CAR-TR-894 N00014-95-1-0521
CS-TR-3925 September 1998

Qualitative Description of Camera Motion and Scene Depth from Histograms of Normal Flow[1]

Peter Cucka[2], Zoran Duric[3], Ehud Rivlin[4], and Azriel Rosenfeld

Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

Abstract

If we histogram the normal flow vectors in images of a scene viewed by a moving observer, we can use the time-varying histogram to derive qualitative information about the observer's motion—for example, whether it is (primarily) translational or rotational, and whether the direction of translation or axis of rotation is (roughly) parallel or perpendicular to the camera axis. This is illustrated using flow histograms obtained from a variety of real image sequences. If the motion is translational, qualitative information about the scene depth can also be obtained from the flow histograms—for example, whether the scene depth is unimodal or bimodal. This is illustrated for real scenes containing a layer of vegetation seen against a textured background, or two layers of vegetation.

Keywords: Camera motion, flow histogram, normal flow, optical flow, scene depth

[1]The support of the Office of Naval Research under Contract N00014-95-1-0521 is gratefully acknowledged, as is the help of Janice Perrone in preparing this paper. The authors also thank Brad Stuart for providing the panning sequence.

[2]Present address: DreamWorks SKG, Universal City, CA, 91608

[3]Also with the Department of Computer Science, George Mason University, Fairfax, VA 22030

[4]Also with the Department of Computer Science, Technion, Haifa, Israel 32000

1 Introduction

A moving agent that obtains a sequence of images of a stationary scene can infer both its motion and the layout of the scene, up to a range/speed ambiguity, by analyzing the images. The classical "structure from motion" problem (Ullman, 1979) attempts to determine the motion and layout completely (except for the ambiguity). However, as is well known, it is quite difficult to obtain this complete information accurately. (Thomas et al., 1993; Daniilidis and Spetsakis, 1997). It may therefore be useful to attempt to derive only partial information.

A number of researchers have shown how to derive partial information about the agent's motion or the scene structure from images. Examples include relative distances of objects (Nitzberg and Mumford, 1990); time to collision with (or rate of approach toward) an object (Lee, 1976; Nelson and Aloimonos, 1989; Subbarao, 1990; Cipolla and Blake, 1992; Meyer and Bouthemy, 1992; Ancona and Poggio, 1993; Tistarelli and Sandini, 1993; Duric et al., 1994; Burlina and Chellappa, 1994); and qualitative information about an object's shape (Koenderink and Van Doorn, 1975; Cipolla and Zisserman, 1992). Other relevant references include (Jain, 1983; Thompson and Kearney, 1986; Nelson and Aloimonos, 1988; Francois and Bouthemy, 1990; Weinshall, 1991; Thompson and Painter, 1992; Aloimonos and Duric, 1994; Fermuller and Aloimonos, 1995).

We show in this paper that a moving agent can obtain basic information about its motion by examining the sequence of two-dimensional histograms of normal flow vectors computed from successive pairs of images. (A related idea is the use of Hough transform methods in flow analysis (Kalvainen, et al. 1992; Bober and Kittler, 1994; Heikkonen, 1995).) Specifically, we show, using a variety of real image sequences, that these histogram sequences are quite different for different types of simple motions: translation along the camera axis, translation in a plane perpendicular to the camera axis, rotation around the camera axis, and rotation around an axis perpendicular to the camera axis. Obviously, combinations of these motions give rise to more complicated histograms; but it is well known that humans find such combined motions difficult to interpret. In any case, methods of selective stabilization (Duric and Rosenfeld, 1996; Yao and Chellappa, 1997) can be used to eliminate "unwanted" components—in particular, rotational components—from an image sequence.

If the motion is a translation, it is also possible to derive partial information about the layout of the scene by examining the normal flow histogram. In particular, if the motion is translation in a direction perpendicular to the camera axis (e.g., a moving vehicle carrying a sideways-pointing camera), parts of the scene at different distances from the camera give rise to different peaks in the normal flow histogram (so that if the distances are bimodal, the histogram is bimodal). On the other hand, if the motion is translation along the camera axis (a moving vehicle carrying a forward-pointing camera), the normal flow vectors all point away from the focus of expansion (FOE), which is located at the center of the image. In this situation, if we compute the magnitudes of the normal flow vectors, and scale them by their distances from the FOE, then parts of the scene at different distances from the camera give rise to different peaks in the one-dimensional histogram of the scaled values. We demonstrate these effects using images of real scenes containing a layer of vegetation seen against a textured background, or two layers of

vegetation. It should be pointed out that such scenes present a severe challenge to traditional structure-from-motion methods, which typically impose constraints on the (piecewise) smoothness of the surfaces in the scene and the optical flow fields to which they give rise.

