International Journal of Computer Vision 55(1), 55–71, 2003
© 2003 Kluwer Academic Publishers. Manufactured in The Netherlands.

Optical Snow

MICHAEL S. LANGER
School of Computer Science, McGill University, Montreal, Quebec H3A 2A7, Canada
langer@cim.mcgill.ca

RICHARD MANN
School of Computer Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
mannr@uwaterloo.ca

Received July 26, 2002; Revised March 25, 2003; Accepted April 1, 2003

Abstract. Classical methods for measuring image motion by computer have concentrated on the cases of optical flow in which the motion field is continuous, or layered motion in which the motion field is piecewise continuous. Here we introduce a third natural category which we call optical snow. Optical snow arises in many natural situations such as camera motion in a highly cluttered 3-D scene, or a passive observer watching a snowfall. Optical snow yields dense motion parallax with depth discontinuities occurring near all image points. As such, constraints on smoothness or even smoothness in layers do not apply. In the Fourier domain, optical snow yields a one-parameter family of planes which we call a bowtie. We present a method for measuring the parameters of the direction and range of speeds of the motion for the special case of parallel optical snow. We demonstrate the effectiveness of the method for both synthetic and real image sequences.

Keywords: motion analysis, Fourier transform, optical flow, layered motion, transparency, motion field, spatiotemporal filters

1. Introduction

It has been known for over half a century that optical flow is a strong cue for human visual navigation (Gibson, 1950). Most studies of optical flow assume that there is a unique velocity vector at each point in the visual field (Koenderink, 1986; Barron et al., 1994). This assumption is only valid over regions of the image in which the depth map is continuous. In this paper we introduce a new natural motion category that we call "optical snow". Optical snow is a generalization of optical flow in which the assumption of local spatial continuity of the motion field is abandoned.

Optical snow arises in many natural situations in which an observer moves relative to a densely cluttered 3-D scene (see Fig. 1). An extreme example of optical snow is that of falling snow seen by a static observer. Snowflakes fall approximately vertically and the image speed of each snowflake varies inversely with its distance from the camera. Because any image region may contain the snowflakes visible at multiple depths, the motion field will be discontinuous within that region. A second natural example of optical snow is the motion seen by an observer moving laterally past a cluttered 3D object such as a bush or through a cluttered scene such as a forest. Any image region contains objects such as leaves and branches at multiple depths and so discontinuities in the motion occur in nearly every local image region.

Despite the apparent complexity this new category of motion, optical snow provides a rich set of visual cues about depth and 3-D spatial layout. People

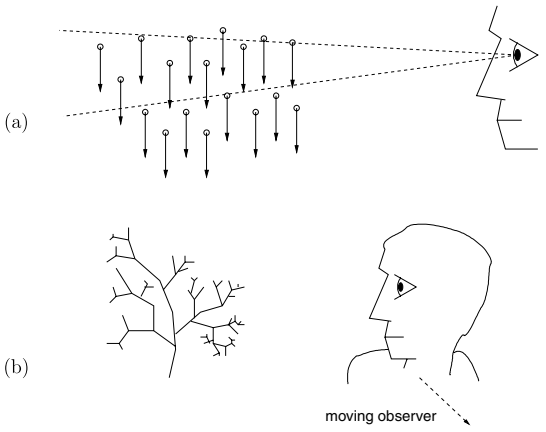

Figure 1. Optical snow. Within each image patch there is a single direction of motion, but a continuous range of speeds generated by motion parallax. (a) A passive observer watches a snow fall. (b) The camera/observer moves relative to a cluttered 3D scene.

observe falling snow and experience a rich sense of depth. People also navigate easily through 3D cluttered environments such as forests and grasslands. This informal observation is supported by formal psychophysical studies which show that human observers can judge direction of heading accurately in a "cloud of dots" environment (Warren and Hannon, 1990). Interestingly, the many kinds of animals that inhabit natural 3D cluttered environments include the rabbit, cat, bird, and monkey.[1] These are among the most heavily studied animals in the neuroscience of motion processing. Since much of what is known about motion processing in biological vision comes from experiments on these animals, and since these animals are presumably specialized for optical snow as well as for optical flow, it makes sense to investigate the computational problems of motion processing in optical snow. That is our goal in the present paper.

1.1. Related Work

How is optical snow related to the two traditional categories of motion used in computer vision, namely optical flow and layered motion? We argue that these three categories of motion can be related to each other on the following 1D continuum. At one end of this continuum is optical flow. The traditional assumption of optical flow is that the image velocity is a piecewise continuous function of position in the image (Barron et al., 1994). In the middle of the continuum of models is layered motion. Traditional layered models assume the velocity field of each layer is either constant (Fennema and Thompson, 1979) or affine, that is, varies linearly with image position (Waxman and Wohn, 1985; Wang and Adelson, 1993; Darrell and Pentland, 1995; Sawhney and Ayer, 1996; Black and Anandan, 1996). Layered models have been used to describe a small number of opaque objects in relative motion. They have also been used for various types of motion transparency, both for constant motion fields (Shizawa and Mase, 1991; Bergen et al., 1992; Jepson and Black, 1993; Milanfar, 1996; Weiss, 1997) and for affine motion fields (Black and Anandan, 1996; Irani et al., 1994; Ju et al., 1996). In the case of transparency, layered models do not assume that the velocity field is smooth in the image, rather only that the velocity field is smooth within each layer (Weiss, 1997). Optical snow is found at the other extreme end of the 1D continuum of models, where the number of motion layers would have to be so high that any assumption of layers including smoothness in layers would be inappropriate.

To reinforce these ideas, consider this 1D continuum of motion models for a laterally moving observer. For this example, assume the camera motion is horizontal. If the scene contained a single smooth surface, then the image motion would be continuous optical flow in the horizontal direction. If the scene contained a small number of surfaces, possibly transparent and with different depths, then layered motion would result with one surface per layer and with horizontal motion in each layer. Examples are the standard flower garden (Wang and Adelson, 1993; Darrell and Pentland, 1995) and SRI tree data set (Barron et al., 1994), and the reflection-in-the-window transparency examples such as in (Bergen et al., 1992). If the scene were densely cluttered in 3D, for example with bushes or trees, then optical snow would result. The velocity field would contain a dense set of discontinuities, and the velocity of each visible point would be horizontal. We emphasize that a layered model would be inappropriate here because the number of layers would be too high.

What can an observer infer about 3D structure in the three cases? In each case, the horizontal laterally moving observer observes horizontal image motion. In the case of optical flow, the observer could apply a smoothness constraint and recover the velocity field with great accuracy (Barron et al., 1994). In the case of layered motion, the observer must reduce the space of allowable motions (e.g. to affine models for each layer) in order to

cope with depth discontinuities or with transparencies. In the case of optical snow, the observer must abandon layered models entirely. The only remaining constraint is that the velocity vectors are restricted to a line in velocity space, namely the horizontal line through the origin.

Optical flow and layered motion analysis have received much attention in the computer vision literature and the computational goals of the problem and algorithms for solving the problem have been defined and developed. To our knowledge, optical snow has not been studied before as a computational problem, despite the fact that it is a common motion category in nature and motions of this category give rise to rich qualitative percepts in human vision. Our goal in this paper is to provide the groundwork for analysis of optical snow. We define the computational goal to be the recovery of the parameters of the motion, namely the set of velocities occurring an image region. As we will describe later, the set of image velocities lies on a line in velocity space. We will use this constraint to solve the computational problem for a particular class of optical snow motions that we call parallel optical snow. The generalization to non-parallel snow will be addressed in a subsequent publication.

1.2. Overview

Section 2 provides the background. We review the image flow constraint equation and several previous results about optical flow in the frequency domain. Section 3 introduces optical snow. We show that it has a characteristic "bowtie" pattern in the frequency domain. This generalizes a classical frequency domain model for optical flow. Section 4 considers the specific case of parallel optical snow in which all velocity vectors in an image region are parallel. The algorithm and experiments in the paper focus on this case. Section 5 discusses how the bowtie pattern is affected by occlusions. Section 6 shows how to estimate the parameters of optical snow for a given image sequence. A two stage solution is presented: the first stage is to estimate the direction of the motion. This corresponds to estimating the axis of the bowtie. The second stage is to estimate the range of speeds present in the motion. Section 7 analyzes the aperture problem and shows how it is manifest in optical snow. Some of the results of this paper appeared in preliminary form in Langer and Mann (2001, 2002) and Mann and Langer (2002).

2. Background: Motion in Frequency Domain

Analysis of image motion in the spatiotemporal frequency domain has a long history in motion understanding research. Much of the analysis is based on the *motion plane property* (Watson and Ahumada, 1985) which states that an image pattern that translates with a uniform image velocity produces a plane of energy in the frequency domain.

