Optical snow and the aperture problem

Richard Mann
School of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1 CANADA

Michael S. Langer
School of Computer Science
McGill University
Montreal, Quebec H3A 2A7 CANADA

Abstract

Classical studies of measuring image motion by computer have concentrated on the case of optical flow, in which there is a unique velocity near each point of the image. In [5], we introduced a generalization of optical flow in which a range of parallel velocities can occur near each point in the image. Such image motion arises in many natural situations, such as camera motion in a cluttered 3-D scene or a stationary camera viewing falling snow. We refer to these image motions as optical snow. In the present paper, we show how the aperture problem manifests itself in this type of image motion.

1 Introduction

It is well-known that image motion is a strong cue for visual navigation and 3D scene reconstruction. Most computer vision methods for measuring image motion assume that there is a unique velocity vector at each point in the visual field (see review [1]). This type of motion is called optical flow. Recently we considered a more complex form of motion that we call *optical snow,* in which there is a 1D set of parallel velocities near each image point [5]. Such image motion arises when an observer moves relative to a densely cluttered 3-D scene. An example is a stationary observer viewing falling snow, or an observer moving laterally past a tree (Fig. 1).

Despite the seeming complexity of optical snow, informal observation tells us that this type of motion provides a rich set of visual cues about depth and 3-D spatial layout. Humans can navigate well in cluttered environments as can other animals such birds and squirrels, for example, moving through the branches of a tree. These observations motivate the novel computer vision problem of how to recover the properties of optical snow. In [5], we showed how to recover the direction of optical snow (see Sec. 2 below). In the present paper, we analyze the case that the objects in the scene have oriented spatial structure. This results in a mo-
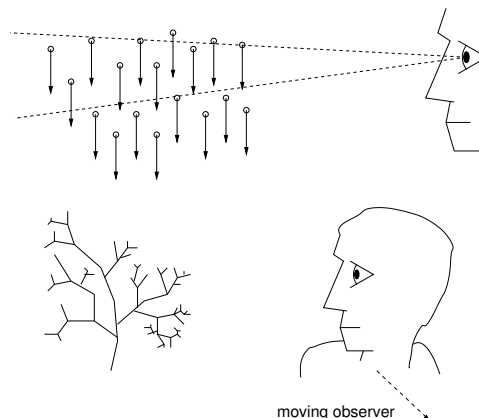

Figure 1. Optical snow arises when a camera/observer moves relative to cluttered 3D scene.

tion ambiguity much like the aperture problem in classical optical flow. We show how an analogous problem emerges for optical snow.

2 Background

Let $I(x, y, t)$ be a time varying image. The image flow constraint equation [4] says that if a local image patch is translating with velocity (v_x, v_y), then this local velocity is constrained by:

$$v_x \frac{\partial I}{\partial x} + v_y \frac{\partial I}{\partial y} + \frac{\partial I}{\partial t} = 0 \qquad (1)$$

Eq. (1) can be expressed in the frequency domain as follows. The 3D Fourier transform of $I(x, y, t)$ is defined:

$$\hat{I}(f_x, f_y, f_t) = \int \int \int I(x, y, t) \, e^{-2\pi i (f_x x + f_x y + f_t t)} \, dx \, dy \, dt$$

Using the derivative property of Fourier transforms,

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

we can take the 3D Fourier transform of Eq. (1). This yields:

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

Eq. (2) implies that all non-zero frequency components of $\hat{I}(f_x, f_y, f_t)$ lie on the plane:

$$v_x \ f_x + v_y \ f_y + f_t \ = \ 0 \ . \quad (3)$$

The observation that a translating image yields a plane in the 3D frequency domain was first made in [8]. Many computer vision algorithms have since been based on this observation. These algorithms recover a velocity (v_x, v_y) in each local patch of the image by finding the plane that best fits the 3D power spectrum of that local patch [3, 2, 7].

In [5], we showed how to extend this translation model of image motion to the case in which there is not one, but rather a large set of parallel velocities within an image patch. Consider an image patch in which the objects all move in the same direction but at different speeds. For example, if the observer is moving then the image speed of each object that is visible in the patch varies with the depth of that object. Since the velocity vectors are assumed to be parallel, each velocity is of the form $(\alpha v_x, \alpha v_y)$. From Eq. (3), a set of parallel image velocities produce multiple planes in the frequency domain,

$$\alpha v_x \ f_x \ + \ \alpha v_y \ f_y + f_t \ = \ 0 \ . \quad (4)$$

Each of these motion planes intersects the (f_x, f_y) plane, i.e. $f_t = 0$, along the line

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

Because the motion planes all intersect at this line, we say the planes have a *bow tie* distribution of power in the 3D frequency domain (see Fig. 2a).

