Measurement of Motion-Induced Image Deformations: Spatio-Temporal Operators for Translation, Divergence, Curl, and Shear

Roy Eagleson

Centre for Cognitive Science
The University of Western Ontario,
(rm SS7332) London, Ontario, Canada, N6A 5C2
elroy@uwo.ca, elroy@julian.uucp, elroy@uwovax.bitnet

KEYWORDS: Motion Analysis, Lie Transformation Group, Spatio-Temporal Operators, Group Theoretic Approach.

ABSTRACT

Many approaches to motion understanding in computer vision are based first on computing the 'optical flow' of a set of time-varying images. Ideally, this vector field corresponds to the 2D translations of local regions of the intensity profile which shift as a result of the 3D motions in the visible world. If this vector field could be computed, it is argued that is would provide an intermediate representation that could be used as input data to higher-level motion analysis algorithms. However; this is an ill-posed problem, due to the underconstrained nature of the projection from 3D to 2D. Many researchers have shown (cf. Koenderink and van Doorn[1], Longuet-Higgins and Prazdny[2], Kanatani[3],[4]) that it is *in principle* possible to estimate the local motion and low-level structure of a visible surface if additional information is added by obtaining higher-order spatio-temporal derivatives of the changing image signal. This information must be combined with additional knowledge about rigid bodies and the projection from 3D to 2D.

The six degrees-of-freedom of the Euclidean rigid-body motion group [Given by the semi-direct product of 3D translations and rotations, (cf. Carlton and Shepard[5]), $E^+ = \Re^3 \circledS SO(3)$] project perspectively to a unique six-dimensional vector field group. (cf. Blicher and Omohundro[6]). It is proposed that the structure and motion of a visible surface can be computed by integrating these infinitesimal deformations over space and time; incorporating the objects' spatio-temporal continuity properties. However; this is not a "structure-from-motion" paper. Rather; it deals with the estimation of a six-element group of local coordinate transformations in the image. The group is proposed as a set of measures for computing structure and relative motion. Despite this formulation, the problem would remain ill-posed when implemented using spatio-temporal derivatives, due to the instability of gradient operators with respect to added noise. Because of this, *six* orthogonal spatio-temporal *filters* are proposed. They extract the translation, divergence, curl, and 'shear', which are independent deformations that can be measured locally in the image. More importantly, they cover the six degrees-of-freedom of motion in the 3D scene.

[1]This research was supported in part by NSERC operating grant number A2600 awarded to Dr. Zenon Pylyshyn.

1 Introduction

"Motion Understanding" in the field of Computer Vision requires that sequences of 2D images be analysed for changes that result from the relative 3D motions of objects with respect to the observer. In order to implement this process on a machine vision system, it must be described with mathematical and computational rigour. Due to the ill-posed nature of this 'inverse optics' problem, there are two major hurdles; the first being the measurement of these motion-induced image deformations, and the second being the interpretation of this vector field.

Many approaches to motion understanding in computer vision are based first on computing the 'optical flow' of a set of time-varying images. Classically, this vector field corresponds to the 2D translations of local regions of the intensity profile which shift as a result of the 3D motions in the visible world. If this vector field could be computed, it is argued that is would provide an intermediate representation that could be used as input data to higher-level motion analysis algorithms. This paper departs from this approach essentially by proposing that the 'optical flow' field can be regarded as a 2D field of six-dimensional vectors, rather than a 2D field of two-dimensional vectors.

1.1 Ill-posed Nature of Inverse Optics

Perceiving the 3D structure of the world from a 2D image can be described as solving an inverse problem. Since the projection removes one dimension (depth), solving for the inverse is inherently underconstrained. Additional information must be added to constrain the missing degree of freedom. The inverse problem of finding the structure of an object given only its projected motion in the image is an ill-posed problem unless other constraints or information is added. The requirements for a problem to be well-posed are that; (1) A solution exists, (2) The solution is unique, and (3) There are no discontinuities in the relationship between the solution result and the amount of added input noise. ("Graceful degradation")

There are many cues to structure and environment layout that can be used by a vision system. A system which uses two spatially-displaced images of the same scene can recover depth from stereopsis. In addition, there are monocular cues to the shape of an object from its shading in the image. In fact, Cavanagh[7] shows evidence that stereopsis, motion, texture, colour, and 'form' perception all take part in complex interactions to determine 3D structure. The motion parallax field contains information about structure that requires relatively few assumptions for its calculation, as compared for example with "shape from shading" algorithms which require assumptions about lighting and reflectance properties. As many researchers have indicated, the 3D structure of a rigid object can be determined in some cases from the deformations which occur in its projected 2D image alone.

1.2 Constraints for Solution of Inverse

Two of the most common sources for adding information to the structure from motion problem are the assumption that a surface is locally rigid, and that the geometry of the projection from 3D to 2D is known. Nakayama[8] asserts that the information from the motion parallax field, plus information about translational velocity or absolute distance to one object point is sufficient to fix absolute distances (ie. environment layout) with certainty. In addition, Koenderink and van Doorn[1] note that the motion parallax field "contains rich information as to slant of surfaces", from which the structure of individual objects can be estimated. As has been stated earlier, a major problem in vision is the fact that the 2D data from the world severely underconstrains the solution to the possible scenes in 3D that could have resulted in the same projection. As a result, additional data or constraints are needed. One that is often imposed is the "rigidity assumption[9][page 516.]."

Helmholtz[10] and Gibson[11] were perhaps the first to suggest that we may be able to extract information about the structure of the world from changes in the field of optical flow. In fact, Gibson suggested that the spatial gradient of the velocity field might be an invariant property of the image of an object. Since the velocity vectors are two-dimensional, there will be four spatial derivatives of 2D velocity: u_x, u_y, v_x, and v_y.

Kanatani[3] showed that the nine "structure and motion parameters" can be extracted using a technique similar to that shown by Longuet-Higgins and Prazdny[2], and Longuet-Higgins[12], by adding second order spatial derivatives of optical flow. He calculates the 3D translational velocities of a surface patch, the 3D components of rotation, ω_1, ω_2, and ω_3, as well as the gradient of the surface, p and q. In this paper, he refers to absolute distance, r, as an indeterminate parameter (although Kanatani points out that r is not special in this sense, and could be calculated if one of the other parameters were fixed or used as the indeterminate parameter).