The intuition behind the motion plane property is as follows. If an image sequence is created by a translating a single image frame over time, say with velocity (v_x, v_y), then each of the 2D spatial frequency components of the single image frame itself travels with this velocity. Each of these translating 2D sine waves produces a unique spatiotemporal frequency component in the translating image sequence. Thus, the velocity vector (v_x, v_y) induces a specific relationship (see Eq. (5) later) between the temporal frequency of the translating wave as seen at a pixel over time and the spatial frequencies of the wave as seen at one time and over all pixels.

The motion plane property may be formally derived as follows. Let $I(x, y, t)$ be a time varying image. Assume that $I(x, y, t)$ has been smoothed by a low pass filter so that derivatives are well-defined. Assume the image is translating with velocity $(v_x, v_y) = (dx/dt, dy/dt)$, that is, assume

$$I(x, y, t) \;=\; I(x + v_x dt, y + v_y dt, t + dt). \quad (1)$$

From the image flow constraint equation (Horn and Schunck, 1981), the velocity (v_x, v_y) is constrained by:

$$v_x \frac{\partial I}{\partial x} + v_y \frac{\partial I}{\partial y} + \frac{\partial I}{\partial t} = 0. \quad (2)$$

To derive the motion plane property, we use the derivative property of Fourier transforms:

$$\int \frac{\partial I(u)}{\partial u} e^{-2\pi i f u} du = -2\pi i f \int I(u) e^{-2\pi i f u} du. \quad (3)$$

Treating (v_x, v_y) as a constant, we apply Eq. (3) to Eq. (2) which yields:

$$-2\pi i \;(v_x f_x + v_y f_y + f_t)\; \hat{I}(f_x, f_y, f_t) = 0 \quad (4)$$

where $\hat{I}(f_x, f_y, f_t)$ is the Fourier transform of $I(x, y, t)$. Equation (4) implies that

$$v_x f_x + v_y f_y + f_t = 0 \quad (5)$$

whenever $\hat{I}(f_x, f_y, f_t) \neq 0$. That is, all motion energy lies on the plane of Eq. (5). This is the motion plane property.

Several methods for measuring translational image motion (or constant optical flow) have been based on this motion plane property. For example, frequency-based optical flow methods recover a unique velocity (v_x, v_y) in a local region of the image by finding the motion plane that best fits the 3D power spectrum of that local region (Heeger, 1987; Grzywacz and Yuille, 1990; Simoncelli and Heeger, 1998; Huang and Chen, 1995). Methods for recovering layered transparent motion have also used the motion plane property. These methods assume linear superposition of two or more motion planes in the frequency domain and attempt to recover these planes for a given image sequence (Shizawa and Mase, 1991; Milanfar, 1996). The model of optical snow that we develop in the next section can be thought of as a generalization of layered motion transparency, in which the number of layers is very high.

3. Optical Snow

We begin the technical discussion of optical snow with the following observation. The motion plane property can be extended from the case of pure translation to the case in which there is a one-parameter set of velocities within an image region. Suppose that each velocity vector in an image region is of the form

$$(v_x, v_y) = (u_x + \alpha\tau_x, u_y + \alpha\tau_y) \tag{6}$$

where $\{u_x, u_y, \tau_x, \tau_y\}$ are constants and the parameter α varies within the region. We may assume without loss of generality that (τ_x, τ_y) is a unit vector and that (u_x, u_y) is perpendicular to (τ_x, τ_y) (see Fig. 2).

Substituting Eq. (6) into Eq. (5) produces a one-parameter family of planes in the frequency domain,

$$(u_x + \alpha\tau_x)f_x + (u_y + \alpha\tau_y)f_y + f_t = 0 \tag{7}$$

where α is the free parameter. The central property of our model of optical snow is the following.

Proposition 1. *The one parameter family of motion planes in Eq.* (7) *intersects at a common line that passes through the origin* (*see Fig. 3*).

Proof: Each of motion planes in Eq. (7) has a normal vector $(u_x + \alpha\tau_x, u_y + \alpha\tau_y, 1)$. These normal vectors all lie on a line in the plane $f_t = 1$. Let us call this line

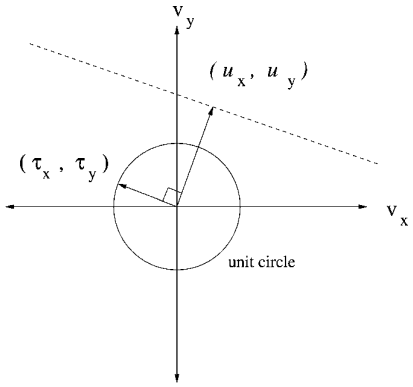

Figure 2. Equation (6) defines a one-parameter family of velocites. We assume without loss of generality that (τ_x, τ_y) is a unit vector and that (u_x, u_y) is perpendicular to (τ_x, τ_y).

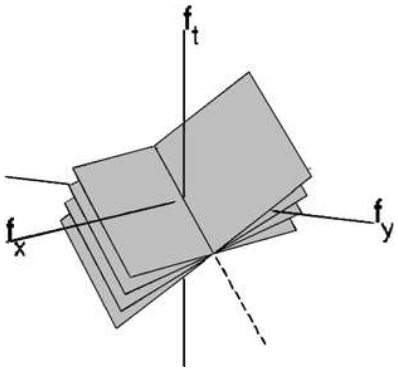

Figure 3. The motion described by Eq. (7) results in a family of planes in the frequency domain. The planes intersect at a line that passes through the origin. We refer to this as a "bowtie" signature.

l. The line l, together with the origin, define a plane π in the frequency domain. The vector perpendicular to π is, by definition, perpendicular to each of the vectors in l. Hence, the line from the origin in the direction of this perpendicular vector must lie in each of the motion planes. □

Because the planes intersect at a common line, we say that the family of planes has a *bowtie signature* and we say the common line of the planes is the *axis of the bowtie*.

Proposition 2. *The axis of the bowtie is in the direction* $(-\tau_y, \tau_x, \sqrt{u_x^2 + u_y^2})$.

Proof: The direction of the axis of the bowtie can be computed by taking the cross product of any two of

the normal vectors in l. Taking the two normal vectors defined by $\alpha = \{0, 1\}$ yields

$$(u_x, u_y, 1) \times (u_x + \tau_x, u_y + \tau_y, 1) = (-\tau_y, \tau_x, u_x\tau_y - u_y\tau_x).$$

The third component of the bowtie axis is obtained by setting (τ_x, τ_y) to be a unit vector, and choosing its direction such that $(u_x, u_y) \times (\tau_x, \tau_y) > 0$. □

3.1. *Lateral Observer Motion*

As illustrated in Fig. 1, optical snow arises naturally when an observer moves laterally in a static 3D cluttered scene. The general equations of the image velocity field for an observer moving relative to an arbitrary scene are presented in Longuet-Higgins and Prazdny (1980) and Trucco and Verri (1998). Here we relate these equations to optical snow by considering the specific scenario of lateral motion.

Assume a camera is moving through a static scene, and let the instantaneous motion of the camera be represented by a translation vector (T_x, T_y, T_z) and a rotation vector $(\Omega_x, \Omega_y, \Omega_z)$. Suppose that the camera is moving near laterally, that is, the translation vector is near perpendicular to the optical axis. Formally,

$$|T_z| \ll \|(T_x, T_y)\|.$$

Following Lappe and Rauschecker (1993), we further suppose that the camera can pan and tilt, but not roll, that is,

$$\Omega_z \approx 0.$$

When these two conditions hold and when the projection plane of the camera is at unit depth, the image velocity field may be well-approximated by:

$$(v_x, v_y) = (-\Omega_y, \Omega_x) + \frac{1}{Z}(T_x, T_y), \qquad (8)$$

where Z is the depth of the point visible at pixel (x, y). The model of Eq. (8) ignores terms that are second order in the image coordinates x, y. This is valid for pixels that are within roughly ± 20 degrees from the optical axis.

The physical intuition behind Eq. (8) is that the camera rotation generates a constant velocity component $(-\Omega_y, \Omega_x)$ and the lateral translation generates a velocity component that is in the direction of the translation and that has speed inversely proportional to depth. Importantly, Eq. (8) has a general property that the set of velocity vectors in the image lie on a line in velocity space, and thus is of the form of Eq. (6) discussed earlier. This model holds regardless of the scene geometry, for example, whether the scene consists of a single smooth surface or whether it is densely cluttered in 3D with depth discontinuities occurring nearly everywhere.

3.2. *Example: Lateral Observer Motion with Tracking*

A special case of Eq. (8) occurs when the image velocities are parallel. Such parallel motion arises in two natural cases:

(i) There is no camera rotation and so $\Omega_x = \Omega_y = 0$;
(ii) The image motion due to camera rotation is parallel to direction of translation, that is, $(-\Omega_y, \Omega_x)$ is parallel to (T_x, T_y).