The detection of peaks in velocity histograms was used by Jasinschi et al. (1992) in studies of "motion transparency" in synthetic patterns; an earlier technical report (Jasinschi et al., 1991) also contained the vegetation results presented here, but these results were omitted from (Jasinschi et al., 1992).

In Section 2 we give equations for the normal flow resulting from a given camera motion, and describe our method of computing normal flow histograms. In Section 3 we illustrate the appearance of normal flow histogram sequences obtained from different types of camera motions, and show how selective stabilization methods can be used to eliminate the effects of complex motions. In Section 4 we give examples of bimodal flow histograms obtained from translatory motions relative to scenes that have bimodal depth distributions.

2 Flow computation and histogramming

The instantaneous velocity of the image point (x, y), the image of the scene point (X, Y, Z) under perspective projection, is given by

$$\dot{x} = \frac{-Uf + xW}{Z} + \omega_x \frac{xy}{f} - \omega_y \left(\frac{x^2}{f} + f \right) + \omega_z y \tag{1}$$

$$\dot{y} = \frac{-Vf + yW}{Z} + \omega_x \left(\frac{y^2}{f} + f \right) - \omega_y \frac{xy}{f} - \omega_z x \tag{2}$$

where $\vec{T} = (U \;\; V \;\; W)^T$ is the translational velocity and $\vec{\omega} = (\omega_x \;\; \omega_y \;\; \omega_z)^T$ is the rotational velocity of the camera.

Let $\vec{\imath}$ and $\vec{\jmath}$ be the unit vectors in the x and y directions, respectively; $\dot{\vec{r}} = \dot{x}\vec{\imath} + \dot{y}\vec{\jmath}$ is the image motion field at the point $\vec{r} = x\vec{\imath} + y\vec{\jmath}$. If we choose a unit direction vector $\vec{n}_r = n_x\vec{\imath} + n_y\vec{\jmath}$ at the image point $\vec{r}$ and call it the normal direction, then the *normal motion field* at $\vec{r}$ is $\dot{\vec{r}}_n = (\dot{\vec{r}} \cdot \vec{n}_r)\vec{n}_r = (n_x\dot{x} + n_y\dot{y})\vec{n}_r$. $\vec{n}_r$ can be chosen in various ways; the usual choice (and the one that we use) is the direction of the image intensity gradient $\vec{n}_r = \nabla I / \|\nabla I\|$.

Note that the normal motion field along an edge is orthogonal to the edge direction. Thus, if at time t we observe an edge element at position $\vec{r}$, the apparent position of that edge element at time $t + \Delta t$ will be $\vec{r} + \Delta t \dot{\vec{r}}_n$. This is a consequence of the well-known *aperture problem*. We base our method of estimating the normal motion field on this observation.

For each image frame (say collected at time t) we find edges using an implementation of the Canny edge detector. For each edge element, say at $\vec{r}$, we resample the image locally to obtain a small window with its rows parallel to the image gradient direction $\vec{n}_r = \nabla I / \|\nabla I\|$. For the next image frame (collected at time $t + \Delta t$) we create a larger window, typically twice as large as the maximum expected value of the magnitude of the normal motion field. We then slide the first (smaller) window along the second

(larger) window and compute the difference between the image intensities. The zero of the resulting function is at distance u_n from the origin of the second window; note that the image gradient in the second window at the positions close to u_n must be positive. Our estimate of the normal motion field is then $-u_n$, and we call it the *normal flow*.

We construct a normal flow histogram by quantizing the x and y components of each u_n. The histogram shows the number of occurrences of each pair of quantized values.

3 Flow histograms for simple motions

In this section we show examples of flow histogram sequences that result from four simple types of camera motion: z-axis rotation, z-axis translation, lateral translation, and pan. (We follow the usual convention that the z-axis is the optical axis of the camera; lateral translation is translation along an axis in the xy-plane, and pan is rotation around such an axis.) We also show flow histograms obtained from a forward-pointing camera carried by a ground vehicle moving on unpaved terrain; when the image sequence is not stabilized, the flow shows a mixture of effects due to translation, roll, and pitch, but stabilization can be used to remove the rotation effects.