To extract the "structure and motion parameters", Kanatani requires approximations to the following optical flow parameters;

$$(u_0, v_0), \quad (u, v), \quad \frac{\partial u}{\partial x}, \frac{\partial v}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial v}{\partial y}, \frac{\partial^2 u}{\partial x \partial y} \text{ and } \frac{\partial^2 v}{\partial x \partial y}$$

> "The flow equations are quadratic in (x, y) for planar surfaces. For general curved surfaces, however, the flow equations are expressed in an infinite Taylor series (Waxman and Ullman[13]). For curved surfaces, see also Subbarao and Waxman[14], and Subbarao[15]." - Kanatani[3]

1.2.1 Perceptual Invariants

Koenderink and Van Doorn[1] showed that certain combinations of the gradients of the optical flow would be invariant properties of the image of a surface. They note that the deformation term remains essentially invariant in their first-order expansion of the flow field:

$$\text{Grad } V = \text{Curl} + \text{Div} + \text{Def}$$

Longuet-Higgins and Prazdny[2] attack the problem of computing the 3D translation and rotation of an observed environment from the optical flow field alone.

> "The problem is to determine the translational and rotational motion of a given object, and the gradient of its surface at any point (Marr, 1976), from the optic flow field due to the nearby texture elements. It is shown that all these unknowns may be computed from the field and its first and second spatial derivatives at the corresponding point on the retina."[1]

They join with Koenderink and Van Doorn[1] in pointing out that there are four combinations of the optical flow gradients that can be given the following meanings as shown in figure 1. (After Longuet-Higgins and Prazdny[2], p.395.)

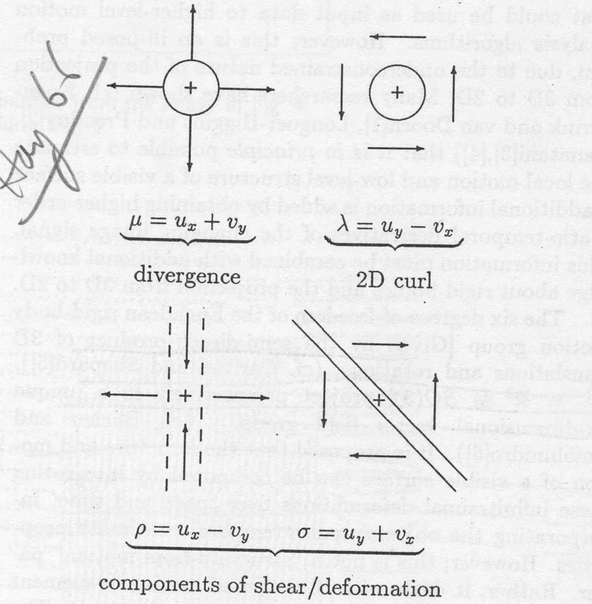

Figure 1: Combinations of spatial derivatives of optical flow

The divergence and curl ("dilation and vorticity") remain invariant under image rotation, while the two components of shear combine to form one oriented quantity, $(\rho \cos 2\theta + \sigma \sin 2\theta)$. They also note that the second spatial derivatives may provide a means by which a system can check the local rigidity assumption, and point to the possibility of finding receptors which may respond to "local

[1] In fact, they note that "the observer gains certain computational advantages from tracking with his eye any surface whose gradient and relative motion are of special interest to him; this is an intuitively reassuring result."

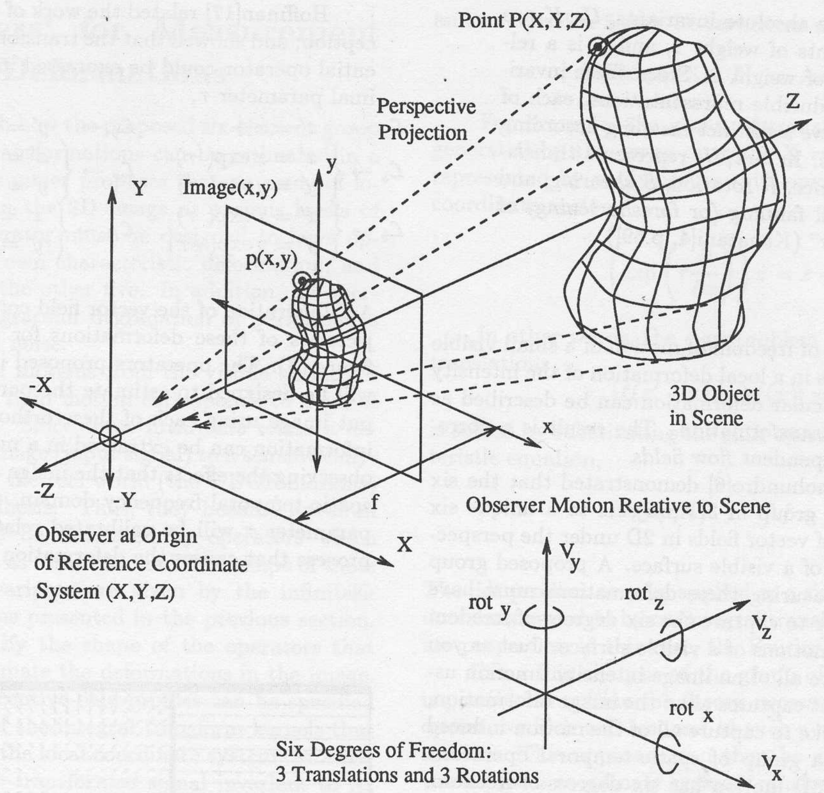

Figure 2: Projection from 3D Scene to 2D Image

deformations of the optic flow field," as well as dilation and vorticity.

Kanatani[4] underscores the importance and benefits of posing this problem in terms of its geometrical invariants. He describes this as exploiting the "underlying geometrical structure" of the perspective projection and object motions. Kanatani first describes a formal mathematical representation of the motion of the object in 3D. Then, according to Weyl's thesis, it is desirable to find an irreducible representation with respect to this transformation group.

