the interaction **F** must also transform as a second rank tensor, as we showed previously.

Furthermore, since

$$\operatorname{tr}\left[\mathbf{\Omega}^T\mathbf{F}\mathbf{\Omega}\right] = \operatorname{tr}[\mathbf{\Omega}^T\mathbf{F}\mathbf{\Omega}]^T = \operatorname{tr}[\mathbf{\Omega}^T\mathbf{F}^T\mathbf{\Omega}], \tag{16}$$

it follows that only the symmetric part of **F** contributes to the trace. As a consequence we lose no generality by limiting ourselves to symmetric interactions:

$$\mathbf{F} = \mathbf{F}^T. \tag{17}$$

Since both the identity matrix $\mathbf{1}_2$ and the antisymmetric matrix $\mathbf{J} \equiv \mathbf{R}(\pi/2)$ commute with **R** (see Appendix A), then if $\mathbf{F} = \sigma\mathbf{1}_2$ or $\mathbf{F} = \sigma\mathbf{J}$, where σ is an ISO(2) scalar, **F** will transform like a "tensor" (11). But since $\mathbf{F}$ is proportional to either $\mathbf{1}_2$ or $\mathbf{J}$, both of which commute with $\mathbf{R}$, it is also (in this case) true that

$$\mathbf{F} \longrightarrow \mathbf{F}' = \mathbf{R}\mathbf{F}\mathbf{R}^T = \mathbf{F}\mathbf{R}\mathbf{R}^T \equiv \mathbf{F}, \tag{18}$$

i.e., that $\mathbf{F}$ transforms like a scalar. Such a "tensor" interaction is a possible interaction that gives rise to an invariant smoothness density. We therefore have two classes of interactions which satisfy requirements **I** and **III**:

- **"scalar–based" interactions**, in which the interaction $\mathbf{F}$ commutes with $\mathbf{R}$, i.e., is of the form $\mathbf{F} = \sigma\mathbf{1}_2$ or $\mathbf{F} = \sigma\mathbf{J}$, and hence transforms like (18), where σ is an ISO(2) scalar,

- **"tensor–based" interactions**, in which $\mathbf{F}$ does not commute with $\mathbf{R}$, and transforms like (11) under ISO(2).

We emphasize that since $\mathbf{J}$ is antisymmetric, and only the symmetric part of $\mathbf{F}$ contributes to $\mathrm{tr}\left[\mathbf{\Omega}^T\mathbf{F}\mathbf{\Omega}\right]$, the only "scalar–based" interactions we need consider are those proportional to $\mathbf{1}_2$, with proportionality factor an ISO(2) scalar. The smoothness density which results from such a scalar–based interaction will therefore be a multiple of

$$\Sigma_{sm} = \mathrm{tr}\left[\mathbf{\Omega}^T\mathbf{\Omega}\right] = u_x^2 + u_y^2 + v_x^2 + v_y^2, \tag{19}$$

which is just the smoothness density originally proposed by Horn and Schunck [Horn81].

Any scalar–based interaction, therefore, will give rise to a Horn and Schunck–like smoothness density, modulated by an overall ISO(2) scalar factor. Such an interaction smooths $\mathbf{U}$ isotropically. It cannot, therefore, give rise to any "orientation–dependent" smoothness constraints (in the sense in which it is used by Nagel [Nage86]).

Looking ahead, we prove (Corollaries C.2.1 and C.3.1 of Appendix C) that there is only one independent scalar–based interaction quadratic in 1^{st} derivatives of the image intensity I, and only two independent such interactions quadratic in 2^{nd} derivatives of I, as was originally shown by Brady and Horn [Brad83]. (Note, however, that they showed this only for the rotational subgroup SO(2) of ISO(2).)

In the rest of this work, we will not consider scalar–based interactions, since they are trivially related to the Horn and Schunck interaction.

3 The Determination of All Possible Smoothness Densities Quadratic in First or Second Derivatives of I

Since we have assumed that all the dependence of Σ_{sm} on derivatives of the flow field is contained in the matrix $\mathbf{\Omega}$, it follows that in the absence of any significant pre–processing such as grouping or model recognition, the interaction $\mathbf{F}$ can only depend on the grey–level image intensity function $I(x, y)$. If we are to limit ourselves to tensor–based interactions, then we must construct out of I and its derivatives objects which transform like ISO(2) tensors. This is done in the next two Sections for the case of 1^{st} and 2^{nd} derivatives of I. The reader should consult Appendix A for the background necessary to understand the arguments of these sections.

3.1 Interactions that are Quadratic in First Derivatives of I

Since I is a scalar, it follows that there are precisely two vectors which can be constructed from the (two) first derivatives of I, namely, ∇I and its dual vector $\mathbf{J}\nabla I = \widetilde{\nabla} I$. It is shown in Theorem C.1 of Appendix C that these are the only two independent vectors which can be so constructed. Now a second rank tensor must, in particular, have two indices (the row and column of the matrix which represents it). This means that the minimal tensor interaction must be quadratic in first derivatives of I. We know from Appendix A that if $\mathbf{A}$ and $\mathbf{B}$ are vectors, then the outer product $\mathbf{A}\mathbf{B}^T$ transforms as a tensor. Consequently, we can construct four tensors that are quadratic in 1^{st} derivatives of I:

$$\nabla I \nabla^T I \equiv \mathbf{K} = \mathbf{K}^T = \begin{pmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{pmatrix} \tag{20}$$

$$\widetilde{\nabla} I \nabla^T I = \mathbf{J}\mathbf{K} = \begin{pmatrix} I_xI_y & I_y^2 \\ -I_x^2 & -I_xI_y \end{pmatrix} \tag{21}$$

$$\nabla I \widetilde{\nabla}^T I = \mathbf{K}\mathbf{J}^T = (\mathbf{J}\mathbf{K})^T = \begin{pmatrix} I_xI_y & -I_x^2 \\ I_y^2 & -I_xI_y \end{pmatrix} \tag{22}$$

$$\widetilde{\nabla} I \widetilde{\nabla}^T I \equiv \widetilde{\mathbf{K}} = \mathbf{J}\mathbf{K}\mathbf{J}^T = \begin{pmatrix} I_y^2 & -I_xI_y \\ -I_xI_y & I_x^2 \end{pmatrix} \tag{23}$$

The tensor $\widetilde{\mathbf{K}} = \mathbf{J}\mathbf{K}\mathbf{J}^T$ is called the *dual* of the tensor $\mathbf{K}$. Since $\mathbf{J} = \mathbf{R}(\pi/2)$, we see that $\widetilde{\mathbf{K}}$ is just the tensor $\mathbf{K}$, rotated by 90°. We show in Theorem C.2 of Appendix C that these are the **only** second rank tensors not proportional to $\mathbf{1}_2$ or $\mathbf{J}$ which are quadratic in the 1^{st} derivatives of I.

Therefore, the quantities (20)—(23) constitute a complete list of all the tensor–based interactions consistent with Requirements **I** and **III**.

We now impose the ancillary requirement (discussed in Section 2) that the interaction $\mathbf{F}$ be symmetric. Since $\mathbf{K}$ and $\widetilde{\mathbf{K}}$ are symmetric, they both pass this test. The interactions $\mathbf{J}\mathbf{K}$ and $\mathbf{K}\mathbf{J}^T$, however, are not symmetric. Since they are transposes of each other, it follows that they have the same symmetric part, namely (one–half of):

$$\mathbf{JK} + (\mathbf{JK})^T = \mathbf{JK} + \mathbf{KJ}^T = \mathbf{JK} - \mathbf{KJ} \equiv [\mathbf{J},\mathbf{K}], \tag{24}$$

where we have used $\mathbf{J}^T = -\mathbf{J}$. There are, therefore, precisely three independent tensor–based interactions quadratic in 1^{st} derivatives of I:

$$\mathbf{K}, \ \ \widetilde{\mathbf{K}}, \ \ [\mathbf{J},\mathbf{K}]. \tag{25}$$

We have used Requirements I and III to restrict the number of possible interactions to three in this case, but we have not yet demanded Requirement II, that the density be positive definite. This question is addressed by the following, which are proved in Appendix B:

Theorem B.1 Let $\Theta = \mathrm{tr}\left[\mathbf{\Omega}^T\mathbf{F}\mathbf{\Omega}\right]$. Then Θ is positive definite if and only if $\mathrm{tr}\,\mathbf{F} \geq 0$ and $\det \mathbf{F} \geq 0$.

Corollary B.1.1 If $\mathbf{F}$ is a real, traceless, symmetric 2×2 matrix different from 0. Then the associated $\Theta = \mathrm{tr}\left[\mathbf{\Omega}^T\mathbf{F}\mathbf{\Omega}\right]$ is not positive definite.

Corollary B.1.2 If $\mathbf{F} = \mathbf{M}^2$, where $\mathbf{M}$ is a symmetric, real 2×2 matrix different from zero, then $\Theta = \mathrm{tr}\left[\mathbf{\Omega}^T\mathbf{F}\mathbf{\Omega}\right] > 0$.

It is easy to see that the trace and determinant of $\mathbf{K}$ and $\widetilde{\mathbf{K}}$ are the same, and that

$$\mathrm{tr}\,\mathbf{K} = \mathrm{tr}\,(\nabla I \nabla^T I) = \mathrm{tr}\,(\nabla^T I \nabla I) = \nabla^T I \nabla I = \|\nabla I\|^2 \geq 0, \tag{26}$$

$$\det \mathbf{K} = \det \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix} = 0. \tag{27}$$

(We show in Section 5.1 that the singularity of $\mathbf{K}$ and $\widetilde{\mathbf{K}}$ is because both are projection operators.) Hence, both $\mathbf{K}$ and $\widetilde{\mathbf{K}}$ give rise to positive definite densities. On the other hand, the commutator of two (finite rank) matrices is always traceless, and since $[\mathbf{J},\mathbf{K}]$ is symmetric by construction, $[\mathbf{J},\mathbf{K}]$ is a symmetric, traceless, real matrix. Hence, from Corollary B.1.1, the associated density is not positive definite.