Figures 1a-b show two frames from the "Robot" sequence, taken at the University of Massachusets; here the motion is essentially clockwise z-axis rotation. This motion results in counterclockwise image rotation around the center of the image plane; at each point of the image, the direction of motion is tangential, and its magnitude increases with distance from the center. If the scene contained edges oriented in all directions, the resulting normal flow would contain vectors of all magnitudes in all directions; but since most of the edges in the scene in Figure 1 are oriented in two perpendicular directions, most of the normal flow vectors are perpendicular to those directions (Figures 1c-d). This results in a +-shaped flow histogram (Figures 1e-f); as the camera rotates, this histogram rotates (since the edges are revolving around the image center).

Figures 2a-c show three frames of the "Coke can" sequence, taken at NASA Ames Research Center; here the motion is essentially z-axis translation toward the scene. This results in an expansion of the image; at each point of the image, the direction of motion is radial, and its magnitude increases with distance from the center (and with closeness of the scene point to the camera). If the scene contained edges oriented in all directions, the resulting normal flow would contain vectors of all magnitudes in all directions; but since most of the edges in the scene in Figure 2 are oriented horizontally or vertically, most of the normal flow vectors are horizontal or vertical (Figures 2d-f), once again resulting in a +-shaped flow histogram (Figures 2g-i). As the camera approaches the scene, the flow magnitude increases, so that the flow histogram expands.

Figures 3a-b show two frames from the SRI "Tree" sequence, in which the camera is pointing forward and translating to the left. This results in rightward horizontal image flow whose magnitude increases with closeness of the scene point to the camera. Since the scene contains many near-vertical edges (the tree trunks), the normal flow is predominantly rightward (Figures 3c-d and e-f).

Figures 4a-b show two frames from a panning sequence taken in our laboratory. Since the camera motion was leftward (counterclockwise around a vertical axis, as seen from above), the image flow is predominatly rightward (Figures 4c-d). The scene contains

