Tracking through optical snow

M. S. Langer[1] and R. Mann[2]

[1] School of Computer Science, McGill University, Montreal, H3A2A7, Canada
email:langer@cs.mcgill.ca
http://www.cim.mcgill.ca/~langer
[2] School of Computer Science, University of Waterloo, Waterloo, Canada
email:mannr@uwaterloo.ca
http://www.cs.uwaterloo.ca/~mannr

Abstract. Optical snow is a natural type of image motion that results when the observer moves laterally relative to a cluttered 3D scene. An example is an observer moving past a bush or through a forest, or a stationary observer viewing falling snow. Optical snow motion is unlike standard motion models in computer vision, such as optical flow or layered motion since such models are based on spatial continuity assumptions. For optical snow, spatial continuity cannot be assumed because the motion is characterized by dense depth discontinuities. In previous work, we considered the special case of parallel optical snow. Here we generalize that model to allow for non-parallel optical snow. The new model describes a situation in which a laterally moving observer tracks an isolated moving object in an otherwise static 3D cluttered scene. We argue that despite the complexity of the motion, sufficient constraints remain that allow such an observer to navigate through the scene while tracking a moving object.

1 Introduction

Many computer vision methods have been developed for analyzing image motion. These methods have addressed a diverse set of natural motion categories including smooth optical flow, discontinuous optical flow across an occlusion boundary, and motion transparency. Recently we introduced a new natural motion category that is related to optical flow, occlusion and transparency but that had not been identified previously. We called the motion *optical snow.* Optical snow arises when an observer moves relative to a densely cluttered 3-D scene (see Fig. 1).

Optical snow produces dense motion parallax. A canonical example of optical snow is falling snow seen by a static observer. Although snowflakes fall vertically, the image speed of each snowflake depends inversely on its distance from the camera. Since any image region is likely to contain snowflakes at a range of depths, a range of speeds will be present. A similar example is the motion seen by an observer moving past a cluttered 3D object such as a bush. Any image region will contain leaves and branches at multiple depths. But because of parallax, multiple speeds will be present in the region.

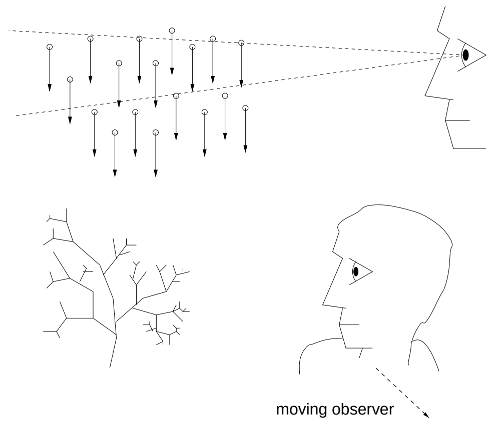

Fig. 1. Optical snow arises when a camera/observer moves relative to cluttered 3D scene.

Optical snow is a very common motion in nature which makes it especially relevant for computer vision models that are motivated by biological vision. Animals that are typically studied by visual neuroscientists include the rabbit, cat and monkey. These animals inhabit environments that are densely cluttered, for example, grasslands or forest. Since most of our knowledge of motion processing in the visual brain of mammals is obtained from experiments done on these animals, it is important to understand the computational problems of motion perception in the environments these animals inhabit, namely cluttered 3D environments.

In earlier papers on optical snow[1, 2], we developed a mathematical model of the motion that extended the classical frequency domain analysis of Watson and Ahumada [3]. In the present paper, we generalize that model to the case of non-parallel optical snow, and we explicitly relate the model to classical equations of motion of a moving observer of Longuet-Higgins and Prazdny [4]. Implications for the problem of tracking are discussed.

2 Background

Previous research on image motion that uses the spatiotemporal frequency domain is based on the following motion plane property [3]: 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 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 will itself travel with that velocity. Each of these translating 2D sine waves will produce a unique spatiotemporal frequency component in the translating image sequence. The velocity vector (v_x, v_y) induces a specific relationship (see Eq. 2 below) between the temporal and spatial frequency of each translating component.

Formally, let $I(x, y, t)$ be a time varying image that is formed by pure translation, so that

$$I(x,\ y,\ t) \ = \ I(x + v_x dt,\ y + v_y dt,\ t + dt)\ . \tag{1}$$

Taking the Fourier transform, $\hat{I}(f_x, f_y, f_t)$, one can show [3, 2] that

$$\hat{I}(f_x, f_y, f_t)\ (v_x\ f_x + v_y f_y + f_t) \ = \ 0\ .$$

Thus, any non-zero frequency component of the translating image satisfies

$$v_x\ f_x + v_y f_y + f_t \ = \ 0\ . \tag{2}$$

This is the motion plane property.