We summarize these results in the following theorem:

Theorem 1 *If a smoothness density satisfies Requirements* **I**, **II**, *and* **III**, *and is quadratic in* 1^{st} *derivatives of* I, *then the tensor-based interaction which gives rise to it must be of the form:*

$$\mathbf{F} = a_1\mathbf{K} + a_2\widetilde{\mathbf{K}},$$

where a_1 *and* a_2 *are constants.*

We note that our results here are slightly more general than they appear at first sight. Namely, we can construct tensors of order $2n$ in 1^{st} derivatives of I by simply constructing products of n tensors quadratic in 1^{st} derivatives of I, i.e., $\mathbf{K}, \mathbf{JK}, \mathbf{KJ}^T$, and $\widetilde{\mathbf{K}}$. It can be shown that such tensors of order $2n$ in 1^{st} derivatives of I are just equal to $|\nabla I|^{2n-2}$ (an ISO(2) scalar) times one of the original four tensors.

This is easy to show. We first denote the four tensors (20)—(23) as:

$$\mathbf{K}_1 = \mathbf{K}\ ;\ \ \mathbf{K}_2 = \mathbf{JK}\ ;\ \ \mathbf{K}_3 = \mathbf{KJ}^T\ ;\ \ \mathbf{K}_4 = \widetilde{\mathbf{K}}\ . \tag{28}$$

Each of these four is a product of the form

$$\mathbf{K}_i = \mathbf{A}_i\mathbf{B}_i{}^T \quad \text{(no sum)},$$

where $\mathbf{A}_i$ and $\mathbf{B}_i$ are either ∇I or $\widetilde{\nabla} I$. A typical term $\mathbf{K}_i\mathbf{K}_j$ quartic in 1^{st} derivatives of I is then of the form

$$(\mathbf{A}_i\mathbf{B}_i{}^T)(\mathbf{A}_j\mathbf{B}_j{}^T) = \mathbf{A}_i(\mathbf{B}_i{}^T\mathbf{A}_j)\mathbf{B}_j{}^T \quad \text{(no sum)}, \tag{29}$$

where $\mathbf{B}_i^T$ is either $\nabla^T I$ or $\tilde{\nabla}^T I$, and $\mathbf{A}_j$ equals either ∇I or $\tilde{\nabla} I$, depending on what i and j are. The term $(\mathbf{B}_i^T \mathbf{A}_j)$ in (29) is therefore simply a number, equal to one of the following:

$$\begin{aligned} \nabla^T I \nabla I &= \tilde{\nabla}^T I \tilde{\nabla} I = \|\nabla I\|^2, \\ \nabla^T I \tilde{\nabla} I &= \tilde{\nabla}^T I \nabla I = 0. \end{aligned}$$

consequently, the product $\mathbf{K}_i \mathbf{K}_j$ either vanishes identically (a trivial ISO(2) scalar), or is equal to

$$\mathbf{K}_i \mathbf{K}_j = \|\nabla I\|^2 \, \mathbf{A}_i \mathbf{B}_j^T.$$

But $\mathbf{A}_i \mathbf{B}_j^T$ must be one of the original four $\mathbf{K}_i$'s given by (28). It is then obvious that a non-vanishing tensor of order $2n$ in 1^{st} derivatives of I, given by a product of n of the $\mathbf{K}_i$'s, must be of the form

$$\mathbf{K}_{i_1} \mathbf{K}_{i_2} \cdots \mathbf{K}_{i_n} = \|\nabla I\|^{2n-2} \mathbf{K}_k,$$

for some k, which is exactly what we wanted to prove. We emphasize, however, that we have *not* shown that *all* tensors of order $2n$ in 1^{st} derivatives of I are of this form.

3.2 Interactions Quadratic in Second Derivatives of I

In this section, we consider interactions which are quadratic in second derivatives of I, i.e., $\mathbf{F}$ is of the form

$$\mathbf{F} = \mathbf{A}^{(ijk\ell)} \partial_i \partial_j I \partial_k \partial_\ell I, \tag{30}$$

where $\mathbf{A}^{(ijk\ell)}$ are *matrices* (*not* matrix elements!).

We can construct such tensors by noting that the quantities

$$\nabla\nabla^T I \equiv \mathbf{L} = \begin{pmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{pmatrix}, \tag{31}$$

$$\tilde{\nabla}\nabla^T I = \mathbf{JL} = \begin{pmatrix} I_{xy} & I_{yy} \\ -Ixx & -I_{xy} \end{pmatrix}, \tag{32}$$

$$\nabla\tilde{\nabla}^T I = \mathbf{LJ}^T = \begin{pmatrix} I_{xy} & -I_{xx} \\ I_{yy} & -I_{xy} \end{pmatrix}, \tag{33}$$

$$\tilde{\nabla}\tilde{\nabla}^T I = \mathbf{JLJ}^T \equiv \tilde{\mathbf{L}} = \begin{pmatrix} I_{yy} & -I_{xy} \\ -I_{xy} & I_{xx} \end{pmatrix} \tag{34}$$

are all tensors which are linear in 2[nd] derivatives of I. Hence, the $4 \times 4 = 16$ products of each of these tensors with each other are a set of tensors which are quadratic in 2[nd] order derivatives of I. Let us denote the above tensors by

$$\mathbf{L}_1 \equiv \mathbf{L}; \quad \mathbf{L}_2 \equiv \mathbf{JL}; \quad \mathbf{L}_3 \equiv \mathbf{LJ}^T; \quad \mathbf{L}_4 \equiv \tilde{\mathbf{L}}. \tag{35}$$

Upon computing all the possible products $\{\mathbf{L}_i\mathbf{L}_j; i, j = 1\ldots 4\}$, we find that there are only 8 different ones, given by:

$$\mathbf{L}^2,\ \tilde{\mathbf{L}}^2\ \mathbf{JL}^2,\ \mathbf{L}^2\mathbf{J}^T,\ \mathbf{L}\tilde{\mathbf{L}},\ \tilde{\mathbf{L}}\mathbf{L},\ \mathbf{JL}\tilde{\mathbf{L}},\ \mathbf{L}\tilde{\mathbf{L}}\mathbf{J}^T.$$

However, one easily checks from the explicit form of $\mathbf{L}$ and $\tilde{\mathbf{L}}$ that

$$\tilde{\mathbf{L}} = (\det \mathbf{L})\ \mathbf{L}^{-1},$$

and hence that

$$\mathbf{L}\tilde{\mathbf{L}} = \tilde{\mathbf{L}}\mathbf{L} = (\det \mathbf{L})\mathbf{1}_2; \quad \mathbf{JL}\tilde{\mathbf{L}} = -\tilde{\mathbf{L}}\mathbf{LJ}^T = (\det \mathbf{L})\mathbf{J}.$$

Since $\det \mathbf{L}$ is an obvious ISO(2) scalar, we see that these are in fact scalar–based interactions, and not tensor–based. We therefore drop them from further consideration.

This leaves the four tensors:

$$\mathbf{L}^2, \quad \tilde{\mathbf{L}}^2, \quad \mathbf{J}\mathbf{L}^2, \quad \mathbf{L}^2\mathbf{J}^T. \tag{36}$$

We have no *a priori* guarantee, however, that all second rank tensors quadratic in 2^{nd} order derivatives of I can be obtained in this way. That is, the same question of completeness arises here as it did in the previous section. Theorem C.3 of Appendix C, however, tells us that the arbitrary second rank tensor not proportional to $\mathbf{1}_2$ or $\mathbf{J}$, and quadratic in 2^{nd} order derivatives of I, is a linear combination of the four tensors (36). We therefore are assured that the only possible tensor–based interactions are those we have already found.

As in the previous section, the tensors $\mathbf{J}\mathbf{L}^2$ and $\mathbf{L}^2\mathbf{J}^T$ are transposes of each other and hence have the same symmetric part, which is easily seen to be

$$\mathbf{J}\mathbf{L}^2 + \left(\mathbf{J}\mathbf{L}^2\right)^T = [\mathbf{J}, \mathbf{L}^2].$$

It is also easy to see that $\mathbf{L}^2$ and $\tilde{\mathbf{L}}^2$ are symmetric.