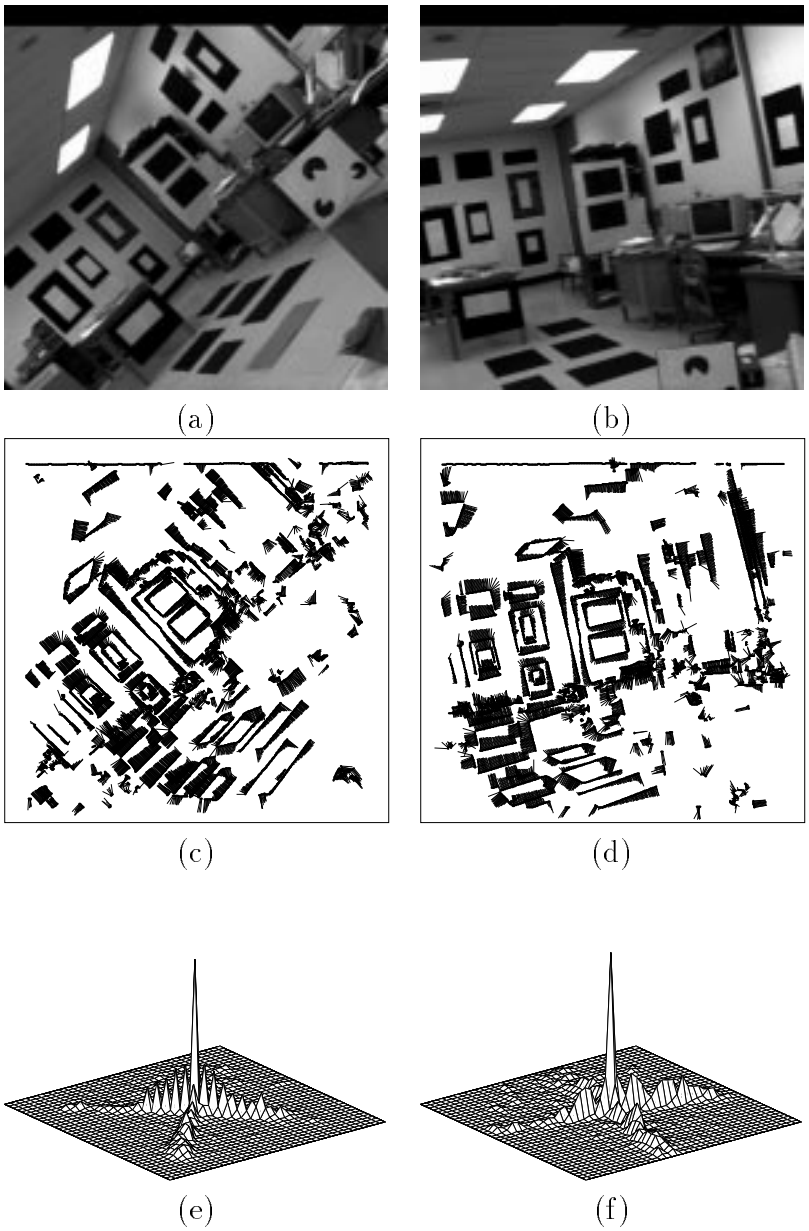


Figure 1: If the motion is z-axis rotatijon, the flow histogram rotates. (a-b) Frames 1 and 10 of the "Robot" sequence. (c-d) Flow fields for these frames. (e-f) Histograms of these flow fields.