An interesting and natural example of case (ii) occurs when camera rotation is used to stabilize (i.e. track) a particular surface patch at some depth Z' in the scene. For this patch, the rotation component exactly cancels the translation component for that surface patch (Lappe and Rauschecker, 1993). For a point at depth $Z \neq Z'$, image velocity will satisfy:

$$(v_x, v_y) = \left(\frac{1}{Z} - \frac{1}{Z'}\right)(T_x, T_y) \qquad (9)$$

Note that near points ($Z < Z'$) have image motion in the opposite direction of distant points ($Z > Z'$). Also note that case (i) is just a special instance of case (ii) in which the tracked point has depth $Z' = \infty$.

An important observation is that Eq. (9) is of the form of Eq. (6) with

$$(u_x, u_y) = (0, 0), \quad (\tau_x, \tau_y) = (T_x, T_y), \quad \alpha = (1/Z - 1/Z').$$

That is, lateral observer motion with tracking produces optical snow, but it is a special case of optical snow in which all the velocity vectors are parallel. We refer to this case as *parallel optical snow*. The remainder of the paper addresses parallel optical snow only.

4. Parallel Optical Snow

Parallel optical snow is a special case of Eq. (6) in which the image velocities are all in the same direction,

$$(v_x, v_y) = (\alpha\tau_x, \alpha\tau_y).$$

From Eq. (7), it follows parallel optical snow produces a bowtie of the form,

$$\alpha\tau_x f_x + \alpha\tau_y f_y + f_t = 0. \tag{10}$$

This bowtie has the following property. This property will be the basis for the method we describe later in Section 6.

Proposition 3. *For parallel optical snow, the axis of the bowtie lies in the* (f_x, f_y) *plane and is in direction* $(-\tau_y, \tau_x)$ *in that plane. The axis is perpendicular to the direction of motion.*

Proof: The proof here is special case of Proposition 1. From Eq. (10), we set $f_t = 0$. This yields the equation of a line in (f_x, f_y) plane. This line passes through the origin and, since (τ_x, τ_y) is perpendicular to this line, $(-\tau_y, \tau_x)$ must be the direction of the line. An alternative proof comes from Proposition 2. Since the (u_x, u_y) vector is zero, the axis of the bowtie is in direction $(-\tau_y, \tau_x, 0)$. □

In Section 6, we will use this property to show how to recover the parameters of a bowtie from an image sequence of parallel optical snow. Before doing so, we present three examples of parallel optical snow and their bowties.

4.1. Examples of Bowtie Pattern

For parallel optical snow, the axis of the bowtie lies in the (f_x, f_y) plane and hence we can parameterize this axis with an angle θ_b. Define θ_b such that $(\cos\theta_b, \sin\theta_b)$ is the axis of the bowtie in the (f_x, f_y) plane (see Fig. 9(a)). Note that $(\tau_x, \tau_y) = (-\sin\theta_b, \cos\theta_b)$ since, from Proposition 3, (τ_x, τ_y) is perpendicular to the axis of the bowtie. As an example, consider an image sequence in which the motion is vertical only. Since the image velocities are of the form $(0, \alpha)$, it follows that $\theta_b = 0$ and so all the motion planes pass through the f_x axis.

To visualize a bowtie signature for a given image sequence $I(x, y, t)$, we compute the power spectrum $|\hat{I}(f_x, f_y, f_t)|^2$ and repeatedly project the power spectrum orthographically onto a set of vertical planes,

$$\cos\theta\ f_x + \sin\theta\ f_y = 0 \tag{11}$$

where $\theta \in [0, \pi)$. For any fixed θ, we compute this projection by summing the power along lines parallel to $(\cos\theta, \sin\theta)$. If a bowtie exists, it should appear at $\theta = \theta_b$.

To reduce image boundary effects in the Fourier transform, we subtract the mean grey level value of the entire sequence from each pixel and window the sequence both in space and time by a Gaussian, prior to computing the power spectrum. The standard deviations $\sigma_x, \sigma_y, \sigma_t$ of Gaussian were chosen so that the width of $I(x, y, t)$ was 6σ along each axis. The window is applied for all image sequences.

4.1.1. Example 1: Falling Spheres (Synthetic Video). A synthetic version of "falling snow" is shown in Fig. 4. The scene consists of a set of spheres of constant 3-D size, placed at random positions within a view volume. The image sequence was generated using OpenGL by moving a virtual camera upwards at a constant 3D velocity.

The spheres were illuminated by collimated light source that was parallel to the line of sight, to minimize visible shadows. The scene was viewed in perspective. The spheres had a radius of 0.2 units and were at depths ranging from $z = 8$ to -8. The camera was at $z = 10$ and the field of view was 30 degrees. The image sequence was created by moving the camera upwards at a constant velocity $V_y = 0.025$ units/frame.

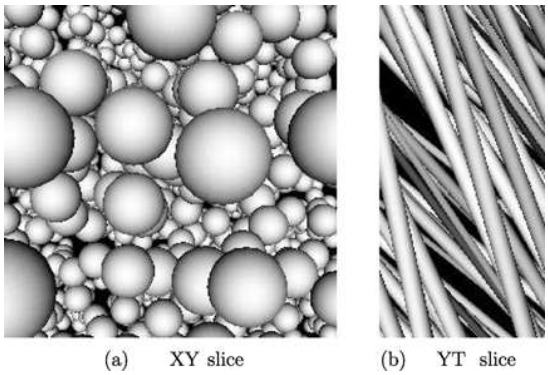

Figure 4. Synthetic "falling spheres" sequence. (a) XY slice of sequence at first frame. (b) YT slice of sequence taken at rightmost pixel column of (a). The data consists of 128 frames, each of which is a 256 × 256 image.

As the camera moves upward, the imaged spheres move downward with speed inversely proportional to their depths. Figure 4(a) shows one frame of the sequence (XY slice) and Fig. 4(b) shows a YT slice (Adelson and Bergen, 1985). Multiple image speeds are evident in Fig. 4(b) in the form of space-time bars of multiple orientations (Bolles et al., 1987).

Figure 5 shows the projected power for several angles θ. In this and subsequent figures, the axes have range $[-N/2 \ldots N/2 - 1]$, where N is the image size and $[-T/2 \ldots T/2 - 1]$, where T is the number of frames in the sequence. To increase contrast, log energy is plotted. White corresponds to high values and black corresponds to low values.

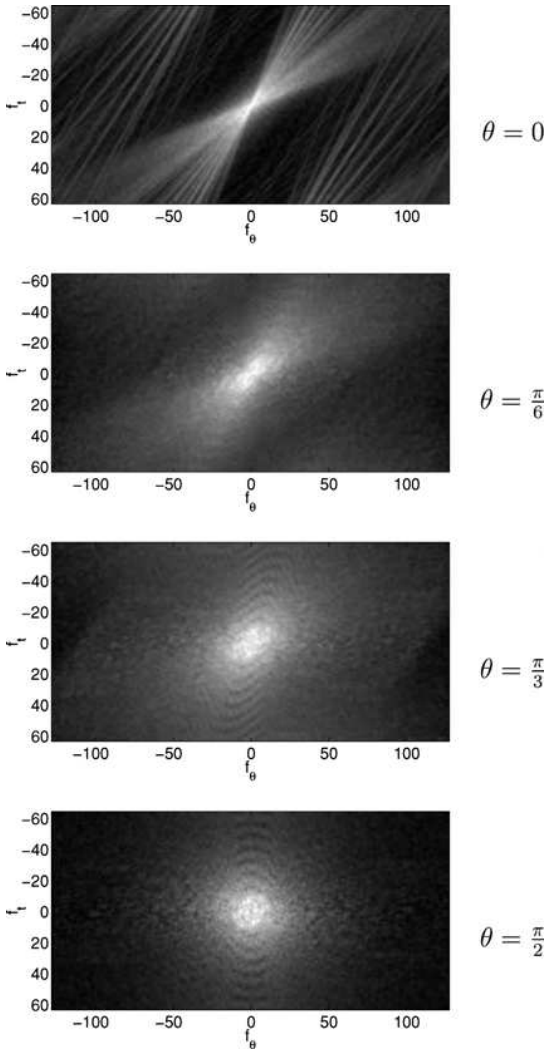

Figure 5. Projected power spectrum for falling sphere sequence in Fig. 4. The bowtie is evident at $\theta = 0$ where the direction of projection is identical to the axis of the bowtie.

The projected power is plotted as a function of (f_θ, f_t) where the f_θ axis is perpendicular to the direction of projection and to the f_t axis,

$$f_\theta \equiv \cos\left(\theta + \frac{\pi}{2}\right) f_x + \sin\left(\theta + \frac{\pi}{2}\right) f_y. \qquad (12)$$

A bowtie is clearly present in Fig. 5 when $\theta = \theta_b = 0$. As θ deviates from $\theta_b = 0$, the bowtie gradually diminishes.