We therefore have the following set of tensors which are symmetric, and which satisfy Requirements **I** and **III**:

$$\mathbf{L}^2, \quad \tilde{\mathbf{L}}^2, \quad [\mathbf{J}, \mathbf{L}^2],$$

We now impose the requirement that the corresponding density be positive definite (Requirement **II**). An analysis identical to that for the set (25) shows that $\mathbf{L}^2$ and $\tilde{\mathbf{L}}^2$ give rise to positive definite densities, and that $[\mathbf{J}, \mathbf{L}^2]$ does not. Hence, there are only two interactions which are quadratic in 2^{nd} derivatives of I and which satisfy requirements **I**, **II**, and **III**. This is summarized in the following theorem:

Theorem 2 *If* $\Theta = \mathrm{tr}\left[\Omega^T \mathbf{F} \Omega\right]$, *where* $\mathbf{F}$ *is quadratic in* 2^{nd} *derivatives of* I, *is a tensor-based smoothness density satisfying Requirements* I, II, *and* III, *then*

$$\mathbf{F} = a_0 \mathbf{L}^2 + a_1 \tilde{\mathbf{L}}^2,$$

where a_0 *and* a_1 *are constants.*

4 Relation to the Work of Brady and Horn on Performance Measures for Surface Reconstruction

Our claim that a smoothness density (or any other object, such as the "performance index" considered by Brady and Horn [Brad83]) be ISO(2) invariant generalizes the requirement proposed by these authors that such objects be "rotationally symmetric." We note that the central mathematical results of Brady and Horn (Propositions 2 and 6 in [Brad83], namely that $I_x^2 + I_y^2$ is the only independent "rotationally symmetric" scalar quadratic in 1^{st} derivatives of I, and that $(\nabla^2 I)^2$ (the "squared Lagrangian") and $I_{xx}^2 + 2I_{xy}^2 + I_{yy}^2$ (the "quadratic variation") are the only independent "rotationally symmetric" quantities quadratic in 2^{nd} derivatives of I appear (with "rotationally symmetric" replaced by the more general "ISO(2) scalar") as immediate corollaries (C.2.1 and C.3.1) of our main theorems. In addition, the cumbersome "tensor product" notation used by them is seen to be unnecessary, and to have a more elegant expression in terms of the tensors and scalars we have introduced here.

5 Relation to the Work of Nagel and Enkelmann

In searching for a smoothness constraint that would be more effective than the isotropic constraint of Horn and Schunck at computing optical flow near motion and depth boundaries, Nagel [Nage83,Nage87], and Nagel and Enkelmann [Nage86], noted that a suppression of the constraint along image gradients and along one of the two principal directions of the image intensity surface, would accomplish that task to a certain degree. Such a smoothness constraint is called by them an "oriented smoothness constraint." What we have called the "interaction" is called by them a "weight matrix." The interaction introduced by them is essentially a linear combination of the tensor interactions $\widetilde{\mathbf{K}}$ and $\tilde{\mathbf{L}}^2$ we introduced in Section 3. Indeed, the matrices $\mathbf{F}$ and $\mathbf{C}^{-1}$ given in [Nage86] are given in our notation by

$$\mathbf{F} = \widetilde{\mathbf{K}} + b^2\tilde{\mathbf{L}}^2 \quad (37)$$

$$\mathbf{C}^{-1} = \frac{\mathbf{F}}{\det \mathbf{F}}. \quad (38)$$

They also considered other normalizations of the interaction $\mathbf{F}$, such as dividing $\mathbf{F}$ by $\operatorname{tr}\mathbf{F}$. It is clear that since the determinant and trace of a tensor are ISO(2) invariant, such normalizations are simply multiplication of a tensor-based ISO(2) density by an ISO(2) scalar. Similar comments obtain for the later smoothness constraint discussed in the recent paper by Nagel [Nage88]. We note also that the smoothness constraint used by Hildreth [Hild83] is shown by Nagel [Nage87] to be a special case of the smoothness constraint (38).

In the next two sections, we discuss the geometry of the four interactions we have found, and show that the approach of Nagel and Enkelmann to "oriented smoothness" constraints is essentially the only possible one.

5.1 The Geometrical Interpretation of K and $\widetilde{\mathbf{K}}$

In this section we investigate the geometrical interpretation of $\mathbf{K}$ and $\widetilde{\mathbf{K}}$ as smoothness interactions. Our analysis makes more explicit the comments of Nagel [Nage86] on these interactions.

We recall from (20) and (23) that $\mathbf{K}$ and $\widetilde{\mathbf{K}}$ are defined as:

$$\mathbf{K} = \nabla I \, \nabla^T I \quad ; \quad \widetilde{\mathbf{K}} = \widetilde{\nabla} I \, \widetilde{\nabla}^T I.$$

We define the unit vectors $\hat{\epsilon}_1$ and $\hat{\epsilon}_2$ as

$$\hat{\epsilon}_1 = \frac{\nabla I}{\|\nabla I\|} \tag{39}$$

$$\hat{\epsilon}_2 = \frac{\widetilde{\nabla} I}{\|\widetilde{\nabla} I\|}. \tag{40}$$

These vectors may be defined everywhere that $\nabla I \neq 0$ (which we assume in the rest of this discussion). They form a local basis for the image plane:

$$\hat{\epsilon}_a{}^T \hat{\epsilon}_b = \delta_{ab}. \tag{41}$$

Recalling that $\widetilde{\nabla} I = \mathbf{J}\nabla I = \mathbf{R}(\pi/2)\nabla I$, we see that $\hat{\epsilon}_1$ is a unit vector normal to the isointensity contour, and $\hat{\epsilon}_2$ is a unit vector tangent to the isointensity contour.

We see that $\mathbf{K}$ and $\widetilde{\mathbf{K}}$ can be expressed as

$$\mathbf{K} = \|\nabla I\|^2 \mathbf{P}, \tag{42}$$

$$\widetilde{\mathbf{K}} = \|\nabla I\|^2 \mathbf{Q}, \tag{43}$$

where the operators (matrices) $\mathbf{P}$ and $\mathbf{Q}$ are given by

$$\hat{\epsilon}_1\hat{\epsilon}_1^T \equiv \mathbf{P} = \mathbf{P}^T, \tag{44}$$

$$\hat{\epsilon}_2\hat{\epsilon}_2^T \equiv \mathbf{Q} = \mathbf{Q}^T. \tag{45}$$

It follows from (41) that the set $\{\mathbf{P}, \mathbf{Q}\}$ obeys the algebra of a complete set of orthogonal projection operators, namely:

$$\begin{aligned} \mathbf{P}^2 &= \mathbf{P} \; ; \; \mathbf{Q}^2 = \mathbf{Q} \\ \mathbf{PQ} &= \mathbf{QP} = 0 \; ; \\ \mathbf{P} + \mathbf{Q} &= \mathbf{1}_2 \, . \end{aligned} \tag{46}$$

It is easy to show from this that the determinants of both $\mathbf{P}$ and $\mathbf{Q}$ are zero—which is characteristic of any projection operator not equal to the identity. This explains our statement in the previous section regarding the singularity of $\mathbf{K}$ and $\widetilde{\mathbf{K}}$.

Since $\{\hat{\epsilon}_1, \hat{\epsilon}_2\}$ form a local basis for the image plane, the effect of $\mathbf{P}$ and $\mathbf{Q}$ on any vector field is uniquely determined by their effect on $\hat{\epsilon}_1$ and $\hat{\epsilon}_2$, which is found to be

$$\begin{aligned} \mathbf{P}\hat{\epsilon}_1 &= \hat{\epsilon}_1 \; ; \; \mathbf{P}\hat{\epsilon}_2 = 0 \; ; \\ \mathbf{Q}\hat{\epsilon}_1 &= 0 \; ; \; \mathbf{Q}\hat{\epsilon}_2 = \hat{\epsilon}_2. \end{aligned} \tag{47}$$

Any vector $\mathbf{V}$ can be expressed as the sum of a vector $\mathbf{V}_\perp$ along $\hat{\epsilon}_1$ (i.e., perpendicular to the image isointensity contour), and a vector $\mathbf{V}_\parallel$ along $\hat{\epsilon}_2$ (i.e., parallel to the image isointensity contour):

$$\mathbf{V} = \mathbf{V}_\perp + \mathbf{V}_\parallel, \tag{48}$$

where

$$\mathbf{V}_\perp = \hat{\epsilon}_1 \left(\hat{\epsilon}_1^T \mathbf{V}\right) = \left(\hat{\epsilon}_1\hat{\epsilon}_1^T\right) \mathbf{V} \equiv \mathbf{PV}, \tag{49}$$

$$\mathbf{V}_{\|} = \hat{\epsilon}_2 \left(\hat{\epsilon}_2^T \mathbf{V}\right) = \left(\hat{\epsilon}_2 \hat{\epsilon}_2^T\right) \mathbf{V} = \mathbf{Q}\mathbf{V}. \tag{50}$$

These relations are, of course, the reason why **P** and **Q** are called "orthogonal projection operators"—**P** projects any vector onto its component in the direction ($\hat{\epsilon}_1$) perpendicular to the isointensity contour, while **Q** projects any vector onto its component in the orthogonal direction ($\hat{\epsilon}_2$) parallel to the isointensity contour.

Since $\|\nabla I\|^2$ is "orientation–independent" (in Nagel's sense), the "orientational" properties of **K** and $\widetilde{\mathbf{K}}$ are identical to those of **P** and **Q**, respectively. Hence, we lose no generality by discussing only the latter.

