

Dimensional analysis of image motion

Michael S. Langer
School of Computer Science
McGill University
Montreal, Canada

Richard Mann
Dept. of Computer Science
University of Waterloo
Waterloo, Canada

Abstract

Studies of image motion typically address motion categories on a case-by-case basis. Examples include a moving point, a moving contour, or a 2D optical flow field. The typical assumption made in these studies is that there is a unique velocity at each moving point in the image. In this paper we relax this assumption. We introduce a broader set of motion categories in which the set of motions at a moving point can be 0-D, 1-D, or 2-D. We consider one new motion category in detail, which we call optical snow. This motion category occurs, for example, when an observer translates relative to a massively cluttered scene. Examples include the motion seen by an observer moving through bushes, or falling snow seen by a stationary observer. Optical snow is characterized by a 1-D set of velocities at each moving point and, as such, it cannot be analyzed using a classical computational method such as optical flow. We introduce a technique for analyzing optical snow which is based on a bow tie signature of the motion in the frequency domain. We demonstrate the effectiveness of the technique using both synthetic and real image sequences.

1 Introduction

This paper addresses categorical properties of time varying images in which the variation is due to motion. We consider motion to be changes in the spatial position of image tokens over time. Image tokens include raw pixel intensities, as well as abstract entities such as the output of edge and line detectors.

Traditional models of image motion address the case of a moving point, a moving curve, or a flow over a region. Such motions are distinguished from each other by the dimension of the set of moving image points: A single point is a 0-D set, a curve is a 1-D set, and a region is a 2-D set. In each of these three cases, one assumes there is a unique velocity at each moving point, and one poses the computational problem as, “how do you recover this unique velocity?”

In this paper, we go beyond the traditional image motion problem by considering a more general set of motions in which the set of velocities at each moving point can be

1-D or 2-D. A 1-D set of velocities at a moving point means that the velocities define a curve in image velocity space (with the curve possibly differing from one moving image point to another). A 2-D set of velocities at an image point means that the velocities define a region in velocity space.

We treat image motion as a mapping from the domain of image positions to the range of image velocities. The novelty of our approach is to allow the subsets of velocity space to be 1-D or 2-D sets in addition to the traditional 0-D set.

The paper makes two contributions to the study of image motion within this theme of dimensional analysis. The first contribution, developed in Section 2, is a formal model of image motion categories which are defined by the dimensionality of the motion. The motion categories include familiar motions, as well motions that have not yet been considered in which the set of velocities at a moving point is 1-D or 2-D.

The second contribution, developed in Section 3, is a detailed analysis of one of the new categories of motion. We call this motion *optical snow*. Examples are the motion of trees and branches in a forest as seen by a translating observer, or the motion of rain or falling snow as seen by a stationary observer. Optical snow produces a range of image velocities at each point of the image. Such motions fail to satisfy the classical optical flow assumption that there is a unique image velocity at each pixel. As such, classical optical flow methods are not appropriate for representing these motions. We introduce an alternative technique for analyzing optical snow, and we provide examples that illustrate the feasibility of the technique.

2 Motion Categories

The idea of analyzing and comparing motion categories has been pursued previously and relationships between several categories of motion have been successfully explored, *e.g.* [1]. The novelty of our approach is to consider a larger set of categories than has been considered previously, and to explicitly address the issue of dimensionality [2].

2.1 Dimensional Analysis

To introduce the idea of our dimensional analysis, we consider the possible dimensions of the set of moving points in an image and the set of velocities of each of these moving points. There are three possible dimensions of the set of moving points, namely 0-D, 1-D, and 2-D and there are three possible dimensions of the set of image velocities at each moving point, also 0-D, 1-D, and 2-D. Taking all 3×3 combinations yields nine possible dimensional categories of motion. These are listed in Table 1. Categories 1–3, 4–6, and 7–9 are motions in which the set of velocities at each moving point is 0-D, 1-D, and 2-D respectively.

Table 1: Nine categories of image motion.

#	dim	set of image points	set of velocities
1	0/0	point (0-D)	point (0-D)
2	1/0	curve (1-D)	point (0-D)
3	2/0	region (2-D)	point (0-D)
4	0/1	point (0-D)	curve (1-D)
5	1/1	curve (1-D)	curve (1-D)
6	2/1	region (2-D)	curve (1-D)
7	0/2	point (0-D)	region (2-D)
8	1/2	curve (1-D)	region (2-D)
9	2/2	region (2-D)	region (2-D)

Let us consider a few examples. Category 1 is the motion of a single point, tracing out a curve in space-time. An ideal example would be bird flying across the sky. Category 2 is the motion of a curve such that each point on the curve has a unique velocity. Such a curve traces out a surface in space-time. An example of such a moving image curve is the boundary of a moving object. Category 3 is the motion of 2-D image region such that each point in the region has a unique velocity. Such a motion traces out a volume in space-time. These first three categories are familiar to us. (See [2] for more examples.) Categories 4–9 are not so familiar and require elaboration.