Note that aliasing effects are visible in Fig. 5 when $\theta = 0$, that is, when the bowtie wraps around the boundaries of the plot. Aliasing is due to the edges of the spheres which are represented with floating point accuracy. For real image sequences, aliasing effects are reduced because of optical blurring at the sensor level prior to spatial sampling. An example is the following.

4.1.2. Example 2: Holly Bush (Real Video). Figure 6 shows two frames from a 128 frame sequence of a real holly bush seen by a camera moving horizontally on a sliding platform. Figure 7 shows the projected spectra for various values of θ. The bowtie appears at $\theta = \pi/2$ as expected because the motion is horizontal. Aliasing effects are less severe than in the falling spheres of Fig. 5.

4.1.3. Example 3: Tracking a "Snowflake" (Synthetic Video). In Examples 1 and 2 above, the slopes of the planes of the bowtie all have the same sign.[2] Example 3 shows how both positive and negative slopes are possible. Consider again the synthetic falling sphere sequence but now we rotate the virtual camera about the x-axis direction as the camera translates in the y direction. The rotation is such that the optical axis passes through a 3D point near the center of the view volume in all frames, that is, the camera "tracks" this 3D point. This corresponds to case (ii) of parallel optical snow described in Section 3.2. Objects on the near side of the tracked point have downward image velocity and objects on the far side of the tracked point have upward image velocity.

Figure 8 shows XY and YT plots as well as the projected spectrum where the projection is along the bowtie axis. In the YT plot, the slopes are positive for the background objects and negative for the foreground objects. In the projected bowtie, the speeds are both positive and negative.

Figure 6. Two frames from a real image sequence of a holly bush taken from a horizontally translating camera. Image sequence consists of 128 frames of size 240 × 256 pixels. Frames were padded with zeros to give 256 × 256 square images. Frames 21 and 61 are shown.

5. Occlusions

The bowtie model assumes linear superposition of the motion planes. Strictly speaking this assumption does not hold when occlusions are present as in Examples 1–3. When an object is occluded, the power of that object is spread outside of that object's motion plane. The bowtie data from Examples 1–3 suggest that the bowtie signature is robust to occlusion effects. In this section, we briefly discuss why. See Fleet and Langley (1994) and Beauchemin and Barron (2000) for further discussion of how occlusions affect a motion plane.

Consider an image sequence that contains a large number of well-defined opaque objects, translating at a range of image speeds such as in Examples 1 to 3. We can partition the space-time volume of such an image sequence into the set of space-time volumes of the individual objects,

$$I(x, y, t) = \sum_i I_i(x, y, t)\chi_i(x, y, t),$$

where $I_i(x, y, t)$ is the space-time intensity function of the translating object i and $\chi_i(x, y, t)$ is a binary windowing function that defines the space-time volume in which object i is visible. Assume each object is opaque so that the domains of the $\chi_i(x, y, t)$ are disjoint and using Parseval's relation, i.e.

$$\int I(u)^2 du = \int \hat{I}(f)^2 df,$$

we see that the power spectrum of $I(x, y, t)$ must be equal to the sum of the power spectra of the space-time volumes of the individual objects. Let us examine each of these individual power spectra.

Assume object i is translating in the image with velocity (v_x, v_y) and so $\hat{I}_i(f_x, f_y, f_t)$ lies on a motion plane obeying Eq. (5). The Fourier transform of $I_i(x, y, t)\ \chi_i(x, y, t)$ is the convolution product, $\hat{I}_i(f_x, f_y, f_t) * \hat{\chi}_i(f_x, f_y, f_t)$. The effect of the window $\chi_i(x, y, t)$ in the frequency domain is to blur the motion plane. What is the width of the blurring kernel?

If object i is visible for a time duration Δ then, by the Uncertainly Principle, the Fourier transform of $\chi_i(x, y, t)$ has a temporal frequency width proportional to Δ^{-1}, and so the motion plane of object i is blurred by a width proportional to Δ^{-1} in the f_t direction. A similar argument holds for the spatial domain. If object i is visible over a distance Δ in the x or y directions, then the motion plane is blurred by width proportional to Δ^{-1} in the f_x or f_y directions, respectively.

The accuracy of the bowtie model thus depends on the space-time durations over which the objects are visible. In Examples 1–3, most objects are visible over a large number of frames and so the widths of the blurring kernels are small. This is the reason why the blurring of the motion planes has little effect on the bowtie signature. One could construct examples in which each point on each object is visible for a small number of frames only, and this gets one into the scope of "non-Fourier" models used by visual psychophysists (Fleet and Langley, 1994). Whether such optical snow motions produce a percept of motion to human observers is not known. See Qian et al. (1994) for experiments that address related questions, in the context of two layer transparency.

6. Estimating the Motion

The visualization of the bowtie signature presented in Section 4 suggests an algorithm for estimating the parameters of a parallel optical snow motion. Recall

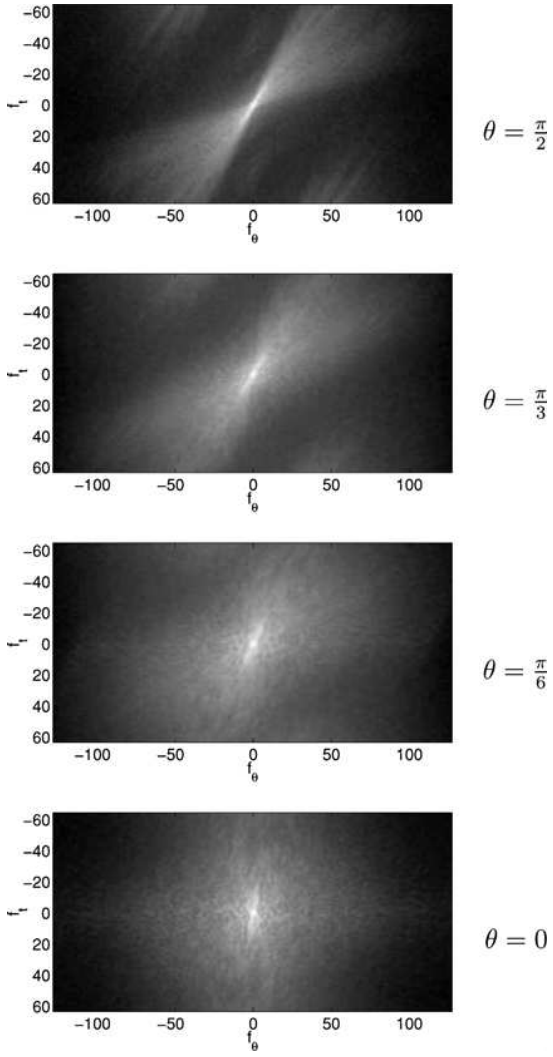

Figure 7. Projected power spectrum for holly sequence in Fig. 6. The bowtie signature appears at $\theta = \pi/2$ which corresponds to horizontal motion.

that θ_b is the direction of the axis of the bowtie (see Fig. 9(a)). The algorithm consists of two stages. First, estimate the angle θ_b. This yields the axis of the bowtie and thus the direction (v_x, v_y) of the motion which is perpendicular to the axis of the bowtie. Second, estimate the range of speeds present in the motion.

6.1. *Estimating the Direction*

To estimate the direction θ_b, we sum the power in a wedge of frequencies where the wedge is oriented at

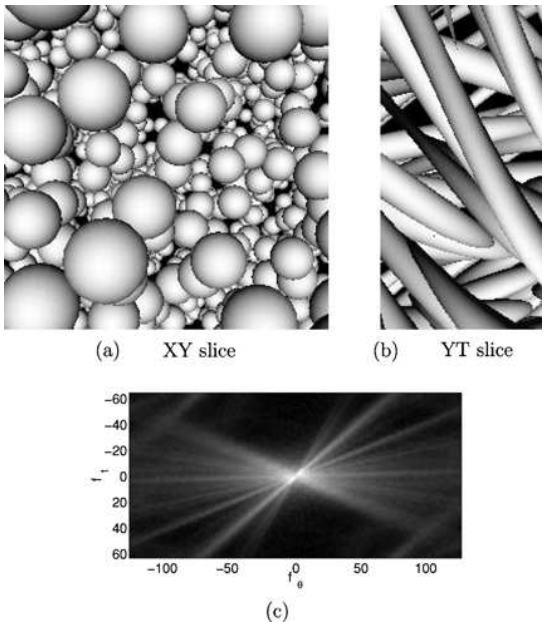

Figure 8. A sequence similar to that shown in Fig. 4, except that as the camera moves upwards, it rotates about the x-axis to "track" a point near the center of the view volume. (a) XY slice. (b) YT slice. (c) Bowtie signature ($\theta = \theta_b = 0$). Positive and negative speeds are present.

an angle θ as in Fig. 9(b). The wedge is defined by a constant slope $v_{\max}$ which is chosen by the user. This slope corresponds to an upper bound on speeds that are assumed to be present in the image sequence. Formally, we define $W(\theta)$ to be the sum of all power in frequencies (f_x, f_y, f_t) that satisfy

$$\left|\frac{f_t}{f_\theta}\right| > v_{\max}\frac{T}{N}$$

where f_θ was defined in Eq. (12), $v_{\max}$ is measured in pixels/frame, N is the image width (in pixels) and T is the number of frames in the sequence. The scaling factor $\frac{T}{N}$ is necesary when the number of pixels is different than the number of frames in the sequence.