If the quantities $\nabla_{\|} u$ and $\nabla_{\perp} u$ (and the corresponding quantities for v) are defined as

$$\begin{aligned} \nabla_{\|} u &= \mathbf{P} \nabla u, \\ \nabla_{\perp} u &= \mathbf{Q} \nabla u, \end{aligned}$$

then since

$$\boldsymbol{\Omega} = \nabla \mathbf{U}^T = (\nabla u\ ,\ \ \nabla v), \tag{51}$$

it follows that

$$\begin{aligned} \mathbf{P}\boldsymbol{\Omega} &= (\mathbf{P}\nabla u\ ,\ \ \mathbf{P}\nabla v) \\ &= (\nabla_{\perp} u\ ,\ \nabla_{\perp} v). \end{aligned} \tag{52}$$

We then see that since $\mathbf{P} = \mathbf{P}^2$,

$$\begin{aligned} \mathrm{tr}\left[\boldsymbol{\Omega}^T \mathbf{P} \boldsymbol{\Omega}\right] &= \mathrm{tr}\left[\boldsymbol{\Omega}^T \mathbf{P}^2 \boldsymbol{\Omega}\right] \\ &= \mathrm{tr}\left[\boldsymbol{\Omega}^T \mathbf{P}^T \mathbf{P} \boldsymbol{\Omega}\right] \\ &= \mathrm{tr}\left[\{\mathbf{P}\boldsymbol{\Omega}\}^T \{\mathbf{P}\boldsymbol{\Omega}\}\right] \end{aligned}$$

$$= \text{tr}\left[\{\mathbf{P\Omega}\}\{\mathbf{P\Omega}\}^T\right]. \qquad (53)$$

Consequently, using (52),

$$\begin{aligned}
\text{tr}\left[\mathbf{\Omega}^T\mathbf{P}\mathbf{\Omega}\right] &= \text{tr}\left(\nabla_\perp u\ ,\ \nabla_\perp v\right)\begin{pmatrix}(\nabla_\perp u)^T\\(\nabla_\perp v)^T\end{pmatrix}\\
&= \text{tr}\left[\nabla_\perp u\,(\nabla_\perp u)^T\right] + \text{tr}\left[\nabla_\perp v\,(\nabla_\perp v)^T\right]\\
&= \text{tr}\left[(\nabla_\perp u)^T\,\nabla_\perp u\right] + \text{tr}\left[(\nabla_\perp v)^T\,\nabla_\perp v\right]\\
&= (\nabla_\perp u)^T\,\nabla_\perp u + (\nabla_\perp v)^T\,\nabla_\perp v
\end{aligned}$$

Hence,

$$\text{tr}\left[\mathbf{\Omega}^T\mathbf{P}\mathbf{\Omega}\right] = \|\nabla_\perp u\|^2 + \|\nabla_\perp v\|^2. \qquad (54)$$

Similar calculations for $\mathbf{Q}$ give

$$\text{tr}\left[\mathbf{\Omega}^T\mathbf{Q}\mathbf{\Omega}\right] = \|\nabla_\| u\|^2 + \|\nabla_\| v\|^2. \qquad (55)$$

Therefore (using (42) and (43)),

$$\begin{aligned}
\text{tr}\left(\mathbf{\Omega}^T\mathbf{K}\mathbf{\Omega}\right) &= \|\nabla I\|^2\left\{\|\nabla_\perp u\|^2 + \|\nabla_\perp v\|^2\right\}\ , &(56)\\
\text{tr}\left(\mathbf{\Omega}^T\widetilde{\mathbf{K}}\mathbf{\Omega}\right) &= \|\nabla I\|^2\left\{\|\nabla_\| u\|^2 + \|\nabla_\| v\|^2\right\}. &(57)
\end{aligned}$$

Recalling the significance of the subscripts "$\|$" and "$\perp$", we see that if $\mathbf{K}$ (or $\widetilde{\mathbf{K}}$) is used for the smoothness interaction, "smoothness" will be demanded only of the components of ∇u and ∇v perpendicular (or parallel) to the image isointensity contours. Although there is no physical basis for *demanding* smoothness along one of these directions, there is at least some justification for *not* demanding smoothness of the flow field components perpendicular to the image isointensity contours. Since these contours often (but not necessarily) correspond to physically meaningful (motion, occlusion) boundaries, it is often the case that the optical flow field varies strongly—perhaps even being discontinuous—perpendicular to such boundaries. Consequently, it would seem that the interaction **K** should most definitely *not* be used for a smoothness interaction. (This argument is due to Nagel [Nage86].) This leaves only the quantity $\widetilde{\mathbf{K}}$ as a possible smoothness interaction quadratic in 1^{st} derivatives of the image intensity.

5.2 The Geometrical Interpretation of $\mathbf{L}^2$ and $\tilde{\mathbf{L}}^2$

In this section we discuss the geometry of choosing $\mathbf{L}^2$ or $\tilde{\mathbf{L}}^2$ as smoothness interactions. We will show that

$$\mathrm{tr}\left[\Omega^T \mathbf{L}^2 \Omega\right] = \lambda_1^2(u_1^2 + v_1^2) + \lambda_2^2(u_2^2 + v_2^2), \tag{58}$$

$$\mathrm{tr}\left[\Omega^T \tilde{\mathbf{L}}^2 \Omega\right] = \lambda_2^2(u_1^2 + v_1^2) + \lambda_1^2(u_2^2 + v_2^2), \tag{59}$$

where u_1 and u_2 are the components of ∇u along the principal directions of the intensity surface corresponding to principal curvatures (proportional to) λ_1 and λ_2, respectively (with similar definitions for v). That is, the interaction $\mathbf{L}$ ($\tilde{\mathbf{L}}$) smoothes the optical flow preferentially in the direction of maximum (minimum) curvature. Note, in contrast to the previous section, that $\mathbf{L}$ and $\tilde{\mathbf{L}}$ are *not* projection operators. This makes more precise the remarks of Nagel [Nage86] in this regard.

The matrix $\mathbf{L}$ is just the Hessian matrix of the grey-level intensity function $I(x,y)$. Since the Hessian matrix appears naturally as the coefficient matrix for the second order terms in the Taylor series expansion of I, it is to be expected that the Hessian matrix is related to the curvature properties of the surface represented by $I(x, y)$. This is because the second order terms in the Taylor series express the difference between the given function and its "tangent plane" approximation.

These remarks are made more precise by considering the fundamental object in the theory of curvature [Thor79], namely the Gauss–Weingarten map Π which maps the tangent space T_P of the surface $I(x, y)$ at the point P into itself, via

$$\begin{aligned} \Pi \;:\; T_P &\longmapsto T_P \\ \Pi \;:\; \mathbf{v} &\longmapsto -(\mathbf{v} \cdot \nabla)\hat{\mathbf{n}}. \end{aligned}$$

Here $\mathbf{v}$ is a vector in the tangent space T_P to $I(x, y)$ at P, and $\hat{\mathbf{n}}$ is the unit normal to the surface at the same point. Physically, Π maps the tangent vector $\mathbf{v}$ into the tangent vector given by $\|\mathbf{v}\|$ times the (negative of the) rate of change of the unit normal in the direction $\mathbf{v}$.

It is physically clear that there are in general two special directions $\mathbf{v}$, namely those in which the unit normal stays in the plane spanned by $\mathbf{v}$ and $\hat{\mathbf{n}}$ (at the point P) as one proceeds in the direction $\mathbf{v}$ away from P. These directions are called the principal directions of the surface at the point P. According to the definition of the Gauss–Weingarten map above, these directions are eigenvectors of the map. The corresponding eigenvalues are the principal curvatures of the surface at the point P.

We can give a matrix interpretation to this by noting that if the following unit tangent vectors are chosen as the (non-Cartesian) basis for the tangent space:

$$\mu_1 = \frac{(1,0,I_x)}{\sqrt{1+I_x^2}};$$

$$\mu_2 = \frac{(0,1,I_y)}{\sqrt{1+I_y^2}},$$

then [doCa76] the matrix representation of the Gauss-Weingarten map is just given by

$$\Pi = \frac{\mathbf{L}}{[1+\|\nabla I\|^2]^{3/2}}.$$

It is clear from this that to within an ISO(2)-invariant factor the matrix $\mathbb{L}$ and the matrix Π are the same. They therefore have the same eigenvectors (the principal directions), and their eigenvalues are proportional to each other.

Let Ψ_1 and Ψ_2 be the normalized eigenvectors of $\mathbf{L}$:

$$\mathbf{L}\Psi_1 = \lambda_1\Psi_1, \tag{60}$$

$$\mathbf{L}\Psi_2 = \lambda_2\Psi_2, \tag{61}$$

with

$$\Psi_a{}^T\Psi_b \equiv \delta_{ab} \quad (a,b=1,2). \tag{62}$$

If the point in question is not an umbilic point ($\lambda_1 \neq \lambda_2$), then the inequality of the eigenvalues of $\mathbf{L}$ guarantees the orthogonality of the eigenvectors. If the point is umbilic ($\lambda_1 = \lambda_2$), then appropriate linear combinations of the eigenvectors can be chosen so that the combinations are orthogonal (the Gram-Schmidt orthogonalization process). At any rate, we can always assume that (62) holds.

We now (in analogy to the previous section) define the projection operators:

$$\mathbf{P}_1 = \mathbf{\Psi}_1 \mathbf{\Psi}_1^T ; \tag{63}$$

$$\mathbf{P}_2 = \mathbf{\Psi}_2 \mathbf{\Psi}_2^T . \tag{64}$$

It is easy to check that the operators $\{\mathbf{P}_1, \mathbf{P}_2\}$ are a set of orthogonal projection operators, in the sense of (46), and that they have the following effect on the eigenvectors of $\mathbf{L}$:

$$\mathbf{P}_1\mathbf{\Psi}_1 = \mathbf{\Psi}_1, \quad \mathbf{P}_1\mathbf{\Psi}_2 = 0,$$

$$\mathbf{P}_2\mathbf{\Psi}_1 = 0, \quad \mathbf{P}_2\mathbf{\Psi}_2 = \mathbf{\Psi}_2.$$

We can say even more, however. Since $\mathbf{\Psi}_2$ is perpendicular to $\mathbf{\Psi}_1$, $\mathbf{\Psi}_2$ must be $\mathbf{\Psi}_1$ rotated by 90°. That is, we can always choose

$$\mathbf{\Psi}_2 = \mathbf{J}\mathbf{\Psi}_1.$$

Furthermore, this implies that

$$\mathbf{P}_2 = \mathbf{J}\mathbf{P}_1\mathbf{J}^T, \tag{65}$$

i.e., the projection operator $\mathbf{P}_2$ is just the dual of the projection operator $\mathbf{P}_1$. (Note that the projection operators $\mathbf{P}$ and $\mathbf{Q}$ of the previous section are similarly related.)