2 Group Theoretic Analysis

The perspective projection of a point (X, Y, Z) from 3D to the 2D image point (x, y) is described by similar triangles:

$$xZ = fX \quad \text{and} \quad yZ = fY$$

where f is the characteristic focal length associated with the imaging system. There are six degrees of freedom of relative motion between the observer (here located at the origin of the coordinate system) and the 3D point. It has been shown (See Waxman and Sinha[16]) that the 3D trajectory of a point projects to a 2D image vector in terms of the depth Z and the six motion parameters, $V_X, V_Y, V_Z, \Omega_X, \Omega_Y$, and Ω_Z, assuming they are small ('differential') motions;

$$v_x = \frac{xV_Z - fV_X}{Z} + xy\Omega_X - (1+x^2)\Omega_Y + y\Omega_Z \qquad (1)$$

$$v_y = \frac{yV_Z - fV_Y}{Z} + (1+y^2)\Omega_X - xy\Omega_Y - x\Omega_Z$$

Using a slightly different notation and focal point, Kanatani[4] derives the equations for optical flow as,

$$u(x,y) = u_0 + Ax + By + (Ex + Fy)x \qquad (2)$$

$$v(x,y) = v_0 + Cx + Dy + (Ex + Fy)y$$

where the coefficients are given in terms of the structure-and-motion parameters of a planar surface as,

$$u_0 = \frac{fa}{f+r}, \qquad v_0 = \frac{fb}{f+r}$$

$$A = p\Omega_Y - \frac{pa+c}{f+r}, \qquad B = q\Omega_Y - \Omega_Z - \frac{qa}{f+r}$$

$$C = -p\Omega_X + \Omega_Z - \frac{pb}{f+r}, \qquad D = -q\Omega_X - \frac{qb+c}{f+r}$$

$$E = \frac{1}{f}\left(\Omega_Y + \frac{pc}{f+r}\right), \qquad F = \frac{1}{f}\left(-\Omega_X + \frac{qc}{f+r}\right)$$

Kanatani demonstrates that the invariants of the perspective transformation from a moving 3D surface to a 2D image are given by the algebraic combinations;

$$U_0 = u_0 + iv_0, \qquad T = A + D, \qquad R = C - B,$$

$$S = (A - D) + i(B + C), \qquad K = E + iF.$$

"where T, R are absolute invariants, U_0, K are relative invariants of weight 1, and S is a relative invariant of weight 2. Since these invariants define irreducible representations, each of them should have a distinct meaning according to Weyl's thesis. In fact, U_0 represents *translation, T divergence, R rotation, S shearing,* and K what we call *fanning (or foreshortening)* of the optical flow" (Kanatani[4, p.59])

The six degrees of freedom of motion of a small visible surface region results in a local deformation of the intensity function. This particular deformation can be described as a *local coordinate transformation.* The result is a corresponding set of independent *flow fields.*

Blicher and Omohundro[6] demonstrated that the six dimensional motion group in 3D projects to a unique six dimensional group of vector fields in 2D under the perspective transformation of a visible surface. A proposed group of operators for measuring these deformations must have six elements in order to capture the six degrees-of-freedom of the group of 3D motions of a visible surface. Just as you would like to capture all of an image intensity function using a basis space that captures all of the image information, so you should also like to capture all of the motion-induced deformations with a group of spatio-temporal operators. Since the group of 3D motion has six degrees-of-freedom, the group of 'basis deformations' should have six elements.

3 Lie Algebra of 2D Vector Fields

The perceptual invariants described here for the motion group take the form of localised 2D image deformations. It is necessary to describe the *representations* of these flow fields. A six-element group of 2D deformations, forming a "Lie Algebra" has been proposed by many researchers for encoding perceptual invariants (cf. Hoffman[17], Dodwell[18], Caelli and Ferraro[19]). The algebra has two common representations; as a group of differential operators, or a set of localised coordinate transformations.

3.1 Differential Form and Infinitesimal Trajectories

$$\mathcal{L}_x = \frac{\partial}{\partial x}, \quad \mathcal{L}_s = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}, \quad \mathcal{L}_b = x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y} \qquad (3)$$

$$\mathcal{L}_y = \frac{\partial}{\partial y}, \quad \mathcal{L}_r = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}, \quad \mathcal{L}_B = y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}$$

This set of six operators forms an algebraic group, which can be referred to as the horizontal and vertical translation ($\mathcal{L}_x, \mathcal{L}_y$), the divergence ($\mathcal{L}_s$), the rotation($\mathcal{L}_r$), and two components of 'shear' ($\mathcal{L}_b, \mathcal{L}_B$), respectively. An additional operator can be included to extract the first temporal derivative, $\mathcal{L}_t = \frac{\partial}{\partial t}$.

Hoffman[17] related the work of Sophus Lie[20] to perception; and showed that the transformations of each differential operator could be expressed in terms of an infinitesimal parameter τ,

$$\mathcal{L}_x \to \quad x' = x + \tau \qquad \mathcal{L}_y \to \quad y' = y + \tau \qquad (4)$$

$$\mathcal{L}_s \to \begin{cases} x' = x\exp(\tau) \\ y' = y\exp(\tau) \end{cases} \quad \mathcal{L}_r \to \begin{cases} x' = x\cos(\tau) - y\sin(\tau) \\ y' = x\sin(\tau) + y\cos(\tau) \end{cases}$$

$$\mathcal{L}_b \to \begin{cases} x' = x\exp(\tau) \\ y' = y\exp(-\tau) \end{cases} \quad \mathcal{L}_B \to \begin{cases} x' = x\cos(\tau) - y\sin(\tau) \\ y' = x\sin(-\tau) + y\cos(-\tau) \end{cases}$$

An illustration of the vector field corresponding to the trajectories of these deformations for small τ are shown in figure (3). The operators proposed in this implementation will be designed to estimate the parameter τ from the input image using each of these orthogonal operators. This information can be extracted in a multiresolution sense by observing the effects that the image deformation has in the spatio-temporal frequency domain. (In a later section, the parameter τ will be calibrated relative to the 3D motion process that causes the deformation).