We further restrict the sum to frequencies satisfying

$$\|(f_\theta, f_t)\|_2 > \tau$$

which removes frequencies with low f_θ and f_t. The speed of such frequencies is poorly defined because of quantization. This τ threshold "truncates" the wedge.

Before stating our next proposition, we observe that $W(\theta)$ has a 180 degree periodicity. We also observe that because $I(x, y, t)$ is real, $\hat{I}(f_x, f_y, f_t)$ and

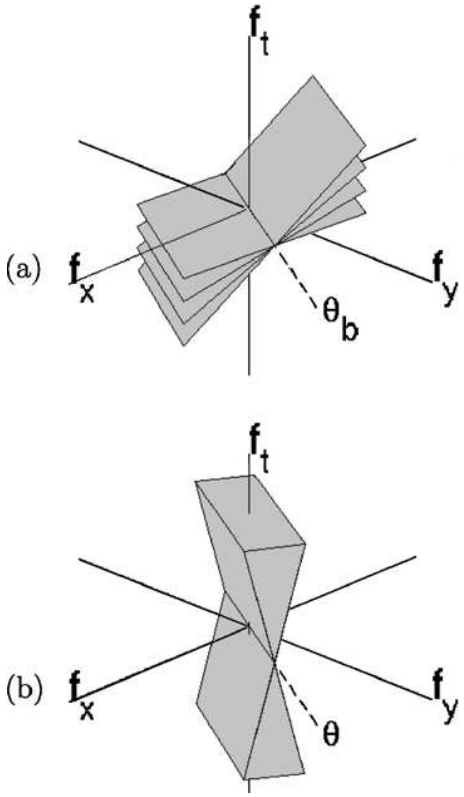

Figure 9. Parallel optical snow. (a) Bowtie signature in the frequency domain. The axis of the bowtie is in the f_x–f_y plane. The angle θ_b is measured from the x-axis. (b) Wedge used to estimate the orientation of the bowtie. The power within the wedge reaches a minimum when the wedge is aligned with the bowtie in (a) (i.e., $\theta = \theta_b$).

$\hat{I}(-f_x, -f_y, -f_t)$ are complex conjugates and so the power in the upper wedge is the same as the power in the lower wedge.

Proposition 4. *Suppose that power spectrum of an image is a family of planes of the form of Eq.* (10) *and that* $|(\alpha\tau_x, \alpha\tau_y)| < v_{max}$ *for all* α. *Then,* $W(\theta) = 0$ *when* $\theta = \theta_b$.

The proof follows directly from the definition of $W(\theta)$.

One caveat is that, when occlusions are present (recall Section 5), the motion planes are thickened slightly and so the pre-conditions for Proposition 4 may not hold exactly. In this case, a minimum of the function $W(\theta)$ occurs roughly but perhaps not exactly at $\theta = \theta_b$ and this minimum is non-zero.

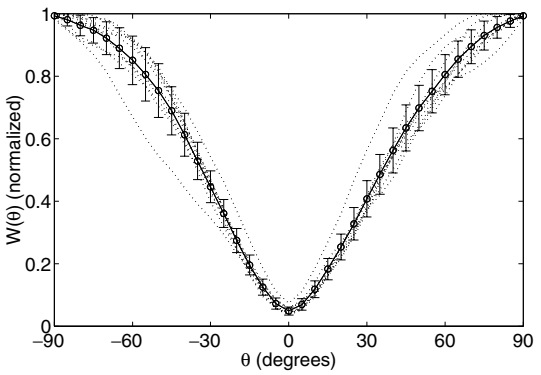

Figure 10. $W(\theta)$ for the synthetic sequence in Example 1. The minimum occurs at $\theta = 0^\circ$ corresponding to vertical motion. The results of ten runs are shown (dotted lines), along the the mean response (solid lines) and error bars (at one standard deviation).

Analogous to the visualization of Section 4, we compute $W(\theta)$ for a given image sequence by rotating through angles θ at fixed angular increments.[3] Figure 10 shows $W(\theta)$ for the sequence in Example 1. In this and subsequent experiments, we chose $v_{\max} = 8$ pixels/frame (slope of 4, since $\frac{N}{T} = 2$) and $\tau = 8$ pixels. We show response curves for ten different sequences, as well as the mean curve and the error bars at one standard deviation. Note that while there is some variability in the response, $W(\theta)$ has a well defined minima at $\theta = 0^\circ$.

Figure 11 shows the plot of $W(\theta)$ for real data, collected from the holly bush image sequence in Example 2. As expected, there is a well-defined minimum at $\theta = \pm 90^\circ$, which corresponds to the horizontal motion.

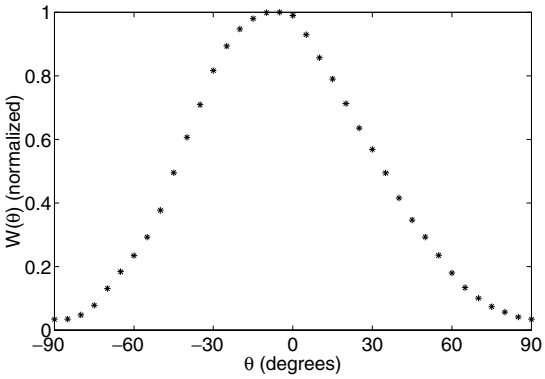

Figure 11. $W(\theta)$ for the holly sequence of Example 2. The minimum occurs at $\theta = \pm 90^\circ$, corresponding to horizontal motion.

6.2. *Estimating the Range of Speeds*

In addition to estimating the motion direction, we also may want to estimate the range of speeds within an image region. The range of speeds is related to the range of depths of visible surfaces in the region, and so it could be used to estimate this range of depths. The range of speeds could also be used to distinguish optical snow from other classes of motion, such as frontoparallel optical flow or a small number of discrete motion layers.

The range of speeds can be estimated directly from the projected power spectrum at the minimum of $W(\theta)$, by making a histogram of power in this projected power spectrum as a function of speed. As a proof-of-concept demonstration, we fit single Gaussian distribution to a histogram and report mean and variance of the speeds. More sophisticated models (e.g., a mixture of Gaussians) may be used for discriminating optical snow from motion with a finite number of layers.

Figure 12 shows two histograms of power as a function of speed. The plot in Fig. 12(a) is from the bowtie of Fig. 5(a) (Example 1) and the plot in Fig. 12(b) is from the bowtie of Fig. 8 (Example 3). The histograms of speeds in the two cases are approximately related by a shift. This shift is due to the rotation of the camera in Example 3 which adds an approximately constant component to the image speed of each object (recall Section 3.1). The mean speeds in (a) and (b) are approximately 1.0 and .14 pixels/frame, and the standard deviations are approximately 1.3 and 1.5, respectively.

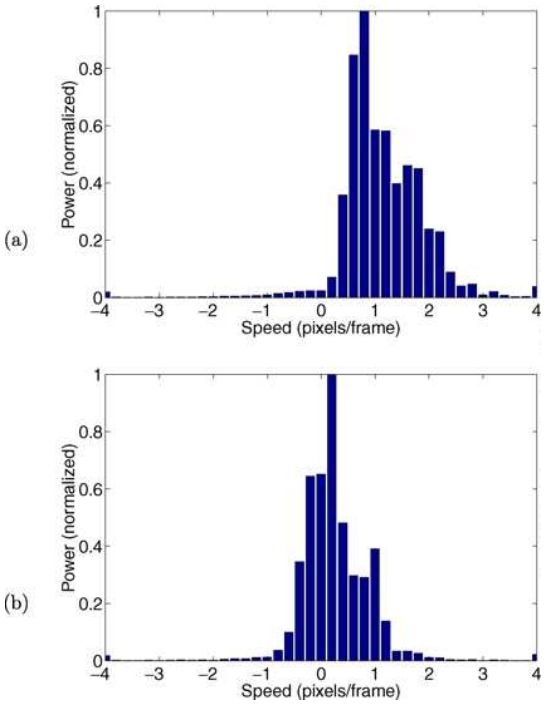

Figure 12. Image power as a function of speed for Examples 1 and 3.