Let the bowtie angle θ_b be the angle from the f_x axis to the line of Eq. (5). This angle is measured from the f_x axis towards the f_y axis. The line of Eq. (5) is in direction $(\cos \theta_b, \sin \theta_b)$ in the (f_x, f_y) plane. Since by definition $(\alpha v_x, \alpha v_y)$ is perpendicular to the line of Eq. (5), the direction of (v_x, v_y) is $(-\sin \theta_b, \cos \theta_b)$.

For example, consider an image sequence of vertically falling snow (recall Fig. 1). Since the image velocities would be of the form $(0, \alpha)$, it follows that $\theta_b = 0$, and so all the motion planes pass through the f_x axis.

In [5], we introduced a technique for recovering the direction θ_b for the motions present in an image patch. We sum the power in a *wedge* of frequencies in (f_x, f_y, f_t) (see Fig. 2b), specifically, the frequencies above and below the origin. [1] The wedge is then rotated through various angles

---

[1] The wedge is defined by a slope v_{max} which is chosen by the user. In [5], we used $v_{max} = 4$ pixels/frame which is a very fast motion.
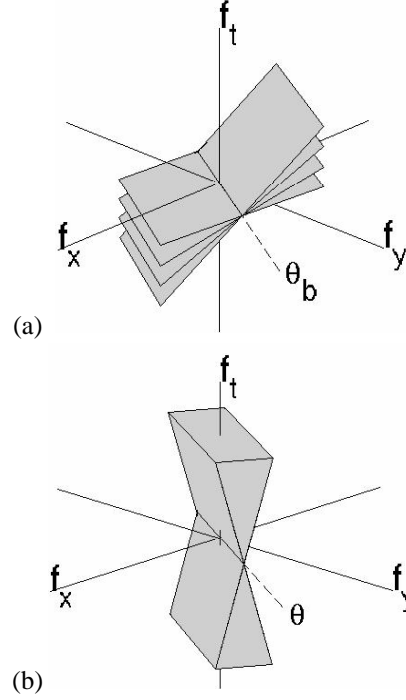

(a)

(b)

Figure 2. (a) Bow tie distribution of power in frequency domain. (b) A wedge is used to estimate the orientation of the bow tie .

θ at fixed angular increments. For each θ, the total power in the wedge is measured.

Let $W(\theta)$ denote the power in the wedge as function of θ. $W(\theta)$ has a 180 degree periodicity.

KEY OBSERVATION 1: *The function $W(\theta)$ has a minimum when $\theta = \theta_b$, that is, when the wedge is in the same direction as the bow tie*, as in Fig. 2.

The only power in the wedge at $\theta = \theta_b$ is due to image velocities that are greater than v_{max}. These results were presented in [5].

3 Aperture problem

The contribution in the present paper is to show how the classical "aperture problem" manifests itself in optical snow. In many scenes, objects will be dominated by particular orientations. For example, in a crowd scene or in a forest, the dominant orientation is vertical. In the context of optical flow, it is well-understood that if an image has oriented structure then only the component of velocity that is normal (perpendicular) to this oriented structure can be measured. This *aperture problem* can be expressed using Eq. 1. If the direction of the image gradient $(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})$ is constant over a local space-time image patch, then only

the velocity component that is parallel to the image gradient can be recovered within that patch [6]. What is the corresponding phenomenon in optical snow?

3.1   Example 1: parallel cylinders

We begin by considering the case of a 3D scene consisting of infinitely long parallel cylinders like trees in a forest. These parallel cylinders are assumed to be perpendicular to the optical axis of the camera so that the cylinders are parallel in the image as well in 3D. Let θ_n be such that the image direction $(\cos\theta_n, \sin\theta_n)$ is *normal* to the axis of the cylinders. The spatial gradient of intensities in the image must thus be parallel to $(\cos\theta_n, \sin\theta_n)$. Since all power in the 2D frequency domain (f_x, f_y) is parallel to the spatial gradient direction, all power must lie on frequencies (f_x, f_y) parallel to $(\cos\theta_n, \sin\theta_n)$, that is,

$$(f_x, f_y) \cdot (-\sin\theta_n, \cos\theta_n) \; = \; 0. \qquad (6)$$

For example, if the trees are vertical in the image then $\theta_n = 0^o$ and $f_y = 0$.