2.2 Axis-aligned motions

We now consider a refinement of the motion categories in Table 1, which we call *axis-aligned motions*. For the axis aligned motions, each 1-D curve is either a horizontal line or a vertical line. This restriction to horizontal and vertical lines is quite natural, since the force of gravity in the physical world defines a natural coordinate system: objects fall vertically, cameras pan and translate horizontally, etc. Moreover, the restriction comes at little cost. By deforming (morphing) an axis-aligned motion, the 1-D curve becomes non-linear, and one obtains the more general corresponding motion in Table 1. The advantage of working with axis-

aligned motions is that they enable us to articulate several subtle distinctions that can occur within the categories of Table 1, involving either a 1-D set of moving image pixels or a 1-D set of velocities.

Axis-aligned motions are defined as follows. Consider the four coordinates x, y and v_x, v_y in the usual way. The coordinates x and y denote image positions, and the coordinates v_x and v_y denote components of an image velocity vector. An axis-aligned motion is one in which, for each of the four coordinates, x, y, v_x , and v_y , there is motion either at a large interval of values of that coordinate or at one value of that coordinate only.

Table 2 illustrates 2^4 (*i.e.* 16) binary labelings of image motion, according to the definition of axis-aligned motion. These labelings are 4-bit binary codes defined by the x, y, v_x , and v_y coordinates. For any row in Table 2 and for any column, a “1” means there is image motion over a large interval of values of that coordinate, and a “0” means there is image motion at one value of that coordinate only. (A similar model was used in [3] to represent lighting categories.)

In Table 2, we have grouped the sixteen binary labeling into the nine categories that we saw in Table 1. Two rows in Table 2 are in the same category if they have the same number of 1’s in their (x, y) columns and the same number of 1’s in the (v_x, v_y) columns. That is, the nine categories are determined by the dimension (0, 1, or 2) of the moving points in the image domain and by the dimension (0, 1, or 2) of the velocities at each of the moving points. For example, in Category 1/0, each of the rows has one 1 in the (x, y) columns and no 1’s in the (v_x, v_y) columns. In Category 1/1 each of the rows has one 1 in the (x, y) columns and one 1 in the (v_x, v_y) columns. Note that, in the case of 1-D sets of image points and/or motions, Table 2 distinguishes between horizontal and vertical alignments. The result is a subdivision of categories 2, 4, 5, 6, and 8 in Table 1. We will now describe each category in Table 2 in more detail and highlight several subtle distinctions and commonalities between categories.

Category 0/0

Category 0/0 is defined by image motion that occurs at a unique value of each of the coordinates x and y , and at a unique value of the coordinates v_x and v_y . An example is a point moving across a uniform background.

Note that this category of motion does not distinguish between motion paths such as linear, circular, or undulatory [4]. Such types of motions are sub-categories of Category 1. These motions are equivalent in our representation because they have the same dimensionality.

Category 1/0

Category 1/0 describes a moving curve in which the motion of each point on the curve is unique. In the axis-aligned

Table 2: Axis aligned motions. Categories representing instantaneous events are marked with *.

#	dim	Example	x	y	v_x	v_y
1	0/0	moving point	0	0	0	0
2	1/0	moving curve	1	0	0	0
			0	1	0	0
3	2/0	optical flow	1	1	0	0
4	0/1 *	line pointing to camera	0	0	0	1
			0	0	1	0
5	1/1 a *	moving past a corner	1	0	0	1
			0	1	1	0
	1/1 b	texture behind a linear aperture	1	0	1	0
			0	1	0	1
6	2/1	optical snow	1	1	1	0
			1	1	0	1
7	0/2 *	splash	0	0	1	1
8	1/2 *	exploding line	1	0	1	1
			0	1	1	1
9	2/2	speckle noise	1	1	1	1

case, the curve can be either a horizontal or vertical line: motion occurs for an large interval of x values and for a unique y (case 1000), or for a large interval of y and for a unique x (case 0100). In either case, for each point on the moving line, there is a unique velocity vector (a point in velocity space). This vector may vary from point to point along the moving line. An example of this type of motion would be the motion of the bounding contour of an object with a horizontal or vertical edge.