It is well known [Stra86] that any symmetric matrix can be expressed in terms of its eigenvectors and eigenvalues. In our case this takes the form:

$$\mathbf{L} = \lambda_1\mathbf{P}_1 + \lambda_2\mathbf{P}_2. \tag{66}$$

It is then obvious, upon using (65), that

$$\tilde{\mathbf{L}} = \lambda_2\mathbf{P}_1 + \lambda_1\mathbf{P}_2. \tag{67}$$

This then implies that

$$\tilde{\mathbf{L}}\mathbf{\Psi}_1 = \lambda_2\mathbf{\Psi}_1, \quad (68)$$

$$\tilde{\mathbf{L}}\mathbf{\Psi}_2 = \lambda_1\mathbf{\Psi}_2. \quad (69)$$

Comparing this with the properties (60) and (61), we see that $\mathbf{L}$ and $\tilde{\mathbf{L}}$ have the same eigenvectors, but with the eigenvalues switched.

If we use the fact that the square of an operator has the same eigenvectors as the original operator, but with the corresponding eigenvalues squared, we see that $\mathbf{L}^2$ and $\tilde{\mathbf{L}}^2$ have the expression:

$$\mathbf{L}^2 = \lambda_1^2\mathbf{P}_1 + \lambda_2^2\mathbf{P}_2 \quad (70)$$

$$\tilde{\mathbf{L}}^2 = \lambda_2^2\mathbf{P}_1 + \lambda_1^2\mathbf{P}_2 \quad (71)$$

An analysis identical to that given in the previous section then shows that

$$\mathrm{tr}\left[\mathbf{\Omega}^T\mathbf{L}^2\mathbf{\Omega}\right] = \lambda_1^2(u_1^2 + v_1^2) + \lambda_2^2(u_2^2 + v_2^2), \quad (72)$$

$$\mathrm{tr}\left[\mathbf{\Omega}^T\tilde{\mathbf{L}}^2\mathbf{\Omega}\right] = \lambda_2^2(u_1^2 + v_1^2) + \lambda_1^2(u_2^2 + v_2^2), \quad (73)$$

where u_1 (or u_2) is the projection of ∇u along the principal direction $\mathbf{\Psi}_1$ (or $\mathbf{\Psi}_2$), with similar definitions for v_1 and v_2:

$$u_1 = \mathbf{P}_1\nabla u \quad ; \quad u_2 = \mathbf{P}_2\nabla u \quad (74)$$

$$v_1 = \mathbf{P}_1\nabla v \quad ; \quad v_2 = \mathbf{P}_2\nabla v \quad (75)$$

Owing to the proportionality of the eigenvalues λ_1 and λ_2 to the principal curvatures, it is clear that the interaction based on $\mathbf{L}$ "smoothes" the optical flow field preferentially in the direction corresponding to the maximum eigenvalue of the Gauss–Weingarten operator, i.e., the direction of *maximum* principal curvature, whereas the interaction based on the dual $\tilde{\mathbf{L}}$ "smoothes" the optical flow field preferentially in the direction of *minimum* principal curvature. Since the direction of largest principal curvature is often perpendicular to a physically significant (motion or occlusion) boundary, it would seem that the use of $\mathbf{L}$ as the smoothness interaction would be exactly the wrong thing to do; hence it does not make sense to use $\mathbf{L}$, and $\tilde{\mathbf{L}}$ is the only other possibility.

Combining the results of this and the previous section, we conclude that (modulo ISO(2) invariant quantities), the approach of Nagel and Enkelmann is the only physically reasonable one, assuming a first degree smoothness constraint quadratic in either 1^{st} or 2^{nd} derivatives of I.

6 Conclusions

We have shown in this work that by using three simple and reasonable assumptions about the characteristics of smoothness constraints, there are essentially only 4 independent smoothness constraints that are quadratic in 1^{st} derivatives of the optical flow field, and quadratic in either 1^{st} or 2^{nd} derivatives of the grey–level image intensity function. Only two of these four are physically plausible, and they correspond to those chosen by Nagel and Enkelmann. All other such smoothness constraints can be obtained as linear combinations of these 4, perhaps multiplied by ISO(2) scalar functions of the image intensity and its derivatives.

We also derived generalized versions of the results of Brady and Horn on the possible performance measures that can be used for surface reconstruction, and found them to be simple corollaries of our main results for optical flow.

In the continuation of this work [Snyd89], we investigate the more complicated problem of classifying smoothness densities quadratic in 2^{nd} derivatives of the optical flow field, and the implications of relaxing Requirement III, that the optical flow components are decoupled. Perhaps a coupling of these components in the smoothness constraint can lead to interesting smoothness constraints. For instance, the physically sensible smoothness constraint should reflect a smoothness in the three-dimensional flow field. Upon central projection, this will become a smoothness constraint on the two-dimensional optical flow. Because of the projection, such a two-dimensional smoothness constraint should be of the coupled variety. Consequently, perhaps it is the coupled smoothness constraints which are the most interesting [P. Anandan, personal communication]. We are presently investigating this idea.

Acknowledgements

I thank Prof. E. Riseman for his encouragement and support, and the members of the motion group at the University of Massachusetts for comments that led to the improvement of the ideas contained here. I also thank P. Anandan for several illuminating discussions.

References

[Anan87] P. Anandan, "A Unified Perspective on Computational Techniques for the Measurement of Visual Motion," Proc. 1^{st} Int. Conf. on Comp. Vis., London, England (June 1987).

[Anan85] P. Anandan and R. Weiss, "Introducing a Smoothness Constraint in a Matching

Approach for the Computation of Displacement Fields," COINS Tech. Rep. 85–38, UMass, Amherst (Dec. 1985).

[Brad83] M. Brady and B. K. P. Horn, "Rotationally Symmetric Operators for Surface Interpolation," CVGIP **22**, 70–94 (1983).

[doCa76] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice–Hall (1976).

[Hild83] E. Hildreth, *The Measurement of Visual Motion*, MIT Press, Cambridge, Mass. (1983).

[Grim81] W. E. L. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*, MIT Press, Cambridge, Mass. (1981).

[Horn87] B. K. P. Horn and E. J. Weldon, Jr. "Robust Direct Methods for Recovering Motion ," Univ. Hawaii at Manoa Tech. Rep. (Apr. 1987).

[Horn81] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow," Art. Intell. **23**, 185–203 (1981).

[Nage83] H.-H. Nagel, "Constraints for the Estimation of Displacement Vector Fields from Image Sequences," Proc. IJCAI83, 945–951, Karlsruhe, W. Germany (1983).

[Nage86] H.-H. Nagel and W. Enkelmann, "An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences," IEEE Trans. Patt. An. Mach. Intell. **PAMI–8**, No. 5, 565–593 (Sept. 1986).

[Nage87] H.-H. Nagel, "On the Estimation of Optical Flow: Relations between Different Approaches and Some New Results," Art. Intell. **33**, 299–324 (1987).

[Nage88] H.-H. Nagel, "Image Sequences–Ten (Octal) Years–from Phenomenology towards a Theoretical Foundation," Int. Jour. Patt. Rec. and Art. Intell. **2** (3), 459–483 (1988).

[Schu84a] B. G. Schunck, "The Motion Constraint Equation and Surface Structure," Nat. Conf. Art. Intell., Austin, Tex. (1984).

[Schu84b] B. G. Schunck, "The Motion Constraint Equation for Optical Flow," Int. Conf. Patt. Rec., Montreal, Canada (1984).

[Snyd89] M. A. Snyder, "The Mathematical Foundations of Smoothness Constraints for the Determination of Optical Flow and for Surface Reconstruction II: The Case of Coupled Components," COINS Tech. Rep., UMass, Amherst, in preparation.

[Stra86] G. Strang, *Introduction to Applied Mathematics*, Wellesley–Cambridge Press, Wellesley, Mass. (1986)

[Thor79] J. A. Thorpe, *Elementary Topics in Differential Geometry*, Springer Verlag, New York (1979).

A The Euclidean Group of the Plane ISO(2)

The group of rigid transformations of the 2-plane $\Re^2$ consists of those transformations of $\Re^2$ into itself which preserve the distance between any two points of the plane. As is well known, this group can be represented as a rotation of the plane around an axis perpendicular to the plane, followed by a translation in the plane. Mathematically this group of transformations is the semi-direct product of the two dimensional rotation group SO(2) and the group of translations in the plane. The group is denoted by the symbol ISO(2), which stands for "I(nhomogeneous) SO(2)," also called the *Euclidean group of the plane* E(2).

Since the coordinates of a point in any two Cartesian coordinate systems are related by just such a rigid transformation, it is clear that any transformation by an element of ISO(2) can be described as a change in the Cartesian coordinate system of the plane.

If $\mathbf{r}$ is the column vector $(x, y)^T$, then an element $(\mathbf{R}, \mathbf{t})$ of this group is defined as having the following effect:

$$\mathbf{r} \longrightarrow \mathbf{r}' \equiv \mathbf{R}\mathbf{r} + \mathbf{t}, \tag{76}$$

where

$$\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix};$$

θ is the angle of rotation, and $\mathbf{t}$ is the translation vector. The 2×2 matrix $\mathbf{R}$ can be defined as

$$\mathbf{R} \equiv \mathbf{R}(\theta) = \cos\theta \, \mathbf{1}_2 + \mathbf{J} \sin\theta,$$

where $\mathbf{1}_2$ is the 2×2 identity matrix

$$\mathbf{1}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and $\mathbf{J}$ is the antisymmetric matrix

$$\mathbf{J} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = -\mathbf{J}^T \; ; \; \mathbf{J}^T\mathbf{J} = \mathbf{1}_2.$$

One easily sees that

$$\mathbf{R}^{-1}(\theta) = \mathbf{R}(-\theta) = \mathbf{R}^T(\theta),$$

so that **R** is an orthogonal matrix:

$$\mathbf{R}^T\mathbf{R} = \mathbf{R}\mathbf{R}^T = \mathbf{1}_2.$$

Since $\det \mathbf{R} = \cos^2\theta + \sin^2\theta = 1$, it follows that $\mathbf{R} \epsilon SO(2)$, the group of real orthogonal 2×2 matrices having unit determinant.