If the camera were to translate through such a cluttered scene, then optical snow would result. For any small image patch[2], image power would be contained in a family of planes, namely a bowtie as in Fig. 2. But because the spatial frequencies containing the image power must obey Eq. (6), which in the 3D frequency domain is a plane, it follows that the bow tie distribution of power must be restricted to a set of radial lines within the plane of Eq. (6). (See Fig. 3.) These lines define a set of *normal velocities* in the image. From the discussion up to Eq. (6), these normal velocities are in direction $(\cos\theta_n, \sin\theta_n)$.

We emphasize that there is no relationship between the angle θ_n and the angle θ_b. The angle θ_n is the direction of the gradient of image intensities and the angle θ_b is the direction of the image velocities of the objects.

Let us next address how θ_n can be measured. Because the power is concentrated in the plane of Eq. (6), the function $W(\theta)$ has a maximum when this plane is contained in the wedge.

KEY OBSERVATION 2: *The function $W(\theta)$ has a maximum when $\theta = \theta_n$, that is, when the wedge is in the same direction as the gradient of the image intensities.*

Note that this special θ depends entirely on the orientation of the cylinders, rather than on the image velocities $(\alpha v_x, \alpha v_y)$. In particular, the maximum of $W(\theta)$ is in the direction normal to the cylinder axis.

The above observation is not so surprising for the case of the cylinders, since the normal velocity has to be perpendicular to the orientation of the cylinders. The analysis is more

---

[2]We exclude a patch containing the focus of expansion, since the assumption of parallel velocities would fail.
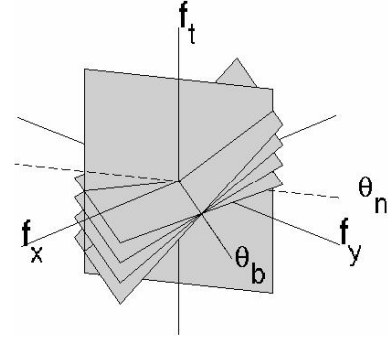

Figure 3. θ_n **is the direction of the gradient of image intensities.**

interesting when we generalize it to cases in which a range of spatial orientations is present in the image.

3.2   Example 2: parallel ellipsoids

A more general synthetic case to consider is a cluttered scene consisting of 3D elongated ellipsoids. For simplicity, suppose that the major axis of each 3D ellipsoid is parallel to the image plane and that the camera motion is vertical. The major axis of the 3D ellipsoids plays the same role as axis of the cylinders in Example 1, namely the angle θ_n is defined to be perpendicular (normal) to this major axis.

Fig. 4 shows the functions $W(\theta)$ for five scenes, which differ in the y/x aspect ratio of the ellipsoids. Fig. 4(a) shows $W(\theta)$ that is produced when the major axis of the ellipsoids is aligned with the camera's x axis. Since the major axis of each ellipsoid is perpendicular to θ_n, the minimum of $W(\theta)$ occurs at $\theta = 0^o$ and the maxima at $\theta = \pm 90^o$.

Fig. 4(b) shows $W(\theta)$ for the same scenes, except that the ellipsoids are now rotated so that the major axis is -30 degrees from the x axis. Since θ_n is perpendicular to the major axis, we have $\theta_n = -30^o + 90^o = 60^o$. When the y/x aspect ratio is 1, we obtain essentially the same curve as in Fig. 4(a). As the aspect ratio shrinks, the position of the minimum of $W(\theta)$ stays the same because of the camera motion hasn't changed, but the position of the maxima of $W(\theta)$ shifts. If the aspect ratio were to shrink to zero, the maxima of $W(\theta)$ would shift all the way to $\theta = \theta_n = 60^o$.