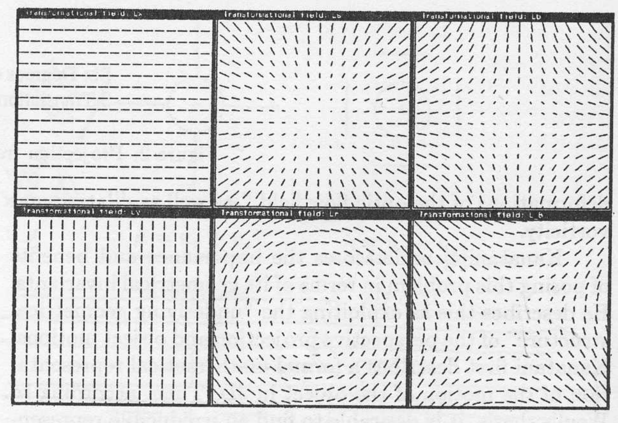

Figure 3: Trajectories of Group{ $\mathcal{L}_x$, $\mathcal{L}_y$; $\mathcal{L}_s$, $\mathcal{L}_r$; $\mathcal{L}_b$, $\mathcal{L}_B$ }

A group of orthogonal 2D deformations has been proposed to represent, or 'account for', the set of invariants of the local coordinate transformations in an image due to the motion in 3D of visible object surfaces. This group structure will dictate the form of the operators that will make this measurement in the image. It is therefore necessary to design a mechanism for extracting these deformations reliably in the presence of noise. The approach adopted in this paper is based on the spatio-temporal properties of certain complex-domain operators. They measure motion which can be viewed in the intensity function as local *signal phase* behaviour. (cf. Fleet[20]). However; instead of measuring the phase shifts of operators tuned for the $\mathcal{L}_x$ and $\mathcal{L}_y$ differential operators, this paper extends to include the orthogonal solutions of the invariants of the differential operators, $\mathcal{L}_s$, $\mathcal{L}_r$, $\mathcal{L}_b$, and $\mathcal{L}_B$.

4 'Eigenfilters' for Measurement of Image Deformations

This section deals with how the proposed six-element group of local coordinate transformations can be estimated in a textured image. This paper proposes that an array of localised operators span the 2D image at various levels of resolution. Each operator must be designed to have optimal response to its own characteristic deformation, and minimal response to the other five. In addition, all operators should exhibit graceful degradation of performance with respect to added noise.

The projection of motions from the visible surface of an object undergoing rigid motion corresponds to a set of perceptual *invariants*. Since these invariants relate a detectable event in the image (input signal) to a semantically-interpretable event in the real world (the 6 DOF motions) according to Weyl's thesis. Thus, they become low-level 'features' that should be estimated by operators which Pylyshyn[22] classifies as "transducers." The shape of these particular motion invariants are given by the infinitesimal trajectory patterns presented in the previous section. Essentially, they specify the shape of the operators that should be used to estimate the deformations in the image. More formally, the receptive-field profiles can be specified as having the shape of the integral transform kernels that are *characteristics*[2] of the local coordinate system deformations. They leave the *transformed* signal invariant to its original pattern. In other words, we are forming the characteristic equation,

$$|f' - \lambda f| = 0$$

and solving the eigenproblem of f' with respect to f (cf. Chien[23, p.22]). Let the eigenspaces be A and the eigenvalues be λ, ie.

$$f'A = \lambda f A$$

If we replace the right-to-left operation of the matrix A with an operator $\mathcal{A}\{\cdot\}$ on the coordinate vector of the function f, we are solving the eigenvalue problem,

$$\mathcal{A}\{f'\} \ = \ \lambda \mathcal{A}\{f\} \qquad (5)$$

(Where $\mathcal{A}$ is an integral transform operator acting on the transformed function f', and λ represents the local coordinate transformation associated with that particular convolution kernel, forming and 'eigen-filter' expression).

4.1 Estimating One-Parameter Coordinate Transformations

'Group-theoretic' approaches (Such as in Olver[24], Ovsiannikov[25], or Belinfante and Kolman[26]) to the analysis of the differential equations of image motion give the following 'exponential map' representation of a local one-parameter coordinate system transformation, $x' = \delta(x, \tau)$, as:

$$I(\delta(x,\tau)) = \exp(\tau X) \cdot I(x)$$

and so equation 5 may be re-written as,

$$\mathcal{A}\{\exp(\tau \mathrm{X}) \cdot f\} \ = \lambda \mathcal{A}\{f\}$$

For example, the group of translations, $x' = x + \tau$, generated by the group operator $\mathrm{X} = \mathcal{L}_x = \partial/\partial x$ can be represented by applying the shift operator to the original coordinate vector,

$$\left(\exp\left(\tau \frac{\partial}{\partial x}\right)\right) x = x + \tau$$

In other words, the eigenproblem for the group of 1D translations,

$$|f(x' - \tau) - \lambda f(x)| = 0$$

is solved by substituting the shift operator into the characteristic equation,

$$\exp(\tau \mathrm{X}) \cdot f - \lambda f = 0$$

The dot product may be applied as a simple product in the domain mapped to by an integral transform with appropriate properties; in this case, the Fourier transform[3].

Using this analysis of the group of 1D translations, it is possible to outline the theory necessary for designing an operator which will respond to such a characteristic deformation. This approach can then be extended to each of the six local coordinate system transformations associated with the complete rigid body motion group. It is thereby proposed that a Lie group of complex operators [or *pairs* of cosinusoidally-modulated operators; each pair separated by a 90 degree phase difference) can be used to measure these particular image deformations. Each of the six local coordinate system transformations are estimated by spatially-localised operators [at varying resolutions] which respond to their characteristic vector fields.