It is easy to see that

$$[\mathbf{R}, \mathbf{J}] = [\mathbf{R}, \mathbf{1}_2] = 0,$$

where $[\mathbf{A}, \mathbf{B}] \equiv \mathbf{AB} - \mathbf{BA}$ is called the *commutator* of **A** and **B**. If $[\mathbf{A}, \mathbf{B}] = \mathbf{0}$, then we say that **A** and **B** *commute*. It is easy to see that if a matrix **M** commutes with **R**, then **M** is a linear combination of $\mathbf{1}_2$ and **J**:

$$[\mathbf{M}, \mathbf{R}] = 0 \Longrightarrow \mathbf{M} = a\mathbf{1}_2 + b\mathbf{J}. \tag{77}$$

A.1 Scalars, Vectors, and Tensors

Suppose that we have a function f defined on the image plane (for instance, $f = I$, the grey–level intensity function). Upon making a change in the coordinate system to which f is referred, we would expect that the value the function takes at the same physical point in the image plane would be unchanged. However, the explicit function f that expresses the value of the function at that point will be different. That is, as a function of the position vector $\mathbf{r}$ of the point P, the value of the function f would be $f(\mathbf{r})$, whereas with respect to the new coordinate system the position vector of P would be $\mathbf{r}'$, and the value of the function f at P would be $f'(\mathbf{r}')$. Since the value of the function at the physical point is independent of the coordinate system in which P is represented, we must have that

$$f'(\mathbf{r}') \equiv f(\mathbf{r}). \tag{78}$$

where both $\mathbf{r}$ and $\mathbf{r}'$ represent the same physical point and hence are related by (76). We will say that a quantity f which satisfies (78) when the transformation (76) of ISO(2) is made transforms as a *scalar* under ISO(2). We will often abbreviate this to the statement that "f is a scalar."

We will also denote the fact that a quantity $Q(\mathbf{r})$ transforms, under ISO(2), into the quantity $Q'(\mathbf{r}')$ by the symbol

$$Q \longrightarrow Q',$$

where the arguments $\mathbf{r}$ (of Q) and $\mathbf{r}'$ (of Q') are suppressed.

We now consider the way that a vector $\mathbf{A}$ transforms under ISO(2). Under a change in Cartesian coordinate system, the physical vector attached to a particular physical point in the image (for instance, the optical flow vector) does not change, but its representation in terms of its components does. In the new coordinate system, the vector is represented as $\mathbf{A}'(\mathbf{r}')$, but this vector is also given by the old vector $\mathbf{A}(\mathbf{r})$, rotated by the rotation matrix $\mathbf{R}$: $\mathbf{RA}(\mathbf{r})$. Since both of these expressions represent the same physical vector, we have that:

$$\mathbf{A}'(\mathbf{r}') = \mathbf{RA}(\mathbf{r}).$$

We express this as

$$\mathbf{A} \longrightarrow \mathbf{A}' = \mathbf{RA}, \tag{79}$$

or, written out in components:

$$\mathbf{A}'_i = \mathbf{R}_{ij}\mathbf{A}_j.$$

We will call a quantity which transforms like (79) an "ISO(2) vector," or more simply just a "vector." We note that the position vector $\mathbf{r}$ does not transform like a vector, as one would expect, since $\mathbf{r}$ depends on the choice of origin, i.e., on the choice of a particular Cartesian coordinate system.

A second–rank tensor M is a quantity with two indices i and j, represented by the set of 4 numbers $\{M_{ij}\}$ which transforms in the following way under ISO(2):

$$M \longrightarrow M',$$

where

$$M'_{ij}(\mathbf{r}') = \mathbf{R}_{ik}\mathbf{R}_{j\ell}M_{k\ell}(\mathbf{r}). \tag{80}$$

If we represent the 4 components $\{M_{ij}\}$ of M as a 2×2 matrix $\mathbf{M}$:

$$M_{ij} \equiv \mathbf{M}_{ij},$$

then the transformation law (80) can be written succinctly as

$$\mathbf{M}'_{ij}(\mathbf{r}') = \mathbf{R}_{ik}\mathbf{M}_{k\ell}(\mathbf{r})(\mathbf{R}^T)_{\ell j},$$

or:

$$\mathbf{M}'(\mathbf{r}') = \mathbf{R}\mathbf{M}(\mathbf{r})\mathbf{R}^T.$$

This may be then expressed as the second rank tensor transformation law under ISO(2):

$$\mathbf{M} \longrightarrow \mathbf{R}\mathbf{M}\mathbf{R}^T. \tag{81}$$

We see immediately that such tensors exist. Indeed, if **A** and **B** are vectors,

$$\mathbf{A} \longrightarrow \mathbf{RA} \ ; \ \ \mathbf{B} \longrightarrow \mathbf{RB} \Longrightarrow \mathbf{B}^T \longrightarrow \mathbf{B}^T\mathbf{R}^T,$$

then the outer product $\mathbf{AB}^T$ transforms as a tensor:

$$\mathbf{AB}^T \longrightarrow (\mathbf{RA})(\mathbf{B}^T\mathbf{R}^T) = \mathbf{R}(\mathbf{AB}^T)\mathbf{R}^T.$$

We now list a number of useful results that are easily proven:

- If **A** is a vector, then $\mathbf{A}^T\mathbf{A} = |\vec{A}|^2$ is a scalar.
- The gradient ∇ transforms as a vector (operator). Therefore if f is a scalar, ∇f is a vector.
- The Laplacian $\nabla^T\nabla$ is a scalar operator, and the quantity $\nabla\nabla^T$ is a tensor operator. Therefore $\nabla^T\nabla I$ is a scalar, and $\nabla\nabla^T I$ is a tensor.
- If **M** and **N** are tensors, then so is **MN**.
- Since the matrix **J** commutes with **R**, if **A** is a vector, then so is $\mathbf{JA} \equiv \widetilde{\mathbf{A}}$, called the dual of the vector **A**.
- If **M** is a tensor, then so are **JM**, $\mathbf{MJ}^T$, and $\mathbf{JMJ}^T \equiv \widetilde{\mathbf{M}}$. The latter is called the dual of the tensor **M**.

B Some Theorems on Positive Definiteness

In this appendix we derive a number of results related to positive definiteness.

Consider the ISO(2) scalar $\Theta \equiv \text{tr}\left[\mathbf{\Omega}^T \mathbf{F} \mathbf{\Omega}\right]$, where $\mathbf{\Omega}$ and $\mathbf{F}$ are tensors, as in the text. We showed in the text that $\mathbf{F}$ could, without loss of generality, be assumed to be a symmetric matrix.

Since $\mathbf{F}$ is a symmetric 2×2 real matrix, there exists an orthogonal matrix (i.e., a rotation) $\mathbf{R}_0 \in SO(2)$ that diagonalizes $\mathbf{F}$:

$$\mathbf{R}_0 \mathbf{F} \mathbf{R}_0^T \equiv \mathbf{D} = \begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix} \tag{82}$$

where δ_1 and δ_2 are the eigenvalues of $\mathbf{F}$. If we rotate the coordinate system by $\mathbf{R}_0$, we know that $\text{tr}\left[\mathbf{\Omega}^T \mathbf{F} \mathbf{\Omega}\right]$ remains invariant. This rotation, however, just transforms $\mathbf{F}$ into the diagonal matrix $\mathbf{D}$. We can, therefore, assume that this rotation has been done, so that Θ has the form:

$$\Theta = \text{tr}\left[\mathbf{\Omega}^T \mathbf{D} \mathbf{\Omega}\right]. \tag{83}$$

Now since $\mathbf{U}^T = (u, v)$, it follows that

$$\mathbf{\Omega} = \nabla \mathbf{U}^T = (\nabla u,\ \nabla v), \tag{84}$$

which implies that

$$\mathbf{D\Omega} = \mathbf{D}(\nabla u,\ \nabla v) = (\mathbf{D}\nabla u,\ \mathbf{D}\nabla v). \tag{85}$$

Hence,

$$\text{tr}\ \left[\mathbf{\Omega}^T \mathbf{D} \mathbf{\Omega}\right] = \begin{pmatrix} \nabla^T u \\ \nabla^T v \end{pmatrix} (\mathbf{D}\nabla u,\ \mathbf{D}\nabla v) \tag{86}$$

$$= \text{tr} \begin{pmatrix} \nabla^T u \mathbf{D} \nabla u & * \\ * & \nabla^T v \mathbf{D} \nabla v \end{pmatrix} \tag{87}$$

$$= \nabla^T u \mathbf{D} \nabla u + \nabla^T v \mathbf{D} \nabla v \tag{88}$$

$$= \delta_1(u_x^2 + v_x^2) + \delta_2(u_y^2 + v_y^2), \tag{89}$$

where we have used (82). It follows that Θ is positive definite if and only if the eigenvalues (δ_1 and δ_2) of $\mathbf{F}$ are positive definite.

We now consider when this situation occurs.