3.3   Example 3: a real forest

Let us next consider a real scene containing oriented structure. The sequence was shot through the window of a horizontally moving car using a Hitachi MPEG MPEG10W camera. The optical axis was perpendicular to the direction of motion of the car. The scene was a densely treed area. The trees were bare of leaves, and the visible branches of the trees were at many different depths. A single frame from the sequence is shown in Fig 5a.
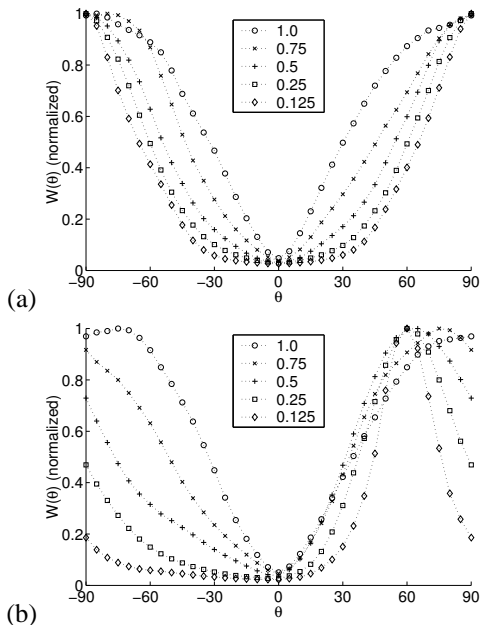
Figure 4. Plots of $W(\theta)$ **for two ellipsoid sequences.**

The camera produces images of size 352×240 at 30 frames/sec. Sequences of 128 images were extracted and converted to grey scale images. To demonstrate the optical snow effects, we considered a 128×128 upper-middle area of each frame (see outlined region in Fig. 5a.). The 3D Fourier transform was calculated and the function $W(\theta)$ was computed. The minima in $W(\theta)$ occurs at $\pm 90°$ which is as expected since the motion is horizontal ($\theta_b = \pm 90^o$). (Recall Key Observation 1.)

The width of the peak at half-height is approximately 50 degrees and so the width of the valley is 130 degrees. This broad minimum is due to the predominance of vertical structure (trees) over non-vertical structure (branches) in the scene. If we compare the plot in Fig. 5 with those in Fig. 4, we see that a half-height peak-width of 50 degrees corresponds to an "effective" aspect ratio of about 0.25. We see that the function $W(\theta)$ provides us with both the direction of normal velocity in the image, as well as an estimate of the range of spatial orientations contributing to the space-time image structure.

4   Conclusions

Classical frequency-based optical flow methods [3, 2, 7] assume that there is a unique velocity in each image patch and attempt to find this velocity by fitting a plane to the power spectrum of the image patch. Here we considered a new type of motion called optical snow in which a range of parallel velocities are present in a patch. We have shown previously [5] how to estimate the direction of these parallel motions by extending the classical frequency-based model.
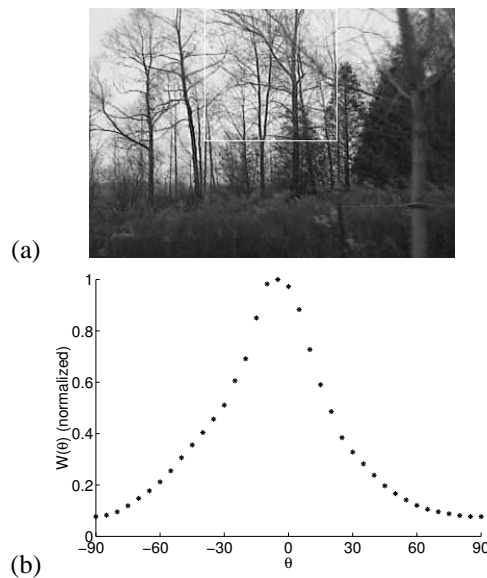

Figure 5. (a) Frame from a image sequence of a forested area. (b) $W(\theta)$**.**

In the present paper, we extended our analysis to cases in which the spatial structure of the image is dominated by particular orientations. We show how the aperture problem manifests itself in these cases and how the direction of the normal velocities can be understood and measured.

References

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

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

[3] D. Heeger. Optical flow from spatiotemporal filters. In *First International Conference on Computer Vision*, pages 181–190, 1987.

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

[5] M. S. Langer and R. Mann. Dimensional analysis of image motion. In *IEEE International Conference on Computer Vision*, pages 155–162, 2001.

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

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

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