Several methods for measuring image motion 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 patch of the image by finding the motion plane that best fits the 3D power spectrum of that local patch [5–8]. The motion plane property has also been used by several methods that recover layered transparency. 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 [9, 10].

3 Optical snow

The motion plane property was originally designed for pure translation, that is, for a unique image velocity (v_x, v_y). We observe that the property can be extended to motions in which there is a one-parameter family of velocities within an image region. Suppose that the velocity vectors in an image region are all of the form $(u_x + \alpha\ \tau_x, u_y + \alpha\ \tau_y)$ where $\{u_x, u_y, \tau_x, \tau_y\}$ are constants and the parameter α varies between points in the region. We do not make any spatial continuity assumptions about α since we are modelling densely cluttered 3D scenes.

From Eq. (2), this one parameter family of image velocities produces a one-parameter family of planes in the frequency domain,

$$(u_x + \alpha\ \tau_x)\ f_x \ + \ (u_y + \alpha\ \tau_y)\ f_y + f_t \ = \ 0 \tag{3}$$

where α is the free parameter. This claim must be qualified somewhat because of occlusion effects which are non-linear, but we have found that as long as most of the image points are visible over a sufficiently long duration, the multiple motion plane property above is a good approximation. See also [11, 12] for discussion of how occlusions can affect a motion plane.

From the family of motion planes above, we observe the following:

Claim: *The one parameter family of motion planes in Eq. (3) intersect at a common line that passes through the origin in the frequency domain* (f_x, f_y, f_t). (see Fig. 2)

Proof: Each of motion planes in Eq. (3) defines a vector $(u_x + \alpha\, \tau_x, u_y + \alpha\, \tau_y,\ 1)$ that is normal to its motion plane. These normal vectors all lie on a line in the plane $f_t = 1$. Let us call this line l. The line l, together with the origin, span a plane π in the frequency domain. The vector perpendicular to π is, by definition, perpendicular to each of the normal vectors in l. Hence, the line from the origin in the direction of this perpendicular vector must lie in each of the motion planes. This proves the claim.

We say that the family of planes has a *bowtie pattern* and we say the line of intersection of the planes is the *axis of the bowtie*. 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 that the axis of the bowtie is in direction $(-\tau_y,\ \tau_x,\ u_x\tau_y - \tau_x u_y)$.

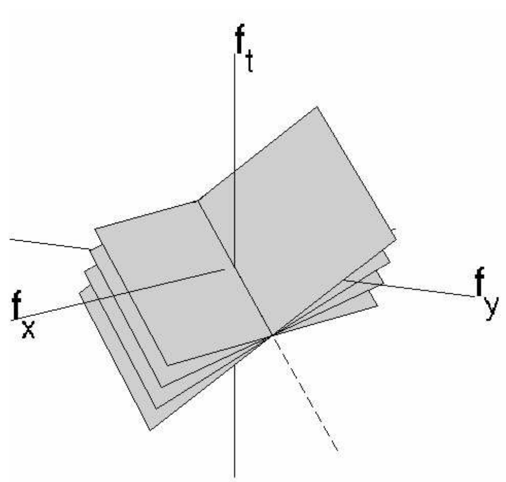

Fig. 2. Optical snow produces a bowtie pattern in frequency domain.

In our previous papers [1, 2], we considered the case that the axis of the bowtie lies in the (f_x, f_y) plane, that is, (u_x, u_y) is parallel to (τ_x, τ_y). In this special case, there is a unique motion direction. We showed how a vision system could recover the parameters of optical snow, namely how to estimate the unique motion direction and the range of speeds α in the motion. In the present paper, we investigate the more general case that (u_x, u_y) is not parallel to (τ_x, τ_y). As we will see shortly, both the special case and the general case just mentioned have a natural interpretation in terms of tracking an object in the scene.