An important issue here concerns the ‘‘aperature problem.’’ If the points on the moving line can be distinguished from each other by different intensities each of which are constant over time, then the velocity at each point is well defined. If, however, the moving line is isoluminant in both space and time, then only the normal component of the velocity is well-defined. One could argue that there is a 1-D set of velocities at each point on a moving isoluminant line, since the image intensities constrain only the normal component of velocity. Historically, this aperture problem has been solved by imposing an additional constraint that there is a unique motion at every point on the moving line, for example, by considering the normal velocity only or by using a regularization constraint such as maximal smoothness. Once such a uniqueness constraint is imposed, one can speak of a Category 1/0 motion.

Category 2/0

Category 2/0 motion is defined by a single velocity at each point in the image, *i.e.* optical flow. Note that the

velocities may vary from point to point in the image. Moreover, motions of different topology such as diverging, translating, or rotating flows [5] are not distinguished. Such topological differences do not affect the dimensionality of the motion, and thus define sub-categories in our representation similar to what we saw in Category 0/0.

We next turn to motion categories in which there is a 1-D set of velocities at each moving point.

Category 0/1

Category 0/1 motion is rare but worth discussing for completeness. It can occur when a linear object rotates about a point in 3-D and the line of the object instantaneously coincides with the optical axis of the camera. At this instant, the image of the object collapses to a point. In the spatio-temporal neighborhood of this singularity points move in the same direction, but a range of image speeds results since the image speed is inversely proportional to the depth along the line. For axis aligned motions, the 1-D set of points is a horizontal line (case 0010) or a vertical line (case 0001).

This category of motion is instantaneous *i.e.* singular. Because of this special property, it is marked with an asterisk in Table 2. We will see other examples of instantaneous motion categories shortly.

Category 1/1

Category 1/1 is defined by motion of a curve (or along a curve) such that each point on the curve has a 1-D set of velocities. In the case of axis-aligned motions, two distinct and interesting examples emerge. The first (Category 1/1a) is a singular motion; it occurs instantaneously only. The second (Category 1/1b) occurs continually.

Category 1/1a

Category 1/1a is defined by two cases. In case 0110, the moving points in the image are on a vertical line and for each point on this vertical line there is a 1-D set of velocities that defines a horizontal line in velocity space. Case 1001 is the opposite. The moving points in the image are on a horizontal line and for each point on this horizontal line there is a 1-D set of velocities that define a vertical line in velocity space.

An example of the first case (0110) is the motion observed by a viewer who walks past an exterior convex corner of a brick building, such that the corner is defined by two outer walls. A singular event occurs at the instant the viewer crosses the extended plane of the far wall. At this instant, the far wall instantaneously becomes visible and a wide range of image velocities occurs. These velocities correspond to the 2-D set of points (e.g. bricks) on the wall that emerge suddenly into view. The continuum of horizontal image velocities is due to motion parallax *i.e.* the bricks on

the wall are at a range of distances from the observer. Since the bricks all become visible at the same instant and along the same vertical line in the image, a range of horizontal velocities occur at each point on this line and at that instant. Hence there is a 1D set of horizontal velocities at each point on the vertical line.

A similar example (the 1001 case) occurs when one climbs a set of stairs. At the instant the eye passes through the extended plane of the landing of the floor above, the horizontal image line that demarcates the landing becomes the source of a range of vertical velocities corresponding to points on the floor.

Notice that these motion events are quite different from those produced near the occluding contour of a rotating smooth surface, such as a rolling soccer ball, or the horizon as one drives a car over a hill. In these latter examples, points emerge continually from behind the occluding contour and the image velocity at each point is unique and continuous, with a limit of zero velocity at the occluding contour. Such motions are thus examples of optical flow (Category 2/0) rather than of Category 1/1a.

Category 1/1b

The second Category 1/1 motion occurs continually over time rather than instantaneously. Consider a textured background that is visible through a linear aperture such as a crack in a doorway. As an observer moves past the aperture, a 1-D speckle pattern is produced. This speckle is due to strips of the textured background as they appear/disappear.

Do such speckle patterns constitute “motion” ? One could argue they do not, since there are no image tokens whose positions are changing over time in a well-defined way, that is, there is no well-defined solution to the correspondence problem. On the other hand, one could argue they do, since such patterns would activate biological motion systems. Any “motion energy detector” [6] whose spatial receptive field overlaps the linear aperture and who is sensitive to velocity components parallel to the aperture would have a response, regardless of its speed tuning. In this sense, we consider that all image velocities parallel to the line present in this motion category.