4.2 2D Image Translation: $\mathcal{L}_x$ and $\mathcal{L}_y$

The Fourier Transform $\mathcal{I}(\jmath\omega_x)$ of a 1D input signal $I(x)$ is,

$$\mathcal{I}(\jmath\omega_x) \triangleq \mathcal{F}\{I(x)\} \triangleq \int_{-\infty}^{\infty} I(x) \cdot e^{-\jmath\omega_x x} dx$$

So, for the infinitesimal transformation given by $\mathcal{L}_x$, the shift property of the Fourier Transform yields, (see Bracewell[29, p. 104])

$$\mathcal{F}\{I(x+\tau)\} = e^{\jmath\omega_x\tau} \cdot \mathcal{I}(\jmath\omega_x)$$

Which means that an $\mathcal{L}_x$ deformation will cause a phase shift of $\tau\omega_x$ of the complex frequency response at each spatial frequency, ω_x. Likewise for the $\mathcal{L}_y$ operator: the phase shift will be $\tau\omega_y$ at each y-oriented spatial frequency, ω_y. The two can be combined to give the image translation at a particular 2D orientation. This can be measured by observing the phase relationship of even and odd Gabor kernels with a particular spatial frequency and orientation tuning. [When the operator has a finite bandwidth,

[2]See Wave Theory for 'special representations of flow', and 'characteristics of flow'. They give a vector or matrix representation which can be solved for Jacobians and Eigenspaces.

[3]Note: The Fourier Transform [and other integral transforms for that matter] has a matrix representation which can be used when describing it as an eigenfunction[27]. (cf. Rosenfeld and Kak[28, p.121], re. Karhunen-Loève transform).

the phase shift with respect to the peak frequency is considered]. (cf. Jenkin, Jepson and Tsotsos[30], Fleet and Jepson[31], and Fleet[21]). One problem with this estimation is that the even Gabor signal will in general have a response to the "DC" component of the signal, whereas the odd Gabor will not. Thus, there will be a phase shift measured even when the intensity function varies uniformly (over the spatial variable).

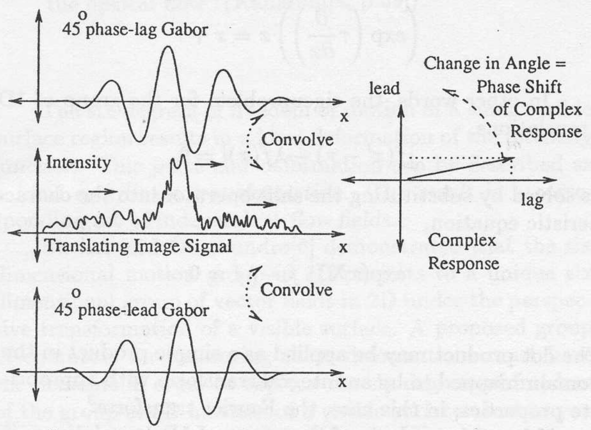

Figure 4: 1D Translation as Phase Shift of Complex Quadrature Pair Response

Another pair of quadrature-shifted sinusoidal Gabor signals can be constructed so as to have *equal* "DC" offset. In fact, they will have the same spectral response amplitude profile. These are a pair of cosinusoids: one with a phase *lead* of $\pi/4$ radians, the other with a phase *lag* of $\pi/4$ radians. The two are mirror-symmetrical about the origin.

4.2.1 1D Translation as Local Phase Behaviour

Fleet[21] and Jepson[32] have characterised normal velocities in image regions as "local phase behaviour". Their formulations are based on the relationship of translation in the image to the phase shift of the individual components making up the frequency-domain representation of the image. Their techniques extract the *normal velocity* (ie. the pure translation component) of motion in the image. This is a fundamental property of the Fourier transform; that translation in the original signal can be represented in the frequency domain with the complex phase shift operator, $e^{-\jmath\omega_x\Delta x}$. The Fourier components making up the original image are all phase-shifted by an amount $\omega_x \Delta x$ [radians] as a result.

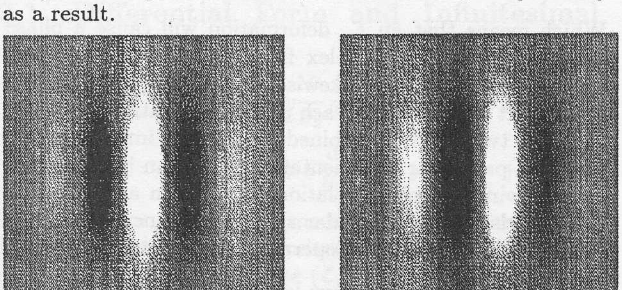

Figure 5: Even and Odd x-Translation Operators: $\exp\{-iux\}$, $u = 1.5$

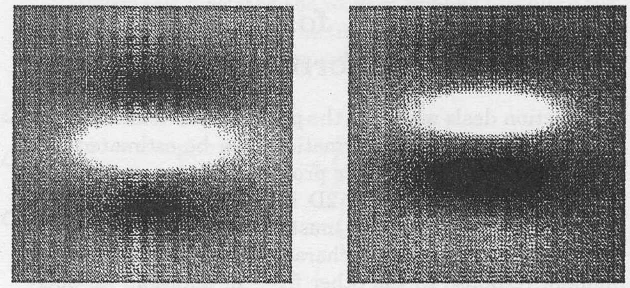

Figure 6: Even and Odd y-Translation Operators: $\exp\{-ivy\}$, $v = 1.0$

4.3 Divergence and Curl: $\mathcal{L}_s$ and $\mathcal{L}_r$

For the $\mathcal{L}_s$ operator, e^τ acts as a frequency-scaling parameter, shown by the following Fourier transform theorem,

$$\mathcal{F}\{I(xe^\tau)\} = \frac{1}{e^\tau}\mathcal{I}\left(\frac{\jmath\omega_x}{e^\tau}\right)$$

An $\mathcal{L}_s$ deformation of τ will cause the frequency domain response to re-scale by a factor of $e^{-\tau}$. In other words, pure motion in depth will result in a spectral shift of information in the frequency domain (As pointed to earlier in Eagleson[33]). Alternatively, the $\mathcal{L}_s$ deformation $x' = x\exp(\tau)$ could be expressed as,

$$\ln \mid x' \mid = \ln \mid x \mid + \tau$$

So, if the spatial intensity function is re-scaled in a logarithmic (or log-polar fashion, since the $\mathcal{L}_s$ deformation is radially symmetric), then diverging flow or shifts in spatial frequency can be regarded as *phase-shift* in log-polar space. ie. by substituting $a = \ln \mid x \mid$, the Fourier transform of this spatially-scaled representation is,

$$\mathcal{F}\{I(a+\tau)\} = e^{\jmath\omega_a\tau} \cdot \mathcal{I}(\jmath\omega_a)$$