4 Lateral motion

A canonical example of optical snow occurs when an observer moves laterally through a cluttered static scene. One can derive an expression for the resulting instantaneous image motion using the general equations of the motion field for an observer moving through a static scene which are presented in [4, 13]. Let the observer's instantaneous translation vector be (T_x, T_y, T_z) and let the rotation vector be $(\omega_x, \omega_y, \omega_z)$. Lateral motion occurs when the following approximation holds:

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

In this case, the focus of expansion is well away from the optical axis. Similar to [14], we also restrict the camera motion by assuming

$$\omega_z \approx 0 \ ,$$

that is, the camera may pan and tilt but may not roll (no cyclotorsion). These two constraints reduce the basic equations of the motion field to

$$(v_x, v_y) = (-\omega_y, \omega_x) + \frac{1}{Z}(T_x, T_y) \qquad (4)$$

where Z is the depth of the point visible at a given pixel. (We have assumed without loss of generality that the projection plane is at $z = 1$.) The model of Eq. (4) ignores terms that are second order in image coordinates x, y. These second order terms are relatively small for pixels that are say ± 20 degrees from the optical axis. Note that Eq. (4) is a particular case of optical snow defined in the previous section, with constants $(u_x, u_y) = (-\omega_y, \omega_x)$, $(\tau_x, \tau_y) = (T_x, T_y)$, and $1/Z$ being the free parameter α.

5 Tracking an object

One common reason for camera rotation during observer motion is for the observer to track a surface patch in the scene, that is, to stabilize it in the image in order to better analyze its spatial properties. We assume first that the entire scene including the tracked surface patch is static and only the observer is moving. Tracking a surface patch at depth Z' stabilizes the projection of the patch on the retina by reducing its image velocity to zero. For this to happen, the camera rotation component of the image motion must exactly cancel the image translation component of that surface patch, that is, from Eq. (4),

$$(v_x, v_y) = (0, 0) \quad \text{if and only if} \quad (-\omega_y, \omega_x) = -\frac{1}{Z'}(T_x, T_y)$$

When the observer tracks a particular surface patch, scene points at other depths will still undergo image motion. From the previous equation and Eq. (4), the image velocity of a point at depth Z in the scene will be:

$$(v_x, v_y) = (\frac{1}{Z} - \frac{1}{Z'}) \ (T_x, T_y).$$

Two observations follow immediately. First, if the observer is tracking a particular point in the scene then Eq. (4) implies that all velocity vectors in the image will be in direction (T_x, T_y) and hence parallel.[3] We call this case of *parallel optical snow.* Notice that parallel optical snow also arises when $(-\omega_y, \omega_x) = (0, 0)$ which is the case of no camera rotation. In this case, the observer is tracking at point at infinity.

The second observation is that a point in front of the tracked point $(Z < Z')$ will have image motion in the opposite direction of a point behind the tracked point $(Z > Z')$. For example, consider walking past a tree while tracking a squirrel that is sitting in the tree. Leaves and branches that are nearer than the squirrel will move in the opposite direction in the image as those that are farther than the squirrel.

The above discussion assumed the observer was tracking a static surface patch in a *static* 3D scene. One natural way to relax this assumption is to allow the tracked surface patch (object) to move in the scene, with the scene remaining static otherwise, and to suppose the observer tracks this moving object while the observer also moves. A natural example is a predator tracking its moving prey [15]. In this scenario, the camera rotation needed to track the object may be in a different direction than the camera's translation component. For example, the predator may be moving along the ground plane and tracking its prey which is climbing a tree. In this case the translation component of motion (T_x, T_y) is horizontal and the rotation component of motion $(-\omega_y, \omega_x)$ is vertical.

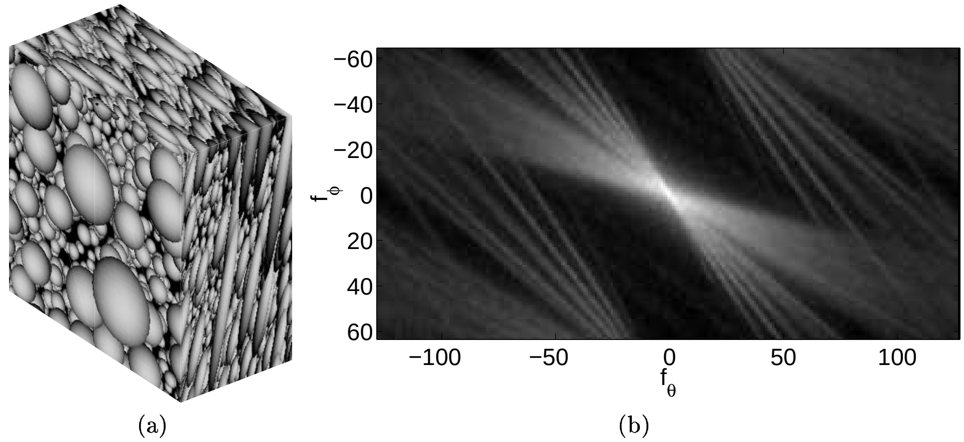

Fig. 3. (a) xyt cube of sphere sequence. (b) Projection of power spectrum in the direction of the axis of the bowtie. Aliasing effects (wraparound at boundaries of frequency domain) are due to the "jaggies" of OpenGL. Such effects are less severe in real image sequences because of optical blur [1, 2].