Category 2/1

Category 2/1 is defined by having a 1D set of velocities at each point in the image. We call this motion *optical snow* and we will analyze it in detail in Section 3. For now, we present a few intuitive examples. One example is falling snow as seen by a stationary observer. All image motion is in the vertical direction and near any image position there is a wide range of vertical speeds. This range of speeds is due to the varying distances to the snowflakes visible to the eye. In the axis-aligned case, this is an example of 1101 motion. For a second example of Category 2/1 motion, consider an observer walking through a forest whose trees are bare of

leaves and whose branches extend in many directions. Further suppose the observer is looking off to the side as he walks. In this scenario, all image motion is in the horizontal direction and there is a range of image speeds at each image position, since multiple objects are visible and the image speed of each object is inversely proportional to its depth. In the axis-aligned case, this is an example of 1110 motion. In both of these examples, there is image motion at each point in the image and the velocities at each point are in a single direction and at multiple speeds.

We next consider motion categories in which the set of velocities at each moving point is 2-D.

Category 0/2

Category 0/2 is similar to Category 0/1 in that it occurs instantaneously rather than continually. In Category 0/2, a 2-D set of velocities is defined at a single point in the image. An example is the image of a stone dropping into water and producing a splash. Another example is a bomb exploding. In the small spatio-temporal neighborhood of this singular event, image motions occur at all velocities.

Category 1/2

Category 1/2 is defined by a line that produces a set of 2D velocities. We have been unable to think of a natural example of such a motion and we mention it for completeness only. An artificial example would be an explosion of a wire. Near the instant of the explosion, all velocities are observed at each point on the wire.

Category 2/2

Category 2/2 motion occurs when all velocities are present at all image positions. An example of this motion would be a swarm of flying insects: within any image region there is a well-defined 2-D set of image velocities. Other examples of motions in this category include a speckle pattern such as white noise on a TV or specular reflections of sunlight off small waves of a swimming pool. Such motions are referred to as “scintillation” in [1]. As in Category 1/1b, there is no well-defined motion correspondence for scintillation, yet we can regard such image sequences as motion because they would stimulate typical energy motion detectors.

2.3 Discussion

We have categorized image motion patterns by a dimensional analysis of the image space (x, y) and velocity space (v_x, v_y) . Several novel motion categories emerged from the analysis, and we were able to relate the categories to each other. Certain categories occur instantaneously (see * in Table 2) whereas others occur continually. Certain categories are defined by a natural spatio-temporal correspondence between points, whereas others are defined by a noise-like

speckle in which no well-defined correspondence exists but the patterns nonetheless stimulate energy motion detectors.

While this categorization does provide a framework for the many motions we have discussed, it still has a number of limitations. Certain types of time-varying images are not considered, such as a flicker of lighting [1] or specularities such as a “flash in the pan.” Also, because our analysis is based on dimensional considerations only, we do not consider detailed properties of the parameters of the motion that might be visually important, such as sign of the direction of motion, specific values of any of the coordinates (extrema or zeroes), or topological sub-categories. Nor do we consider higher order properties of motion such as acceleration. Finally, we do not give special status to motions in which there are a pair of velocities at each point such as in the case of transparency or motion at an occlusion boundary [7]. Despite these limitations, the categories we have discussed provide a concise and coherent description of a wide range of image motions.

3 Optical snow

The second contribution of this paper is to analyze Category 2/1 motion in detail. This category, which we call “optical snow,” is very common in real scenes but has been neglected entirely in previous studies of image motion.

We will study optical snow by analyzing its frequency domain properties. As background, we begin by reviewing frequency domain properties of classical translational motion, a specific case of optical flow in which the velocity field is constant over an image window. We then extend this frequency domain property to optical snow and present a method for analyzing optical snow which is based on a bow tie signature in the frequency domain. To our knowledge, this bow tie signature has not been reported before in the vision literature.

3.1 Motion in the frequency domain

Analysis of motion in the spatiotemporal frequency domain has a long history in motion understanding research. Most of the analysis is based on the following property (see [8]): a textured image pattern $I(x, y)$ that translates with a uniform image velocity (v_x, v_y) produces a plane of energy in the frequency domain. This property is the basis of several optical flow techniques, for example [9, 10, 11], and is formally stated as follows.