Since $\mathbf{F}$ is a symmetric matrix, it can be written in the form

$$\mathbf{F} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}. \tag{90}$$

It is elementary to show that the eigenvalues of $\mathbf{F}$ are given by the solutions $\lambda_\pm$, where

$$2\lambda_\pm = a + c \pm \sqrt{(a-c)^2 + 4b^2}.$$

Both values for λ (i.e., both eigenvalues of $\mathbf{F}$) are positive definite if and only if $\operatorname{tr}\mathbf{F} = a + c \geq 0$, and

$$a + c \geq \sqrt{(a-c)^2 + 4b^2} \Longrightarrow 4\det\mathbf{F} = 4(ac - b^2) \geq 0.$$

Hence we have proved the theorem

Theorem B.1 *The ISO(2) scalar* $\Theta = \operatorname{tr}\left[\boldsymbol{\Omega}^T\mathbf{F}\boldsymbol{\Omega}\right]$, *where* $\boldsymbol{\Omega}$ *and* $\mathbf{F}$ *are tensors, is positive definite if and only if* $\operatorname{tr}\mathbf{F}$ *and* $\det\mathbf{F}$ *are positive definite.*

We then have the important corollaries:

Corollary B.1.1 *If* $\mathbf{F}$ *is a real, traceless symmetric* 2×2 *matrix different from zero, then* $\operatorname{tr}\left[\boldsymbol{\Omega}^T\mathbf{F}\boldsymbol{\Omega}\right]$ *is* not *positive definite.*

Proof: If $\mathbf{F}$ is traceless and symmetric, then $\mathbf{F}$ is of the form (90), where $c = -a$. This implies that $\det\mathbf{F} = -(a^2 + b^2)$. Since a and b are real numbers not both zero, by assumption, it follows that $\det \mathbf{F} < 0$. The corollary then follows immediately from the theorem.

Corollary B.1.2 *If* $\mathbf{F} = \mathbf{M}^2$, *where* $\mathbf{M}$ *is a symmetric real* 2×2 *matrix different from zero, then* $\mathrm{tr}\left[\mathbf{\Omega}^T \mathbf{F} \mathbf{\Omega}\right]$ *is positive and non-zero.*

Proof: If $\mathbf{F} = \mathbf{M}^2$, then $\mathbf{F}_{ij} = \mathbf{M}_{ik}\mathbf{M}_{kj} \equiv \mathbf{M}_{ik}\mathbf{M}_{jk}$, since $\mathbf{M}$ is symmetric. Hence, $\mathrm{tr}\,\mathbf{F} \equiv \mathbf{F}_{ii} = \mathbf{M}_{ik}\mathbf{M}_{ik}$. The right hand side of the latter equation is the sum of the squares of all the matrix elements of $\mathbf{M}$. But by assumption, not all of these matrix elements are zero. Hence, $\mathrm{tr}\,\mathbf{F} > 0$. The corollary then follows immediately from the theorem.

C Some Completeness Results

In this final appendix, we prove a number of "completeness" theorems.

Before beginning the statement and proof of the theorems, we quote a number of results which are easily proven, and whose proof is omitted.

Result # 1: If $\nabla I = (I_x, I_y)^T$ is transformed by $\mathbf{R}(\theta)$, then the quantities

$$\alpha = I_x^2 + I_y^2 \tag{91}$$

$$\beta = I_x^2 - I_y^2 \tag{92}$$

$$\gamma = 2 I_x I_y \tag{93}$$

are a basis for the space spanned by the products $\{I_i I_j;\ i,j = x,y\}$, and transform under ISO(2)) like

$$\alpha \longrightarrow \alpha \tag{94}$$

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} \longrightarrow \mathbf{R}^2 \begin{pmatrix} \beta \\ \gamma \end{pmatrix}. \tag{95}$$

Result # 2: If we define

$$\alpha = I_{xx} + I_{yy} = \nabla^2 I \tag{96}$$

$$\beta = I_{xx} - I_{yy} \tag{97}$$

$$\gamma = 2I_{xy}, \tag{98}$$

then the set of quantities

$$\{\alpha^2,\ \frac{\alpha^2+\beta^2+\gamma^2}{2},\ \alpha\beta,\ \alpha\gamma,\ \beta^2-\gamma^2,\ 2\beta\gamma\} \tag{99}$$

form a basis for the space consisting of all possible products $\{I_i I_j;\ i,j = xx, xy, yy\}$ of second derivatives of I and transform as follows under an ISO(2) transformation:

$$\begin{pmatrix} \alpha^2 \\ \frac{\alpha^2+\beta^2+\gamma^2}{2} \end{pmatrix} \longrightarrow \begin{pmatrix} \alpha^2 \\ \frac{\alpha^2+\beta^2+\gamma^2}{2} \end{pmatrix} \tag{100}$$

$$\begin{pmatrix} \alpha\beta \\ \alpha\gamma \end{pmatrix} \longrightarrow \mathbf{R}^2 \begin{pmatrix} \alpha\beta \\ \alpha\gamma \end{pmatrix} \tag{101}$$

$$\begin{pmatrix} \beta^2-\gamma^2 \\ 2\beta\gamma \end{pmatrix} \longrightarrow \mathbf{R}^4 \begin{pmatrix} \beta^2-\gamma^2 \\ 2\beta\gamma \end{pmatrix}, \tag{102}$$

where we recall that $\mathbf{R}^n(\theta) = \mathbf{R}(n\theta)$.

Result # 3: For any matrix $\mathbf{F}$,

$$2\mathbf{RFR}^T = \left[\mathbf{F} + \tilde{\mathbf{F}}\right] + \cos 2\theta \left[\mathbf{F} - \tilde{\mathbf{F}}\right] + \sin 2\theta [\mathbf{J}, \mathbf{F}].$$

(This is easily shown using the expression $\mathbf{R} = \cos\theta \mathbf{1}_2 + \mathbf{J}\sin\theta$.)

We now go on to prove our "completeness" theorems.

Theorem C.1 *If* $\mathbf{A}$ *is a vector linear in* 1^{st} *derivatives of the scalar function* I*, then*

$$\mathbf{A} = a\nabla I + b\widetilde{\nabla} I,$$

where a *and* b *are ISO(2) scalars.*

Proof: If $\mathbf{A}$ is linear in 1^{st} derivatives of I, then $\mathbf{A}_i$ can be written as

$$\mathbf{A}_i = \mathbf{B}_{ij}\partial_j I,$$

where the $\mathbf{B}_{ij}$ are numbers. We can then write this as a matrix equation:

$$\mathbf{A} = \mathbf{B}\nabla I.$$

But $\mathbf{A}$ is, by assumption, a vector. Hence $\mathbf{A} \longrightarrow \mathbf{RA} \equiv \mathbf{RB}\nabla I$ under an ISO(2) transformation. But under the same transformation, $\nabla I \longrightarrow \mathbf{R}\nabla I$, which implies that $\mathbf{B}\nabla I \longrightarrow \mathbf{BR}\nabla I$. We conclude that

$$\mathbf{RB}\nabla I = \mathbf{BR}\nabla I \Longrightarrow [\mathbf{R},\mathbf{B}]\nabla I = 0.$$

But ∇I is arbitrary. Hence, it follows that $[\mathbf{R},\mathbf{B}] = 0$.

Therefore, $\mathbf{B} = a\mathbf{1}_2 + b\mathbf{J}$. Consequently, $\mathbf{A} = (a\mathbf{1}_2 + b\mathbf{J})\nabla I \equiv a\nabla I + b\widetilde{\nabla} I$, Q.E.D.

Theorem C.2 *If* $\mathbf{H}$ *is a tensor quadratic in* 1^{st} *order derivatives of* I*, then*

$$\mathbf{H} = a_0\mathbf{K} + a_1\mathbf{JK} + a_2\mathbf{KJ}^T + a_3\widetilde{\mathbf{K}},$$

where the tensors $\mathbf{K},\ldots$ *are defined in Section 3, and where the* a_i *are constants.*

Proof: From Result # 1, we see that the general tensor $\mathbf{H}$ can be written as

$$\mathbf{H} = \mathbf{M}_1\alpha + \mathbf{M}_2\beta + \mathbf{M}_3\gamma,$$

where the $\mathbf{M}_i$ are (constant) matrices. When an ISO(2) transformation is made, we know that $\mathbf{H} \longrightarrow \mathbf{RHR}^T$. On the other hand, α, β, and γ transform according to Result # 1. These two ways of looking at the transformation of $\mathbf{H}$ must be equivalent. Hence,

$$\mathbf{RM}_1\mathbf{R}^T\alpha \; + \; \mathbf{RM}_2\mathbf{R}^T\beta + \mathbf{RM}_3\mathbf{R}^T\gamma = \tag{103}$$

$$= \; \mathbf{M}_1\alpha + \mathbf{M}_2(\beta\cos 2\theta + \gamma\sin 2\theta) + \mathbf{M}_3(-\beta\sin 2\theta + \gamma\cos 2\theta) \tag{104}$$