7. Aperture Problem

In many scenes, the objects are dominated by a particular orientation. For example, in a crowd scene or a forest, the dominant orientation is vertical. When the camera moves laterally relative to such a scene, the measurable components of image velocity are in the direction that is normal to the oriented structure. Classically, this is known as the *aperture problem* (Marr and Ullman, 1981). How is this problem manifest in optical snow?

The aperture problem can be expressed in terms of Eq. (2) as follows. Let the spatial gradient of the image intensities be parallel to a constant vector $(\cos\theta_n, \sin\theta_n)$ where the subscript n stands for "normal direction". Formally, we are supposing that

$$\left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right)(x, y, t) = g(x, y, t)\ (\cos\theta_n, \sin\theta_n) \quad (13)$$

where $I = I(x, y, t)$ and where $g(x, y, t)$ is the magnitude of the image gradient,

$$g(x, y, t) = \left|\left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\right)\right|.$$

Proposition 5. *If the direction* $(\cos\theta_n, \sin\theta_n)$ *of the image gradient is constant over* (x, y, t)*, then the maximum of* $W(\theta)$ *occurs when* $\theta = \theta_n$*.*

Proof: We use the derivative property of Fourier transforms. Applying Eq. (3) to Eq. (13) yields

$$2\pi i\ \hat{I}(f_x, f_y, f_t)\ (f_x, f_y) = \hat{g}(f_x, f_y, f_t)(\cos\theta_n, \sin\theta_n).$$

For any (f_x, f_y, f_t), if $\hat{I}(f_x, f_y, f_t) \neq 0$ then (f_x, f_y) is parallel to $(\cos\theta_n, \sin\theta_n)$, i.e.

$$-\sin\theta_n f_x + \cos\theta_n f_y = 0. \quad (14)$$

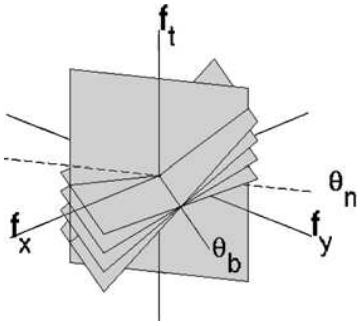

Figure 13. If θ_n is the direction of the gradient of image intensities for all points in the image sequence, then the power is restricted to the intersection of the bowtie with the vertical plane of Eq. (14).

Thus, if the image gradient satisfies Eq. (13), then all power in the frequency domain must lie in the plane of Eq. (14). The geometry of the intersection of the bowtie and the plane of Eq. (14) is illustrated in Fig. 13.

Recall the wedge detector from Fig. 9(b). When $\theta = \theta_n$, the axis of the wedge lies in the plane. By inspection, the plane must be entirely contained in the wedge. Hence, a maximum of $W(\theta)$ occurs when $\theta = \theta_n$. □

This is the aperture problem for optical snow. Let us explore the aperture problem by examining a few examples.

7.1.1. Example 4: Parallel Cylinders. Consider the example of a 3D scene consisting of long parallel cylinders like trees in a forest. Suppose these cylinders are perpendicular to the optical axis of the camera so that the cylinders are parallel in the image plane. Let θ_n be such that $(\cos\theta_n, \sin\theta_n)$ is normal to the axis of the cylinders. The spatial gradient of intensities in the image is thus in direction $(\cos\theta_n, \sin\theta_n)$ and Eqs. (13) and (14) hold. For example, if the trees are vertical in the image then $\theta_n = 0^\circ$ and $f_y = 0$.

If the camera translates laterally through such a cluttered scene, optical snow results. From Proposition 1, the minimum of $W(\theta)$ occurs at θ_b. From Proposition 5, the maximum of $W(\theta)$ occurs at θ_n. We emphasize that there is no relationship between the angle θ_n and the angle θ_b. The angle θ_n is normal to the cylinders and the angle θ_b is the direction of the true image velocity.

Figure 14 shows $W(\theta)$ for a synthetic video of such cylinders. The scene is similar to that of Example 1 except now the objects are cylinders (of radius 0.2) rather than spheres. The cylinders are tilted -30° from the horizontal axis and so $\theta_n = 90^\circ - 30^\circ = 60^\circ$.

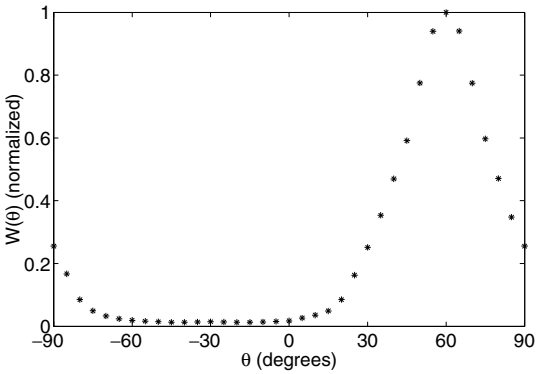

Figure 14. $W(\theta)$ for vertically falling cylinders tilted at 30 degrees from horizontal. The plot shows the average response over ten runs.

Note that $W(\theta)$ is not a delta function. The reason is that, for $\theta = \theta_n$, the plane containing the power lies entirely in the wedge. But for nearby θ, the plane still intersects the wedge, and there is power lying in this intersection.

7.1.2. Example 5: Parallel Ellipsoids. A more general synthetic case is a cluttered scene containing 3D elongated ellipsoids. Suppose that the major axis of each 3D ellipsoid is parallel to the image plane, and assume the aspect ratios of axes are axes is $m : 1 : 1$ where $m \in \{1, 2, 4, 8\}$. Consider vertical camera motion as in Example 1.

Example 5 is just a generalization of Examples 1 and 4 in which the spheres or cylinders are replaced by elongated ellipsoids. The major axis of the ellipsoids plays the same role as the axis of the cylinders in Example 4, namely we define $(\cos\theta_n, \sin\theta_n)$ to be normal to the major axis.

Figure 15(a) shows the functions $W(\theta)$ for four such scenes whose ellipsoids differ in aspect ratio. Each plot shows the mean of $W(\theta)$ over ten example sequences. For each ellipsoid, the major axis is horizontal. Because the motion is vertical, the minimum of $W(\theta)$ is at 0°. Figure 15(b) shows the mean of $W(\theta)$ when the major axis of the ellipsoids is -30° from the horizontal. (This tilt has no effect when the aspect ratio is 1.) Since θ_n is perpendicular to the major axis, and since the major axis is 30 degrees away from the horizontal direction, we have $\theta_n = 90^\circ - 30^\circ = 60^\circ$.

To understand these plots, recall Proposition 4 from Section 6.1 which says that direction of motion is indicated by the minimum of $W(\theta)$. For ellipsoids with

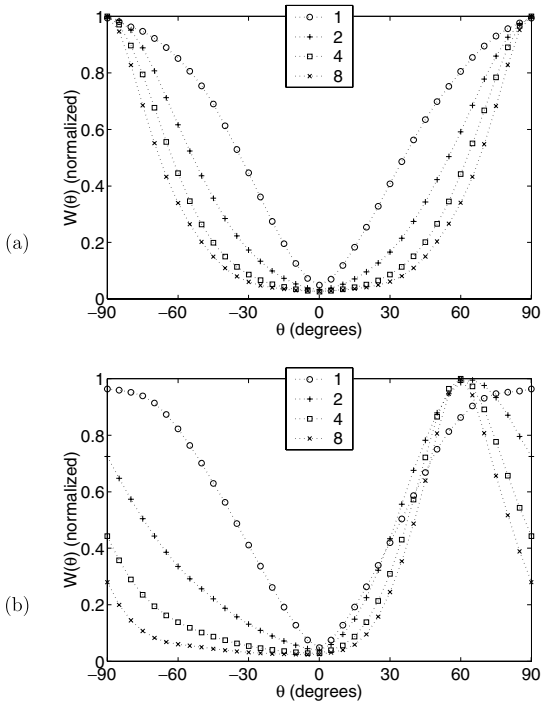

Figure 15. Plots of $W(\theta)$ for sequences containing elongated ellipsoids. The aspect ratios (y/x) are 1.0 (spherical), 2, 4, and 8. (a) Plots of $W(\theta)$ for vertically moving ellipsoid sequences, with horizontal major axis. (b) Plots of $W(\theta)$ when major axis is 30° from the horizontal. The minimum of $W(\theta)$ remains at 0°, but the maximum shifts as the aspect ratio increases. The plots show the average response over ten sequences.

a large aspect ration m, the minimum is broad because that power is concentrated about the normal velocity direction. Because of the broad minimum, the true direction of motion is difficult to estimate.[4] By contrast, the direction of the normal velocity is relatively well defined by the maxima of $W(\theta)$.