The logarithmic spatial scaling of the Laplace or Fourier transform inputs corresponds to the Mellin transform. This relationship will be used in the design of the operator used to measure the divergence of the image flow; ie. its $\mathcal{L}_s$ deformation in the following section.[4]

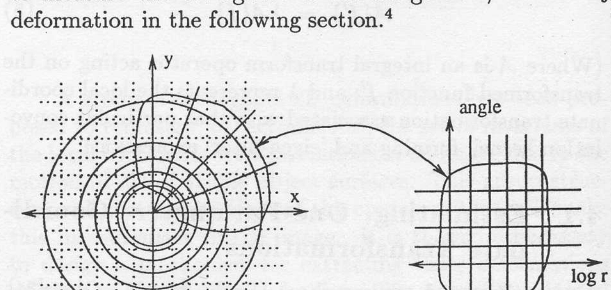

Figure 7: Mapping to log-polar coordinates: divergence and curl become orthogonal translations.

[4]Caelli and Ferraro give a much more formal derivation of a log-polar transform. This is then used to specify the operators for measuring divergence and curl.

4.3.1 Filters based on Log-Polar Transform

The Mellin transform can be described as a log-polar transform, and is defined as:

$$\mathrm{F_M}(s) \triangleq \int_0^\infty f(x) \cdot x^{s-1}\, dx$$

If the negative logarithmic scale of the spatial variable x is taken, it can be easily shown that this is equivalent to the scaled function's Fourier transform if the following substitutions are made: $a = -\log x$ and $s = \jmath\omega_a$. In other words,

$$\mathrm{F_M}(\jmath\omega_a) = \int_{-\infty}^\infty f(e^{-a}) \cdot e^{-\jmath\omega_a a}\, da$$

The Mellin transform has the following property for spatial scaling deformations, (also in Bracewell[29, p. 257])

$$\mathcal{M}\{f(xe^\tau)\} = e^{-\jmath\omega_a\tau} \cdot \mathrm{F_M}(\jmath\omega_a)$$

which is a phase-shifted version of the result shown in the previous section. Thus, the radially-symmetric operators, $\mathcal{L}_s$ and $\mathcal{L}_r$ *could* be extracted as phase shifts on a log-polar map of the local (x, y) neighbourhood; as shown in figure 7. This is consistent with log-polar image coordinates for motion analysis such as that proposed by Jain[34], and Narathong et al[35]. The latter estimated 2D translation as well as the deformations of divergence and curl using ratios of spatio-temporal gradients. Cavanagh[36, p.88] proposed a biological vision model in which complex cells in layers 2 and 3 of the striate cortex perform local log-polar transformations.

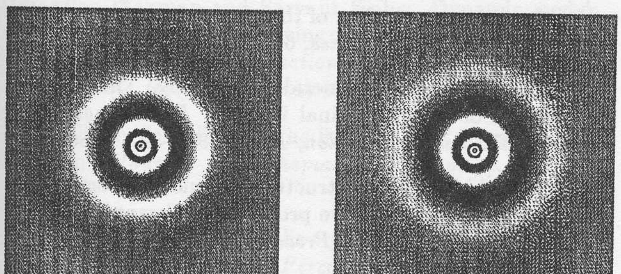

Figure 8: Even and Odd Divergence Operators: $\exp\left\{-iu \cdot \log(x^2+y^2)^{1/2}\right\}$, $u = 6$

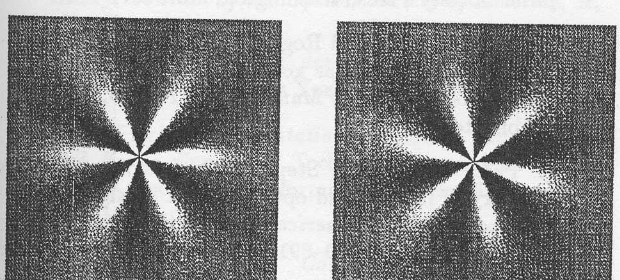

Figure 9: Even and Odd Curl Operators: $\exp\left\{-iv \cdot \tan^{-1}\left(\frac{y}{x}\right)\right\}$, $v = 6$

4.3.2 Implementation of Operators

Caelli and Ferraro[19] derived the kernel of the integral transform invariant for rotations and dilations. They showed that if the Lie bracket $[\mathcal{L}_1, \mathcal{L}_2] = 0$, (ie. $\mathcal{L}_1$ and $\mathcal{L}_2$ are orthogonal) then the transform kernel could be found as follows:

1) Solve for the canonical form of $\eta(x, y)$ and $\xi(x, y)$ by constraining the coordinates to be independent and invariant of their respective Lie operators;

$$\mathcal{L}_1(\eta) = 1, \quad \mathcal{L}_1(\xi) = 0$$

$$\mathcal{L}_2(\eta) = 0, \quad \mathcal{L}_2(\xi) = 1$$

2) Compute the Jacobian determinant of the transformation from (η, ξ) to (x, y) giving the kernel,

$$w(u, v; x, y) = \left| J\begin{pmatrix} \eta & \xi \\ x & y \end{pmatrix} \right| \exp\{-i\ [\eta(x, y)u + \xi(x, y)v\ ]\}$$

which is used as a convolution operator, ie. $g(u, v) = \int\int_{-\infty}^{\infty} f(x, y)\ w(u, v; x, y)\ dxdy$.

$$w_{r,s}(u, v; x, y) = \frac{\exp\left\{-i\left[u \log(x^2+y^2)^{1/2} + v \tan^{-1}\left(\frac{y}{x}\right)\right]\right\}}{x^2+y^2}$$

ie. w_r and w_s are basis functions for a log-polar transformation, just as $\exp\{-\jmath\omega x\}$ and $\exp\{-\jmath\omega y\}$ are basis functions for the Fourier transform. In this case, the coordinate transformation corresponds to,

$$\mathcal{A}\{\exp(\tau X) \cdot f\} = \lambda \mathcal{A}\{f\}$$

where,

$$\mathcal{A}\{f\} \triangleq \int_0^\infty f \cdot x^{s-1}\, dx$$

$$X = \mathcal{L}_s = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} \text{ and } \lambda = e^{-\jmath\omega\tau}$$

In the following subsection (4.4) their approach is taken in order to find the integral kernel for the 'shear' deformations.