[3] Recalling the approximation of Section 4, this result breaks down for wide field of views, since second order effects of the motion field become significant.

A contrived example to illustrate non-parallel optical snow is shown in Figure 3. A synthetic image sequence was created using OpenGL. The scene was a set of spheres distributed randomly in a view volume. The camera translates in the y direction so that (T_x, T_y, T_z) is in direction $(0, 1, 0)$. As the camera translates, it rotates about the y axis so that $(\omega_x, \omega_y, \omega_z)$ is in direction $(0, 1, 0)$ and the rotation component of the image motion is in direction $(1, 0, 0)$. In terms of our tracking scenario, such a camera motion would track a point object moving on a helix

$$(X(t), Y(t), Z(t)) = (r \sin t, t, r \cos t)$$

where r is the radius of the helix and t is time.

For our sequence, the camera rotates 30 degrees (the width of the view volume) in 128 frames. Since the image size is 256×256 pixels, this yields $(u_x, u_y) = (2, 0)$ pixels/frame. Fig. 3(a) shows the xyt video cube [16] and Fig. 3(b) shows a summed projection of the 3D power spectrum onto a plane. The projection is in the bowtie axis direction $(1, 0, 1)$. The bowtie is clearly visible.

6 Future Work

One problem that is ripe for future work is to develop algorithms for estimating the bowtie that are more biologically plausible than the one we have considered which uses an explicit representation of the Fourier transform. To be consistent with motion processing in visual cortex, a biologically plausible algorithms would be based on the measurements of local space-time motion energy detectors [16]. Such algorithms could generalize current biological models for motion processing such as [5, 7, 8] which assume a pure translation motion. These current algorithms estimate the velocity of the pure translation motion in a space-time image patch by combining the responses of motion energy detectors, and estimating a single motion plane in the frequency domain.

Our idea for analyzing optical snow in this manner to estimate a bowtie pattern rather than a single motion plane. Since a motion plane is just a bowtie whose range of speeds collapses to a single speed, the problem of estimating a bowtie generalizes the previous problem of estimating a motion plane. The details remain to be worked out. However, biologically plausible models of layered motion transparency have been proposed already [17]. We expect similar algorithms could be designed for optical snow as well.

Acknowledgments

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

References

1. M. S. Langer and R. Mann. Dimensional analysis of image motion. In *IEEE International Conference on Computer Vision*, pages 155–162, 2001.
2. R. Mann and M. S. Langer. Optical snow and the aperture problem. In *International Conference on Pattern Recognition*, Quebec City, Canada, Aug. 2002.
3. 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.
4. H.C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B)*, B-208:385–397, 1980.
5. D.J. Heeger. Optical flow from spatiotemporal filters. In *First International Conference on Computer Vision*, pages 181–190, 1987.
6. D. J. Fleet. *Measurement of Image Velocity.* Kluwer Academic Press, Norwell, MA, 1992.
7. N.M. Grzywacz and A.L. Yuille. 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, 1990.
8. E P Simoncelli and D J Heeger. A model of neural responses in visual area mt. *Vision Research*, 38(5):743–761, 1998.
9. M. Shizawa and K. Mase. A unified computational theory for motion transparency and motion boundaries based on eigenenergy analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 289–295, 1991.
10. P. Milanfar. Projection-based, frequency-domain estimation of superimposed translational motions. *Journal of the Optical Society of America A*, 13(11):2151–2162, November 1996.
11. D. J. Fleet and K. Langley. Computational analysis of non-fourier motion. *Vision Research*, 34(22):3057–3079, 1994.
12. S.S. Beauchemin and J.L. Barron. The frequency structure of 1d occluding image signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2):200–206, February 2000.
13. E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision.* Prentice-Hall, 1998.
14. M. Lappe and J. P. Rauschecker. A neural network for the processing of optical flow from egomotion in man and higher mammals. *Neural Computation*, 5:374–391, 1993.
15. S. W. Zucker and L. Iverson. From orientation selection to optical flow. *Computer Vision Graphics and Image Processing*, 37:196–220, 1987.
16. 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.
17. R. S. Zemel and P. Dayan. Distributional population codes and multiple motion models. In D. A. Cohn M. S. Kearns, S. A. Solla, editor, *Advances in Neural Information Processing Systems 11*, pages 768–784, Cambridge, MA, 1999. MIT Press.