For the example of Fig. 15(b), when the aspect ratio is one, the maximum of $W(\theta)$ occurs at 90 degrees which is 90 degrees away from the axis of the bowtie for vertical motion. As the aspect ratio increases, the position of the maximum of $W(\theta)$ gradually shifts from 90° towards $\theta_n = 60°$.

Before moving to the next example, we note that humans observing these sequences of falling ellipsoids can easily judge the correct (vertical) motion direction. As long as the endpoints of a long thin ellipsoid are visible, human observers do not confuse the dominant orientation of the ellipsoids with the true direction of motion. Presumably the reasons are twofold. First, strictly speaking, the pre-conditions of Proposition 5 do not apply to the ellipsoidal sequences since intensity gradients are present at all orientations. Second, human observers see ellipsoids and their boundaries. Despite the occlusions, they can follow the true direction of motion from these boundaries. The remarks do not trivialize the computational problem, of course. Finding object boundaries in the presence of dense occlusion in optical snow is an open and difficult computer vision problem.

7.1.3. Example 6: A Real Forest. We next consider a real scene containing oriented structure. An image sequence of a densely treed scene was shot using a Hitachi MPEG MP-EG10W camera pointing out a side window of a forward moving car. The trees in the scene were bare of leaves. Though the trunks of the trees were vertical, visible branches of the trees appeared at a range of angles. A single frame from the sequence is shown in Fig. 16.

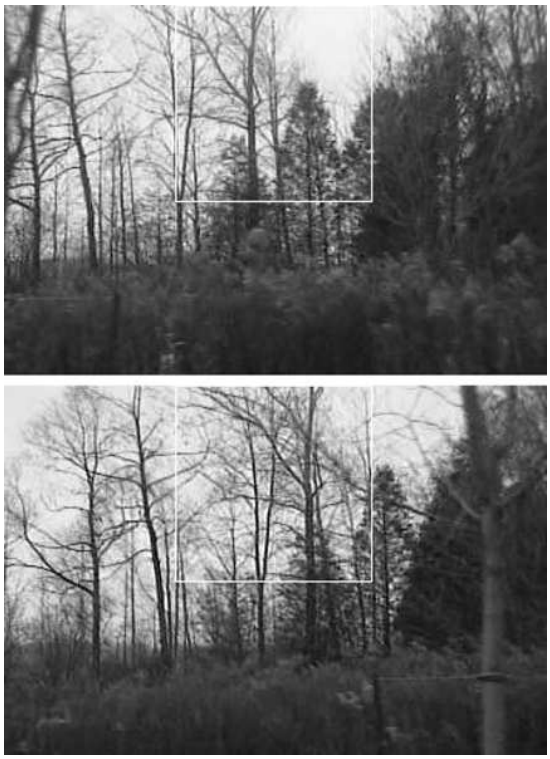

Figure 16. Two frames from an image sequence of a forested area taken with a horizontally translating camera. Image sequence was cropped to 128×128 pixels per frame. The cropped area is outlined.

The camera produces images of size 352 × 240 at 30 frames/sec. A sequence of 128 frames was extracted and converted to grey scale images. Each frame was cropped to an area of 128 × 128 pixels (see outlined region in Fig. 16).

Figure 17 shows the projected power for various angles of θ. The bowtie signature is clearly visible at $\theta = \pi/2$. Figure 18 shows $W(\theta)$. The minimum occurs at $\pm 90^\circ$ which is as expected since the motion is horizontal ($\theta_b = \pm 90^\circ$). The maximum occurs at approximately 0° since the dominant orientation is vertical.

The width of the peak of $W(\theta)$ at half-height is approximately 50 degrees and so the width of the valley is 130 degrees. This broad minimum is due to the vertical structure (tree trucks) which domanates the non-vertical structure (branches) in the scene. If we compare the plot in Fig. 18 with those in Fig. 15(a), we see that a half-height peak-width of 50 degrees corresponds loosely to an effective aspect ratio of about four.

To summarize, the above examples show that the function $W(\theta)$ provides both the direction of normal velocity in the image and an estimate of the range of spatial orientations in the image that are contributing to the motion (the aperture problem). When a large range of spatial orientations is indeed present, the dominant direction of motion can be estimated by the minimum $W(\theta)$. When only a small range of spatial orientations is present, the power spectrum does not provide a good estimate of the true motion direction, but rather could only provides an estimate of the normal velocities.

8. Discussion

Our results in this paper broaden the analysis of image motion to a new natural motion category which we

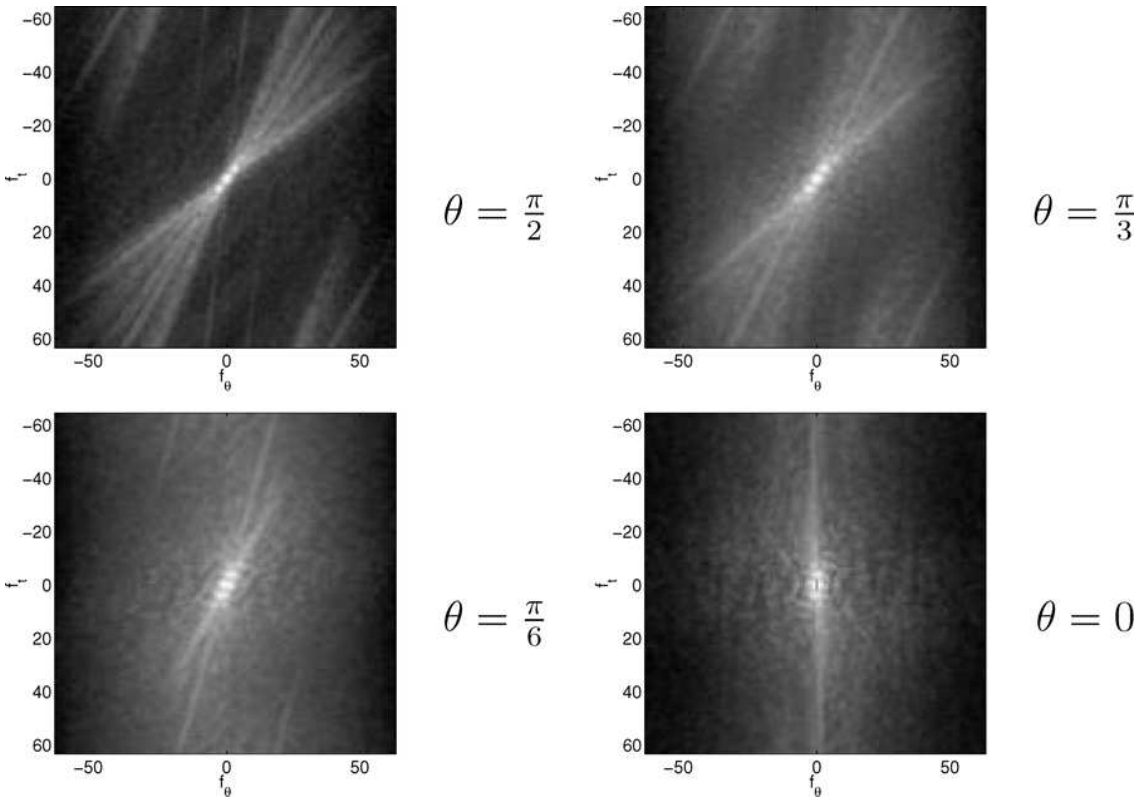

Figure 17. Projected power spectrum for forest sequence in Fig. 16. The bowtie signature appears at $\theta = \pi/2$ which corresponds to horizontal motion.

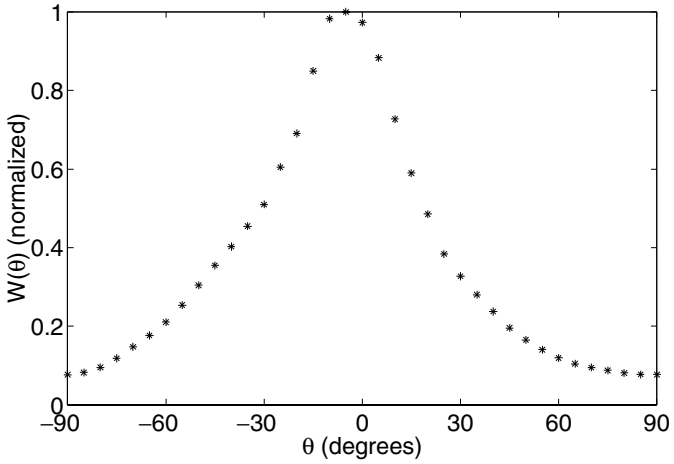

Figure 18. $W(\theta)$ for the forest sequence in Fig. 16. The minima occurs at $\pm 90^{\circ}$, corresponding to horizontal motion.

call optical snow. We concentrate on the case of parallel optical snow. We have presented an algorithm for estimating the parameters of motion for this case. Parallel optical snow arises in many natural situations, for example, tracking a stationary object as the observer moves through a cluttered scene or a statinary observer viewing falling snow.