4.4 Shear Deformations: $\mathcal{L}_b$ and $\mathcal{L}_B$

We would now like to find the coordinate system that transforms the two orthogonal shear deformations in (x, y) to independent phase shifts in (η, ξ).

The invariants of the $\mathcal{L}_s$ and $\mathcal{L}_r$ deformations correspond to the mapping to log-polar space. A point (x, y) can be rewritten as $(l \cos\theta, l \sin\theta)$ where $r^2 = x^2 + y^2$, $l = \log|r|$, and $\theta = \tan^{-1}(y/x)$. Similarly, an hyperbola can be parameterised in terms of (l, θ) as the set of points $(l \cosh\theta, l \sinh\theta)$ where θ must now be interpreted in *hyperbolic radians.* We would therefore expect eigenfunctions based on the inverse hyperbolic tangent function, or its logarithmic equivalent given by the identity,

$$\tanh^{-1} u = \frac{1}{2} \log \frac{1+u}{1-u}$$

Based on this trial function, it can be shown that

$$\mathcal{L}_b \log \left(\frac{x}{y}\right)^{1/2} = 1 \quad \text{and,} \quad \mathcal{L}_B \log \left(\frac{x+y}{x-y}\right)^{1/2} = 1$$

which leads to the equivalent functions[5],

$$\eta_b(x,y) = \frac{-1}{2}\log\left(\frac{y}{x}\right) \quad \text{and,} \quad \xi_B(x,y) = \tanh^{-1}\left(\frac{y}{x}\right)$$

5 Discussion

This paper has presented a class of operators to be used for extimating six local coordinate transformations that can occur in the image due to visible objects moving in 3D with six degrees of freedom. These deformations are commonly referred to as 2D translation, divergence, curl, and 2D shear.

The motivation for measuring this set of deformations is that they cover the six degrees of freedom of motion of moving rigid bodies, as projected to a 2D image perspectively (cf. Blicher and Omohundro[6]). Kanatani[4] showed that these are irreducible representations of the motion occurring in the image. Based on the 'characteristics' of these flow fields, a family of spatial filters have been introduced to estimate the local image transformations. Thus, we can think of these operators as being 'Eigenfilters' for the 6D Lie group covering motion-induced local coordinate system transformations.

Watanabe[37] showed that this type of factorization is similar to a Karhunen-Loève expansion of the input signal, and so we would expect to inherit the same entropy-minimising and error-minimising properties of these techniques. Such expectations demand further evaluation and experimentation for a complete evaluation of these operators. This is currently the subject of further research.

Acknowledgements

I am deeply indebted to my advisor, Dr. Zenon Pylyshyn, for his advice and guidance in this research; supported by NSERC operating grant number A2600. I am also extremely grateful for the direction, criticism, and suggestions offered by Terry Caelli, John Barron, and Judith Hamilton.

The following have made significant comments on earlier drafts: Keith Humphrey and Rick Gurnsey, Peter Dodwell, Michael Leyton, Steve Zucker, Evangelos Milios, David Fleet, Allan Jepson, John Tsotsos, and Alan Mackworth.

[5] According to Caelli and Ferraro[19], one would expect that since $\mathcal{L}_b(x^2 - y^2) = 0$ and $\mathcal{L}_B(xy) = 0$, ξ should be of the form $\xi(x^2 - y^2)$ and η should be of the form $\eta(xy)$. We note that the family of lines of constant (xy) and $(x^2 - y^2)$ are congruent sets of hyperbolae; they are rotated $\pi/4$ radians with respect to each other. It should therefore be noted that these equations can be expressed using these conic terms, giving,

$$\eta(x,y) = \frac{-1}{2}\log\frac{y^2}{xy} \quad \text{or} \quad \eta(x,y) = \frac{1}{2}\log\frac{x^2}{xy}$$

$$\xi(x,y) = \frac{1}{2}\log\frac{2xy + (x^2+y^2)}{x^2 - y^2}$$

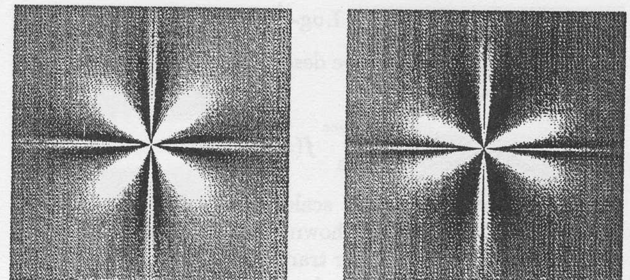

Figure 10: Even and Odd Shear Operators: $\exp\left\{iu \cdot \frac{1}{2}\log\frac{y}{x}\right\}$, $u = 3$

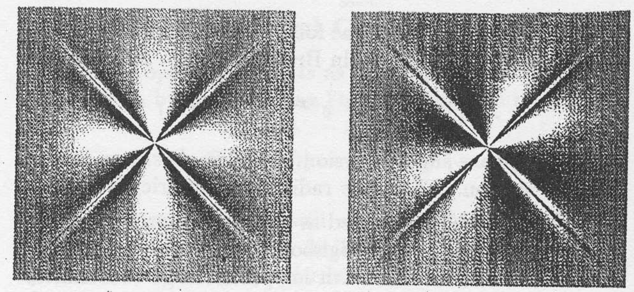

Figure 11: Even and Odd Shear Operators: $\exp\left\{-iu \cdot \tanh^{-1}\frac{y}{x}\right\}$, $u = 3$

References

[1] J.J. Koenderink and A.J. Van Doorn. Local structure of movement parallax of the plane. *Journal of the Optical Society of America*, 66:717–723, 1976.

[2] H.C. Longuet-Higgins and K. Prazdny. The interpretation of moving retinal images. *Proceedings of the Royal Society of London, B.*, 208:385–387, 1980.

[3] Ken-Ichi Kanatani. Structure and motion from optical flow under perspective projection. *Computer Vision, Graphics, and Image Processing*, 38:122–146, 1987.