Let \mathcal{F} be the Fourier transform operator, let $(\omega_x, \omega_y, \omega_t)$ be the spatiotemporal frequency variables, and let $\delta()$ be the Dirac delta function. Then,

$$\begin{aligned} & \mathcal{F}\{I(x - v_x t, y - v_y t)\} \\ = & \mathcal{F}\{I(x, y)\} \delta(v_x \omega_x + v_y \omega_y + \omega_t). \end{aligned}$$

It follows that the image energy in the frequency domain is restricted to lie on a plane,

$$v_x \omega_x + v_y \omega_y + \omega_t = 0. \quad (1)$$

Treating the image velocity components, v_x and v_y , as constants in the above equation, the motion plane of Eq. (1) intersects the plane $\omega_t = 0$ at a line:

$$v_x \omega_x + v_y \omega_y = 0. \quad (2)$$

The velocity vector (v_x, v_y) is perpendicular to this line.

From Eq. (1), if we consider ω_t as a function of ω_x and ω_y , then the gradient of ω_t is a constant vector, *i.e.*

$$\left(\frac{\partial \omega_t}{\partial \omega_x}, \frac{\partial \omega_t}{\partial \omega_y} \right) = (-v_x, -v_y). \quad (3)$$

Thus, the slope of the motion plane in the direction $(\omega_x, \omega_y) = (v_x, v_y)$ is $-\sqrt{v_x^2 + v_y^2}$.

3.2 Bow Tie Signature

How can the above frequency domain properties be applied to the case of optical snow? Assume that the camera is translating in a 3-D direction that is perpendicular to the optical axis. In this case, the image velocities of all the objects in the scene have a common direction, and the speeds of the objects vary inversely with depth [12]. Let the 3D camera motion have velocity $(v_x, v_y, 0)$ relative to the camera coordinate axis $\hat{x}, \hat{y}, \hat{z}$ where \hat{x} and \hat{y} are image coordinates and \hat{z} is the optical axis. In the scenario, the projected image motions of the objects in the scene are parallel to (v_x, v_y) . That is, the image velocity of a point is $(\alpha v_x, \alpha v_y)$ where α depends on the 3D depth of the point.

Fig. 1 shows a synthetic example. The scene is a set of spheres of constant 3D size placed at random positions within a view volume. An image sequence $I(x, y, t)$ was created by moving the camera upwards at a constant velocity.¹ The image sequence was rendered using OpenGL. One frame of the sequence, or xy slice, is shown in Fig. 1a. Following [6], we illustrate the motion using a yt slice in Fig. 1b. Multiple image speeds are evident in the form of space-time bars of multiple orientations [12].

The motion properties described above are manifest in the frequency domain as follows. Since each image velocity produces a plane in the frequency domain, multiple image velocities produce multiple planes. When all the image velocities are in the same direction $(\alpha v_x, \alpha v_y)$, all motion

¹The spheres are illuminated by collimated light source (point source at infinity) in direction parallel to the line of sight, so there are no shadows. The scene is viewed in perspective. The spheres are at depths ranging from $z = 4$ to -4 . The camera is at $z = 10$ and the field of view is 30 degrees. The image sequence was created by moving the camera upwards at a constant velocity $V_y = 0.025$.

planes pass through the line of Eq. (2). In this case, we say that the family of planes yields a *bow tie* distribution of power (see Fig. 2).

For a given image sequence, the bow tie can be visualized by projecting (summing) the 3D power spectrum $|\hat{I}(\omega_x, \omega_y, \omega_z)|^2$ orthographically onto the plane,

$$\cos\theta \omega_x + \sin\theta \omega_y = 0 \quad (4)$$

where $\theta \in [0, \pi)$. The bow tie signature emerges when the projection is in the direction of the line of Eq. (2), that is, when $(\cos\theta, \sin\theta)$ is parallel to (v_x, v_y) . For this θ , each motion plane projects to a line.

For the specific example of Fig. 1, the image velocities are all in the vertical direction *i.e.* (v_x, v_y) is parallel to $(0, 1)$. From Eq. (2), the common line through which the velocity planes pass satisfies $\omega_t = \omega_y = 0$ and therefore has direction $(1, 0, 0)$.