One problem that is ripe for future work is how to estimate the bowtie using methods that are more biologically plausible, for example, methods based on local space-time energy detectors. Such methods could be derived from those of Heeger (1987), Fleet (1992), Grzywacz and Yuille (1990), Simoncelli and Heeger (1998), and Huang and Chen (1995) which estimate a single motion plane in the frequency domain for the case of constant translation optical flow. Indeed they would generalize such methods, since a single motion plane is just a bowtie whose range of speeds is limited to a single speed. Biologically plausible models of layered motion transparency have been proposed e.g. Zemel and Dayan (1999) and we expect similar algorithms could be designed for optical snow as well.

With such algorithms in mind, the reader may be inclined to ask why we have presented an algorithm that is based on a global Fourier analysis in the first place. The answer is that we feel the optical snow problem is mostly cleanly expressed using the Fourier transform. As articulated by Marr (1982), it is important to distinguish a computational problem from a particular algorithm for solving the problem. The computational problem that we posed and solved was to recover the parameters of the bowtie pattern for parallel optical snow. The bowtie pattern itself is derived from a natural constraint, namely that lateral motion of an observer relative to a cluttered scene yields velocity vectors that lie on a line in 2D image velocity space. This constraint relies neither on spatial continuity of the motion field (as in optical flow), nor does it rely on smoothness in layers (as in layered motion). It does not rely on any local structure in the image whatsoever. Hence, a global Fourier approach is natural way to pose and solve the problem.

That said, we admit that the global Fourier algorithm has significant limitations. For example, the algorithm alone is unable to distinguish optical snow from other motion categories such as optical flow or layered motion. When an observer moves laterally through a scene that contains a single horizontal ground plane only, the motion field is a continuous horizontal shear. An affine model with a single layer would be the natural choice here for processing the motion. But since all velocity vectors are horizontal (and hence parallel) and a range of speeds is present, the optical snow algorithm would yield a correct answer as well. It would recover the direction of motion and the range of speeds. It would not recover, however, the systematic spatial relation between speed and image position which the affine model would recover.

This last example illustrates the importance of selecting an appropriate motion model to interpret a given

sequence. This model selection problem is a poorly understood one. A solution to the problem would need to incorporate many models of image motion, such as optical flow, layered motion, optical snow, and other motion categories (Langer and Mann, 2001; Wildes and Bergen, 2000; Davis et al., 2000). A solution would also need to consider issues of motion segmentation and grouping since different models might be appropriate at different regions of the image. Though the challenges in this model selection problem are awesome and enormous, inevitably they must be faced if we are to develop computer vision systems that can interpret unknown raw motion sequence with a competence comparable to that of existing biological motion systems.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Notes

1. See the interesting discussion of the fox in the forest in Zucker and Iverson (1987).
2. In both the input image sequences (Figs. 4 and 6) and the projected power spectra (Figs. 5 and 7) the vertical axis points downward. According to this convention, both bowties are comprised of planes with negative slopes.
3. For all of the experiments reported here $W(\theta)$ was computed at 5 degree angular increments.
4. For these particular scenes, the true direction of motion could be estimated by tracking individual ellipsoids. This is similar to solving the aperture problem in classical optical flow by tracking end points or corners of lines.

References

Adelson, E. and Bergen, J. 1985. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2):284–299.

Barron, J., Fleet, D., and Beauchemin, S. 1994. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77.

Beauchemin, S. and Barron, J. 2000. The frequency structure of 1D occluding image signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2):200–206.

Bergen, J., Burt, P., Hingorani, R., and Peleg, S. 1992. A three-frame algorithm for estimating two-component image motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):886–896.

Black, M. and Anandan, P. 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow-fields. *Computer Vision and Image Understanding*, 63(1):75–104.

Bolles, R.C., Baker, H.H., and Marimont, D.H. 1987. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1:7–55.

Darrell, T. and Pentland, A. 1995. Cooperative robust estimation using layers of support. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):474–487.

Davis, J., Bobick, A., and Richards, W. 2000. Categorical representation and recognition of oscillatory motion patterns. In *IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head, SC, pp. 628–635.

Fennema, C. and Thompson, W. 1979. Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, 9:301–315.

Fleet, D.J. 1992. *Measurement of Image Velocity*. Norwell, MA: Kluwer Academic Press.

Fleet, D.J. and Langley, K. 1994. Computational analysis of non-fourier motion. *Vision Research*, 34(22):3057–3079.

Gibson, J.J. 1950. *The Perception of the Visual World*. Boston: Houghton Mifflin.

Grzywacz, N. and Yuille, A. 1990. A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London B*, 239:129–161.

Heeger, D. 1987. Optical flow from spatiotemporal filters. In *First International Conference on Computer Vision*, London, UK, pp. 181–190.

Horn, B. and Schunck, B. 1981. Determining optical flow. *Artificial Intelligence*, 17:185–203.

Huang, C. and Chen, Y. 1995. Motion estimation method using a 3d steerable filter. *Image and Vision Computing*, 13:21–32.

Irani, M., Rousso, B., and Peleg, S. 1994. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1):5–16.

Jepson, A. and Black, M. 1993. Mixture models for optical flow computation. In *IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, pp. 760–761.

Ju, S., Black, M., and Jepson, A. 1996. Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency. In *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, pp. 307–314.

Koenderink, J.J. 1986. Optical flow. *Vision Research*, 26(1):161–179.

Langer, M.S. and Mann, R. 2001. Dimensional analysis of image motion. In *IEEE International Conference on Computer Vision*, Vancouver, Canada, pp. 155–162.

Langer, M.S. and Mann, R. 2002. Tracking through optical snow. In *Lecture Notes in Computer Science 2525: Biologically Motivated Computer Vision, 2nd International Workshop*, H.H. Buelthoff, S. Lee, T. Poggio, and C. Wallrave (Eds.), Springer, pp. 181–188.

Lappe, M. and Rauschecker, J.P. 1993. A neural network for the processing of optical flow from egomotion in man and higher mammals. *Neural Computation*, 5:374–391.

Longuet-Higgins, H. and Prazdny, K. 1980. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London*, B-208:385–397.

Mann, R. and Langer, M.S. 2002. Optical snow and the aperture problem. In *International Conference on Pattern Recognition*, vol. IV, Quebec City, Canada, pp. 264–267.

Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman.

Marr, D. and Ullman, S. 1981. Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society London Ser.*, B-211:151–180.

Milanfar, P. 1996. Projection-based, frequency-domain estimation of superimposed translational motions. *Journal of the Optical Society of America A*, 13(11):2151–2162.

Qian, N., Anderson, R.C., and Adelson, E.H. 1994. Transparent motion perception as detection of unbalanced motion signals i: Psychophysics. *Journal of Neuroscience*, 14:7357–7366.

Sawhney, H. and Ayer, S. 1996. Compact representations of videos through dominant and multiple motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):814–830.

Shizawa, M. and Mase, K. 1991. A unified computational theory for motion transparency and motion boundaries based on eigenenergy analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, Maui, pp. 289–295.

Simoncelli, E.P. and Heeger, D.J. 1998. A model of neural responses in visual area MT. *Vision Research*, 38(5):743–761.

Trucco, E. and Verri, A. 1998. *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall.

Wang, J. and Adelson, E. 1993. Layered representation for motion analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, pp. 361–366.

Warren, W.H. and Hannon, D.J. 1990. Eye movements and optical flow. *Journal of the Optical Society of America A*, 7(1):160–169.

Watson, A. and Ahumada, A. 1985. Model of human visual-motion sensing. *Journal of the Optical Society of America A*, 2(2):322–342.

Waxman, A. and Wohn, K. 1985. Contour evolution, neighbourhood deformation and global image flow: Planar surfaces in motion. *International Journal of Robotics Research*, 4:95–108.

Weiss, Y. 1997. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, PR, pp. 520–526.

Wildes, R. and Bergen, J. 2000. Qualitative spatiotemporal analysis using an oriented energy representation. In *Sixth European Conference on Computer Vision*, Dublin, Ireland, pp. 768–784.

Zemel, R.S. and Dayan, P. 1999. Distributional population codes and multiple motion models. In *Advances in Neural Information Processing Systems 11*, M.S. Kearns, S. A. Solla, and D.A.C. (Eds.), Cambridge, MA, MIT Press, pp. 768–784.

Zucker, S.W. and Iverson, L. 1987. From orientation selection to optical flow. *Computer Vision Graphics and Image Processing*, 37:196–220.