[4] Ken-Ichi Kanatani. Coordinate rotation invariance of motion characteristics for 3D shape and motion recovery. In *Proceedings of the IEEE First International Conference on Computer Vision*, pages 55–64, Computer Society Press, Washington, June 8-11 1987.

[5] Eloise H. Carlton and Roger N. Shepard. Psychologically simple motions as geodesic pathes: I. asymmetric objects. *Journal of Mathematical Psychology*, 1988. (Submitted article).

[6] A. Peter Blicher and Stephen Omohundro. Unique recovery of motion and optic flow via lie algebras. In *Proceedings of the American Association for Artificial Intelligence*, pages 889–891, 1985(?).

[7] Patrick Cavanagh. Reconstructing the third dimension: Interactions between colour, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and Image Processing*, 37(2):171, February 1987.

[8] Ken Nakayama. Motion parallax sensitivity and space perception. In A. Hein and M. Jeannerod, editors, *Spatial Oriented Behaviour*, pages 223–242, Springer, Berlin, 1983.

[9] Ellen C. Hildreth and Christof Koch. The analysis of visual motion: From computational theory to neuronal mechanisms. *Ann. Rev. Neurosci.*, 10:477–533, 1987.

[10] H. von Helmholtz. *Treatise on physiological optics.* Dover, New York, 1925. (ed. J.P.C. Southall).

[11] J.J. Gibson. *The perception of the visual world.* Houghton Mifflin, Boston, 1950.

[12] H.C. Longuet-Higgins. The visual ambiguity of a moving plane. *Proceedings of the Royal Society of London, Series B*, 223:165–175, 1984.

[13] A.M. Waxman and Shimon Ullman. Surface structure and three-dimensional motion from image flow kinematics. *International Journal of Robotics Research*, 4(3):72–94, 1985.

[14] M. Subbarao and A.M. Waxman. Closed form solution to image flow equations for planar surface in motion. *Computer Vision, Graphics, and Image Processing*, 36:208–228, 1986.

[15] M. Subbarao. *Interpretation of Image Motion Fields: Rigid Curved Surfaces in Motion.* Technical Report CAR-TR-199, Center for Automation Research, University of Maryland, tech 1986.

[16] Allen Waxman and Sarvajit Sinha. Dynamic stereo: Passive ranging to moving objects from relative image flows. *IEEE Transactions on Pattern Analysis and Machine Intellignece*, 8(4):406–412, July 1986.

[17] William C. Hoffman. The lie algebra of visual perception. *Journal of Mathematical Psychology*, 3:65–98, 1966.

[18] Peter C. Dodwell. The lie transformation group model of visual perception. *Perception and Psychophysics*, 34(1):1–16, 1983.

[19] Mario Ferraro and Terry M. Caelli. Relationship between integral transform invariances and Lie group theory. *Journal of the Optical Society of America, A*, 5:738–742, May 1988.

[20] Sophus Lie. *Theorie der Transformationsgruppen.* Chelsea Pub. Co., New York, second edition, 1970.

[21] David Fleet. *Implementation of Velocity-Tuned Filters and Image Encoding.* Technical Report FBI-HH-M-159/88, Fachbereich Informatik, Universität Hamburg, July 1988.

[22] Zenon W. Pylyshyn. *Computation and Cognition.* MIT Press, 1984.

[23] Yi-Tzuu Chien. *Interactive Pattern Recognition. Electrical Engineering and Electronics*, Marcel Dekker, Inc., New York, 1978.

[24] Peter J. Olver. *Applications of Lie Groups to Differential Equations.* Volume 107 of *Graduate texts in mathematics*, Springer-Verlag, New York, 1986.

[25] L.V. Ovsiannikov. *Group Analysis of Differential Equations.* Academic Press, 1982. Translated by W.F. Ames.

[26] Johan G. F. Belinfante and Bernard Kolman. *A Survey of Lie Groups and Lie Algebras with Applications and Computational Methods.* Society for Industrial and Applied Mathematics, Philadelphia, 1972.

[27] Terry Caelli. 1989. Personal Communication.

[28] Azriel Rosenfeld and Avinash C. Kak. *Digital Picture Processing.* Volume 1 of *Computer Science and Applied Mathematics*, Academic Press, second edition, 1982.

[29] Ronald N. Bracewell. *The Fourier Transform and its Applications.* McGraw-Hill, 1978.

[30] Michael J.M. Jenkin, Allan D. Jepson, and John K. Tsotsos. *Techniques for Disparity Measurement.* Technical Report RBCV-TR-87-16, The University of Toronto, September 1987.

[31] David Fleet and Allan Jepson. Hierarchical construction of orientation and velocity selective filters. *IEEE Trans. PAMI*, to appear. (also see Proc. SPIE 594 (Image Coding), pp. 10-20, 1985; and Technical Report: RBCV-TR-85-8, University of Toronto, Dept of Computer Science).

[32] Allan D. Jepson. Personal Communication.

[33] Roy Eagleson. Estimating 3D motion parameters from the changing responses of 2D bandpass spatial frequency filters. In *IEEE Montréal Technologies Conference: Compint '87*, pages 102–105, Montréal, Québec, November 9-12 1987.

[34] Ramesh Jain. Complex logarithmic mapping and the focus of expansion. In Norman I. Badler and John K. Tsotsos, editors, *Motion: Representation and Perception*, pages 137–144, North-Holland, 1986.

[35] C. Narathong, R.M. Iñigo, J.F. Doner, and E.S. McVey. Motion - Vision architectures. In *Computer Vision and Pattern Recognition*, pages 411–416, June 1988.

[36] Patrick Cavanagh. Local log polar frequency analysis in the striate cortex as a basis for size and orientation invariance. In D. Rose and V. G. Dobson, editors, *Models of the Visual Cortex*, pages 85–95, John Wiley and Sons Ltd, 1985.

[37] Satosi Watanabe. Karhunen-Loève expansion and factor analysis: Theoretical remards and applications. In *Trans. Fourth Prague Conf. Infrom. Theory, Statist. Decision Functions, and Random Processes*, pages 635–660, 1965. Prague.