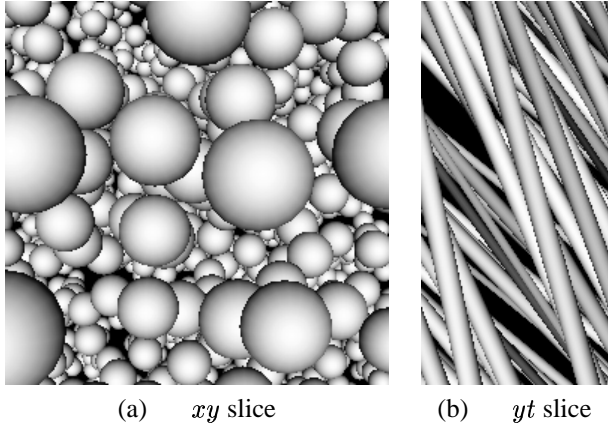


Figure 1: Falling spheres sequence. (a) First frame. (b) yt slice of sequence taken at rightmost pixel column of (a). As the camera moves upward, the imaged spheres move downward. Their speed is inversely proportional to their depth.

Fig. 3 shows projections of the power spectrum of the falling sphere sequence for different θ . When $\theta = \frac{\pi}{2}$, we obtain a bow tie signature as predicted since the image velocities are parallel to $(0, 1)$. The slope of the lines in the bow tie are positive since the spheres are traveling downward relative to the camera and hence the v_y are all negative (recall Eq. (3)). When θ deviates from $\frac{\pi}{2}$, the bow tie signature is diminished since the velocity planes no longer project to lines.

Since the bow tie signature occurs only when the projection of Eq. (4) is perpendicular to the motion planes, we say that the bow tie is a *non-accidental feature* for unidirectional motion in the image plane. Non-accidental features have been proposed for many types of motion including rigid collections of points [13], oscillatory motion [4], and 3D motion of point [14].

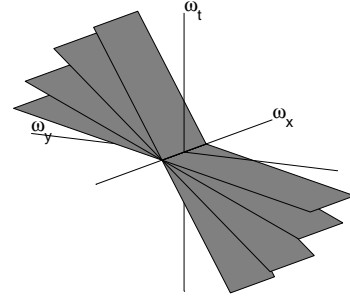


Figure 2: Each image velocity produces a plane in the frequency domain. Multiple image velocities, each having the same direction, produce multiple planes intersecting at a common line, Eq. (2).

In the next section, we introduce a method for recovering the direction angle θ .

3.3 Recovering direction of translation

We propose a two-stage approach to recovering the properties of optical snow. The first stage is to estimate the direction of image motion. The second stage is to estimate the range of image speeds in that direction. In this paper, we address only the first stage.

To estimate the direction of motion, we estimate the orientation θ of the bow tie signature in the frequency domain. For each θ , we sum all power in a *truncated wedge* above the origin. The truncated wedge is defined by frequencies $(\omega_x, \omega_y, \omega_t)$ such that

$$\left| \frac{\omega_t}{\cos\theta \omega_x + \sin\theta \omega_y} \right| > v_{max}$$

and

$$\|(\omega_x, \omega_y, \omega_t)\|_2 > \tau$$

where v_{max} is the maximum expected speed, and τ is a threshold that removes power near the origin. The latter constraint is needed because the direction of a small vector is poorly defined.

In our experiments we chose $\tau = 8$ and $v_{max} = 4.0$ pixels per frame. We varied θ between 0 and 180 degrees at 5 degree increments, *i.e.* the projected power is a function of θ , with a 180 degree periodicity. The function should have a value of zero when $(\cos\theta, \sin\theta)$ is parallel to the direction of image motion. This is the direction in which the bow tie signature emerges and thus there is no power above the origin in the projection plot. The projected power measurement increases to a maximum when $(\cos\theta, \sin\theta)$ is perpendicular to the direction of image motion. That is, the

minima and maxima of the function are separated by 90 degrees. In practice, the minimum is non-zero. This residual power in the truncated wedge is due to aliasing effects and the finite size of the objects (spheres). In addition, for real images, there is noise and camera jitter as well.