We immediately conclude, from the independence of the set $\{\alpha, \beta, \gamma\}$, that

$$\mathbf{RM}_1\mathbf{R}^T = \mathbf{M}_1 \Longrightarrow [\mathbf{R}, \mathbf{M}_1] = 0,$$

which implies that

$$\mathbf{M}_1 = a\mathbf{1}_2 + b\mathbf{J}.$$

We then use Result # 3 to reexpress the left-hand side of (103)—(104), equate Fourier coefficients on either side of the equation, and finally equate coefficients of α, β, and γ. The result is the following independent set of equations:

$$\mathbf{M}_2 \; = \; \mathbf{JM}_2\mathbf{J} \Longrightarrow \{\mathbf{J}, \mathbf{M}_2\} = 0, \tag{105}$$

$$\mathbf{M}_3 \; = \; \mathbf{JM}_3\mathbf{J} \Longrightarrow \{\mathbf{J}, \mathbf{M}_3\} = 0, \tag{106}$$

$$[\mathbf{J}, \mathbf{M}_3] \; = \; 2\mathbf{M}_2, \tag{107}$$

where we have defined the anticommutator $\{\mathbf{A}, \mathbf{B}\} \equiv \mathbf{AB} + \mathbf{BA}$. It is then easy to check that these equations imply that

$$\mathbf{M}_2 = \begin{pmatrix} d & c \\ c & -d \end{pmatrix} \; ; \; \mathbf{M}_3 = \begin{pmatrix} -c & d \\ d & c \end{pmatrix}.$$

The general tensor $\mathbf{H}$ is then of the form:

$$\mathbf{H} = \begin{pmatrix} a\alpha + d\beta - c\gamma & b\alpha + c\beta + d\gamma \\ -b\alpha + c\beta + d\gamma & a\alpha - d\beta + c\gamma \end{pmatrix} \tag{108}$$

$$= \begin{pmatrix} (a+d)I_x^2 + (a-d)I_y^2 - 2cI_xI_y & (b+c)I_x^2 + (b-c)I_y^2 + 2dI_xI_y \\ (-b+c)I_x^2 - (b+c)I_y^2 + 2dI_xI_y & (a-d)I_x^2 + (a+d)I_y^2 + 2cI_xI_y \end{pmatrix}. \tag{109}$$

By looking at the explicit form of $\mathbf{K}, \ldots$, one sees that

$$\mathbf{H} = (a+d)\mathbf{K} + (b-c)\mathbf{JK} + (-b-c)\mathbf{KJ}^T + (a-d)\widetilde{\mathbf{K}}.$$

This proves the theorem.

We now note that if, instead, we were to demand that $\mathbf{H}$ be a scalar, rather than a tensor, all the properties of $\mathbf{H}$ that were used in the theorem would still hold, except that the matrices $\mathbf{M}_i$ would now be only linear combinations of the matrices $\mathbf{1}_2$ and $\mathbf{J}$. The theorem then implies that $\mathbf{M}_2 = \mathbf{M}_3 = 0$, and hence that if $\mathbf{H}$ is an ISO(2) scalar times a linear combination of the matrices $\mathbf{1}_2$ and $\mathbf{J}$, $\mathbf{H}$ must be of the form $\mathbf{H} = (a\mathbf{1}_2 + b\mathbf{J})\alpha$. We therefore have the corollary:

Corollary C.2.1 *If H is an ISO(2) scalar which is quadratic in 1st derivatives of the image intensity I, then H is a multiple of $I_x^2 + I_y^2$.*

This is identical to the result (Proposition 2) of Brady and Horn [Brad83], but with "rotationally symmetric" replaced by the more general "ISO(2) invariant."

Lemma *If $\mathbf{H}$ is a tensor quadratic in 2nd order derivatives of I, then*

$$\mathbf{H} = a_0\mathbf{L}^2 + a_1\mathbf{JL}^2 + a_2\mathbf{L}^2\mathbf{J}^T + a_3\widetilde{\mathbf{L}}^2 + (a_4\mathbf{1}_2 + a_5\mathbf{J})\Delta + (a_6\mathbf{1}_2 + a_7\mathbf{J})Q,$$

where $\mathbf{L}, \ldots.$ are defined in Section 3, and the quantities Δ and Q are defined to be:

$$\Delta = (I_{xx} + I_{yy})^2 = (\nabla^2 I)^2, \tag{110}$$

$$Q = I_{xx}^2 + 2I_{xy}^2 + I_{yy}^2. \tag{111}$$

We note that the quantities Δ and Q have been called the "squared Laplacian" and the "quadratic variation," respectively, by Grimson in his work on surface reconstruction [Grim81].

Proof: Since **H** is, by assumption, quadratic in 2nd order derivatives of I, we use Result # 2 to write **H** in the form:

$$\mathbf{H} = \mathbf{M}_1 \alpha^2 + \mathbf{M}_2 \left(\frac{\alpha^2 + \beta^2 + \gamma^2}{2}\right) + \mathbf{M}_3 \alpha\beta + \mathbf{M}_4 \alpha\gamma + \mathbf{M}_5 (\beta^2 - \gamma^2) + \mathbf{M}_6(2\beta\gamma),$$

where the $\mathbf{M}_i$ are matrices. Proceeding in a way identical to that for Theorem C.2, we make an ISO(2) transformation on **H**, and express this transformed **H** in two ways, one by transforming the matrices $\mathbf{M}_i$ to $\mathbf{R}\mathbf{M}_i\mathbf{R}^T$, and one by transforming the variables $\alpha^2, \ldots$. We equate these two expressions for the transformed **H**, use Result # 3 as before, and then equate Fourier coefficients on either side of the equation. We then equate coefficients of the independent polynomials (99). Finally, we arrive at the set of independent equations which must be satisfied if **H** is to transform as a tensor:

$$\mathbf{M}_5 = 0, \tag{112}$$

$$\mathbf{M}_6 = 0, \tag{113}$$

$$\mathbf{M}_1 = -\mathbf{J}\mathbf{M}_1\mathbf{J} \Longrightarrow \mathbf{M}_1 = a\mathbf{1}_2 + b\mathbf{J}, \tag{114}$$

$$\mathbf{M}_2 = -\mathbf{J}\mathbf{M}_2\mathbf{J} \Longrightarrow \mathbf{M}_2 = c\mathbf{1}_2 + d\mathbf{J}, \tag{115}$$

$$\mathbf{M}_3 = \mathbf{J}\mathbf{M}_3\mathbf{J}, \tag{116}$$

$$2\mathbf{M}_4 = -[\mathbf{J}, \mathbf{M}_3]. \tag{117}$$

The latter two equations taken together imply that

$$\mathbf{M}_3 = \begin{pmatrix} \nu & \mu \\ \mu & -\nu \end{pmatrix}; \quad \mathbf{M}_4 = \begin{pmatrix} -\mu & \nu \\ \nu & \mu \end{pmatrix}.$$

We therefore have found that $\mathbf{H}$ can be expressed in the form

$$\mathbf{H} = (a\mathbf{1}_2 + b\mathbf{J})\,\alpha^2 + (c\mathbf{1}_2 + d\mathbf{J})\left(\frac{\alpha^2 + \beta^2 + \gamma^2}{2}\right) + \mathbf{M}_3\,\alpha\beta + \mathbf{M}_4\,\alpha\gamma, \qquad (118)$$

where $\mathbf{M}_3$ and $\mathbf{M}_4$ are as given above.

Using the explicit forms for the tensors $\mathbf{L}^2, \ldots$ one easily checks that the quantity

$$\mathbf{H} = a_0\mathbf{L}^2 + a_1\mathbf{J}\mathbf{L}^2 + a_2\mathbf{L}^2\mathbf{J}^T + a_3\tilde{\mathbf{L}}^2 + (a_4\mathbf{1}_2 + a_5\mathbf{J})\Delta + (a_6\mathbf{1}_2 + a_7\mathbf{J})Q,$$

is equal to the quantity (118) where a_0, a_1, a_2, and a_3 are arbitrary numbers subject only to the conditions:

$$a_0 - a_3 = 2\nu, \qquad (119)$$

$$a_1 + a_2 = -2\mu, \qquad (120)$$

and a_4, a_5, a_6, and a_7 are given by:

$$a_4 = a, \qquad (121)$$

$$a_5 = b, \qquad (122)$$

$$a_6 = c - \frac{a_0 + a_3}{2}, \qquad (123)$$

$$a_7 = d - \frac{a_1 - a_2}{2}. \qquad (124)$$

We then have the following Theorem

Theorem C.3: *If* $\Theta = \mathrm{tr}\left[\mathbf{\Omega}^T \mathbf{F} \mathbf{\Omega}\right]$ *is a tensor-based ISO(2) scalar, then* $\mathbf{F}$ *is of the form:*

$$\mathbf{F} = a_0 \mathbf{L}^2 + a_1 \mathbf{J}\mathbf{L}^2 + a_2 \mathbf{L}^2 \mathbf{J}^T + a_3 \tilde{\mathbf{L}}^2, \tag{125}$$

and if Θ *is a scalar-based ISO(2) scalar,* $\mathbf{F}$ *is of the form*

$$\mathbf{F} = (a\mathbf{1}_2 + b\mathbf{J})\Delta + (c\mathbf{1}_2 + d\mathbf{J})\,Q \tag{126}$$

Proof: From the Lemma, the general tensor satisfying the conditions of the Lemma is of the form (118). We showed, in the Lemma, that the difference between this tensor and a tensor of the form (125) is just a matrix of the form (126). But a matrix of the latter form transforms as a scalar under an ISO(2) transformation, as discussed in Section 2.2. Hence, we have proved the theorem.

We now show that Proposition 6 of the paper by Brady and Horn [Brad83], generalized to ISO(2) invariance, is an immediate corollary of this theorem:

Corollary C.3.1: *If* H *is an ISO(2) scalar quantity which is quadratic in second derivatives of an ISO(2) scalar quantity* I*, then* H *is of the form*

$$H = a_0 (I_{xx} + I_{yy})^2 + a_1 (I_{xx}^2 + 2I_{xy}^2 + I_{yy}^2),$$

where a_0 *and* a_1 *are constants.*

Proof: We see from the theorem that Δ and Q are necessarily ISO(2) scalars. **Q.E.D.**