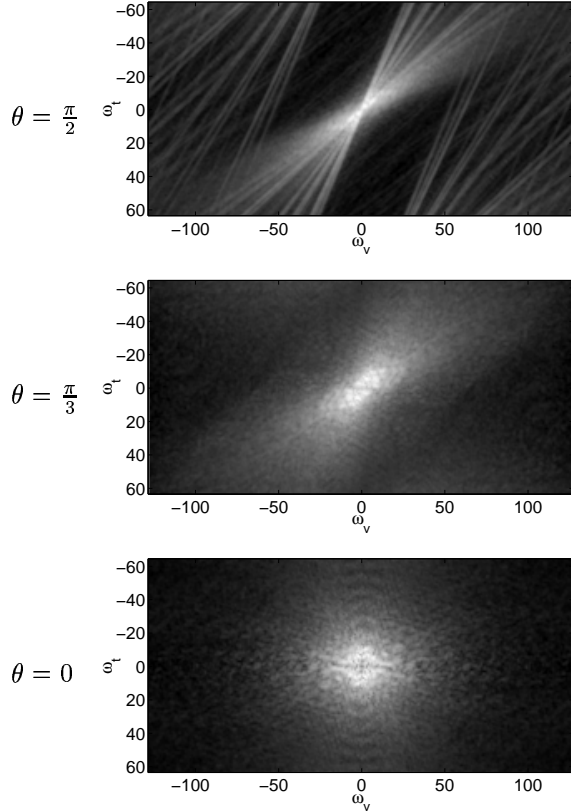


Figure 3: Projected power spectrum for synthetic sequence in Fig. 1 onto (ω_v, ω_t) where $\omega_v = \cos \theta \omega_x + \sin \theta \omega_y$ and $\omega_v \in [-\pi, \pi]$, $\omega_t \in [-\pi, \pi]$. In this and subsequent figures, the axes have range $[-\frac{N}{2} \dots \frac{N}{2} - 1]$, where N is the number of samples. To increase contrast, log power is plotted. White corresponds to high values. The bow tie signature is visible at $\theta = \frac{\pi}{2}$ where the motion planes project to lines.

We applied the truncated wedge, or “bow tie detector” to several synthetic sequences of translating spheres, such as in Fig. 1. Fig. 4 shows a typical set of responses of the detector, as a function of θ . In this case, we rotated the camera -20° around the optical axis prior to rendering the sequence. The minimum response occurs at $\theta = 70^\circ$ as expected. The response curve is normalized so the maximum power is 1. The detector has a minimum response of 0.067, a dynamic range of more than an order of magnitude.

Results for a natural image sequence are shown in Figure 5. The image sequence was obtained using a Hitachi MPEG MP-EG10W camera. The image sequence is of a

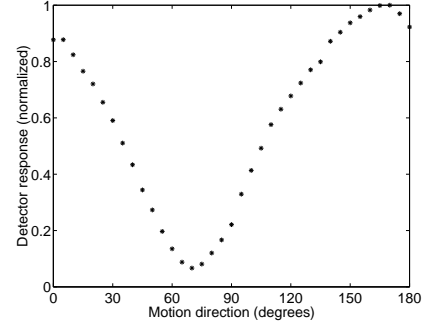


Figure 4: Bow tie detector applied to a synthetic image sequence similar to Fig. 1 but in which the camera was rotated -20° around the optical axis prior to rendering. The minimum power in the truncated wedge occurs at 70° .

holly bush, seen by a camera moving horizontally on a sliding platform. 128 frames were extracted from the middle of the sequence. A 240×256 region was taken from each frame (see outlined areas in Figure 5a). The region was padded with zeros to give a 256×256 image at every frame.

The mean grey level value of the entire sequence was subtracted from each image frame. The frames were windowed in both space and time by a Gaussian to reduce aliasing. The standard deviation σ of the Gaussian for each of the x, y and t axes was chosen so that the image $I(x, y, t)$ had width 6σ for each axis. For each sequence $I(x, y, t)$, the 3D Fourier transform was calculated and the projection of the power spectrum was computed for a range of θ as in Fig. 4.

Figure 5b shows the results of the bow tie detector for the holly sequence. There is a well-defined minimum at 0 degrees, which corresponds to the horizontal motion. Figure 5c shows the projected power spectrum at 0 degrees. The bow tie signature is clearly visible.

4 Summary

One of the fundamental goals of computational studies of image motion is to develop methods for automatically characterizing motion sequences based on both the camera motions (pan, tilt, zoom, translate, etc) and scene content (spatial layout, density and motion of objects). The contributions reported in this paper are to introduce a model of motion categories based on a dimensional analysis, and to analyze an important category of motion that has not been addressed before, namely optical snow, which is due to the translation of an observer through a cluttered 3D scene. We have also presented a computational theory of how to recover parameters of such a motion.

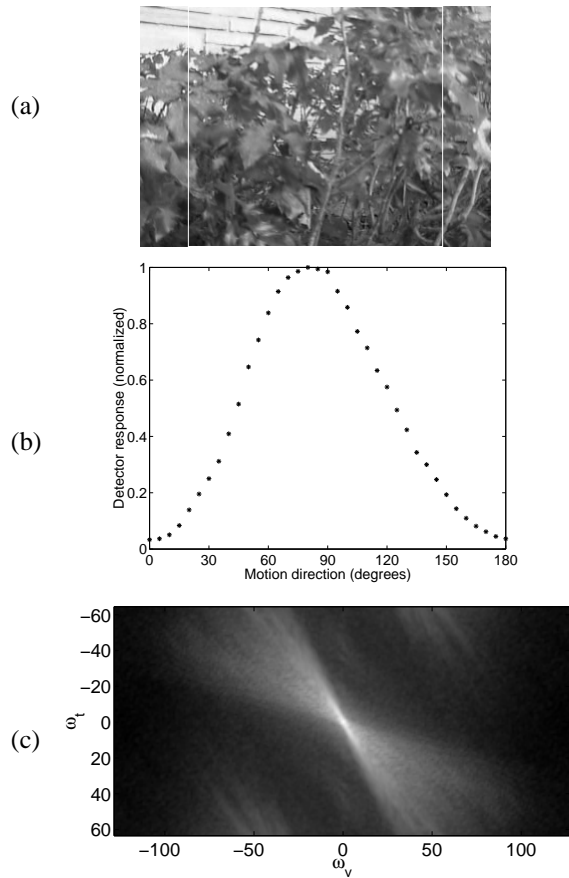


Figure 5: **(a.)** Frame from real image sequence of a holly bush. The white box outlines the 240×256 region used (the remaining rows were padded with zeros). **(b.)** Bow tie detector applied to the *holly* image sequence. The minimum occurs at 0° , corresponding to horizontal motion. **(c.)** The bow tie signature is visible at $\theta = 0$.

Besides opening up new avenues for computer vision, we believe this paper raises important questions about motion processing in biological vision. Cluttered 3D environments such as forests or tall grasslands are abundant in the natural world. Animals that inhabit such environments must solve complex motion understanding tasks in order to navigate, to avoid predators and to track their prey. (See the discussion of “the fox and the forest” in [2].) There have been many studies of human and primate vision which have addressed conditions of motion at multiple depths, or transparent motion. (See recent survey [15]). Stimuli for such studies are typically composed of a small number (usually two) of motion layers, each of which is composed of patterns such as random dots, lines, or transparent layers. While the impression of layers in these stimuli is typically strong, the impression of 3-D depth is relatively weak compared to optical snow

motions such as the sphere sequence in Fig. 1. It would be interesting to measure how well the human or primate visual system can detect the direction of motion in such stimuli and how well the performance can be explained using a computational model such as the bow tie detector that we have introduced.

References

- [1] R. Wildes and J. Bergen. Qualitative spatiotemporal analysis using an oriented energy representation. In *ECCV 2000*, pages 768–784, 2000.
- [2] S. W. Zucker and L. Iverson. From orientation selection to optical flow. *Computer Vision Graphics and Image Processing*, 37:196–220, 1987.
- [3] M.J. Langer and S.W. Zucker. What is a light source? In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 172–178, 1997.
- [4] J. Davis, A. Bobick, and W. Richards. Categorical representation and recognition of oscillatory motion patterns. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 628–635, Hilton Head, SC, 2000.
- [5] J. J. Koenderink. Optical flow. *Vision Research*, 26(1):161–179, 1986.
- [6] E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2):284–299, 1985.
- [7] K.M. Mutch and W.B. Thompson. Analysis of accretion and deletion at boundaries in dynamic scenes. *PAMI*, 7(2):133–138, March 1985.
- [8] A.B. Watson and A.J. Ahumada. Model of human visual-motion sensing. *Journal of the Optical Society of America A*, 2(2):322–342, 1985.
- [9] D.J. Heeger. Optical flow from spatiotemporal filters. In *First International Conference on Computer Vision*, pages 181–190, 1987.
- [10] D. J. Fleet. *Measurement of Image Velocity*. Kluwer Academic Press, Norwell, MA, 1992.
- [11] E. P. Simoncelli. *Distributed Analysis and Representation of Visual Motion*. PhD thesis, Massachusetts Institute of Technology, 1993. Department of Electrical Engineering and Computer Science.
- [12] R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1:7–55, 1987.
- [13] S. Ullman. *The Interpretation of Visual Motion*. 1979.
- [14] A. D. Jepson, W. Richards, and D. Knill. Modal structure and reliable inference. In D. Knill and W. Richards, editors, *Perception as Bayesian Inference*, pages 63–92. Cambridge University Press, 1996.
- [15] R. J. Snowden and F.A.J. Verstraten. Motion transparency: making motion models transparent. *Trends in Cognitive Sciences*, 3:369–377, 1999.