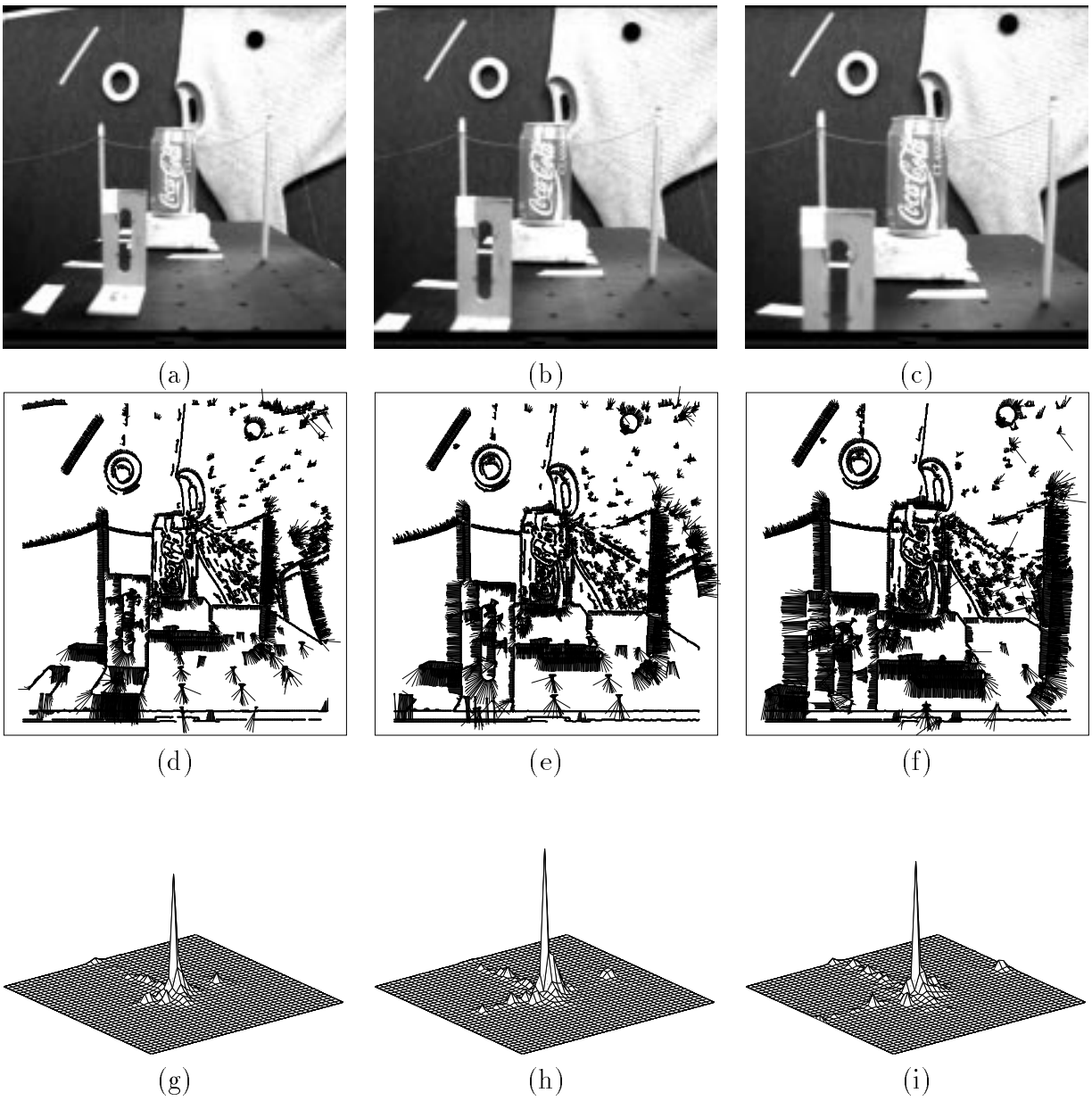


Figure 2: If the motion is z-axis translation, the flow histogram expands. (a-c) Frames 0, 75, and 150 of the NASA Ames "Coke can" sequence. (d-f) Flow fields for these frames. (g-i) Histograms of these flow fields.

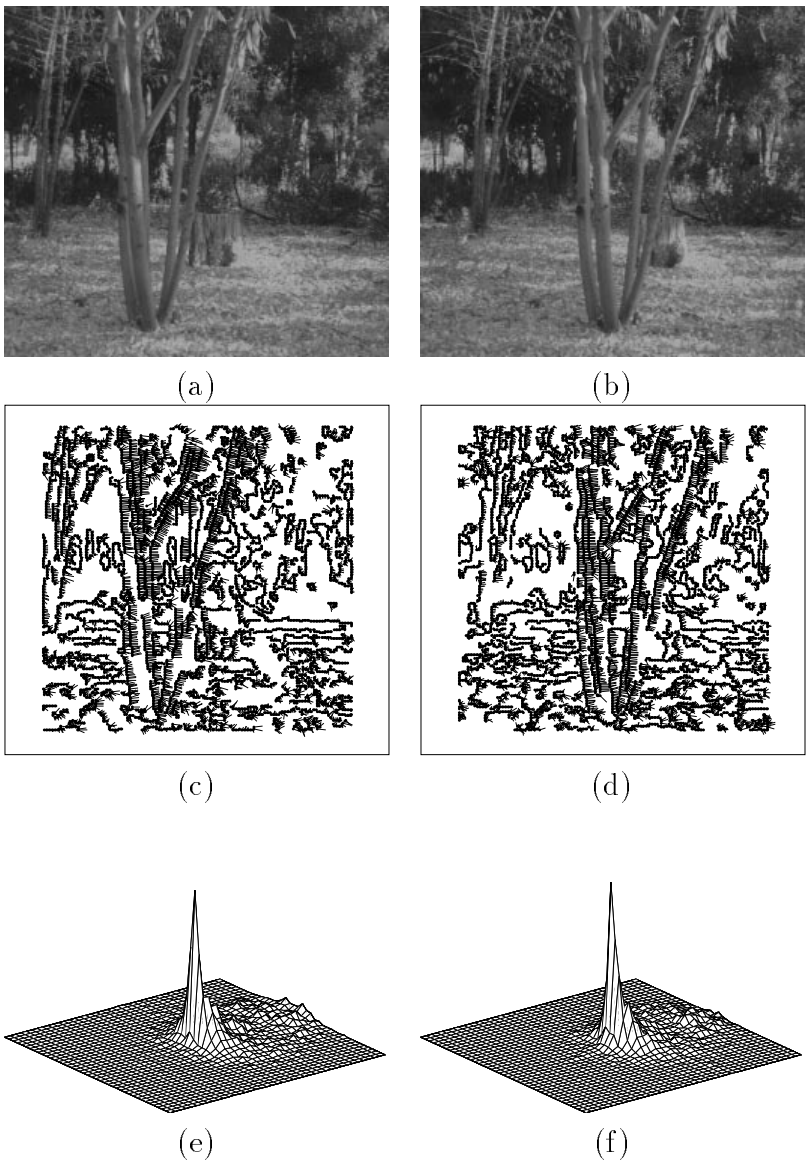


Figure 3: If the motion is lateral translation, the flow histogram is biased in the direction of motion. (a-b) Frames 0 and 18 of the SRI "Tree" sequence. (c-d) Flow fields for these frames. (e-f) Histograms of these flow fields.

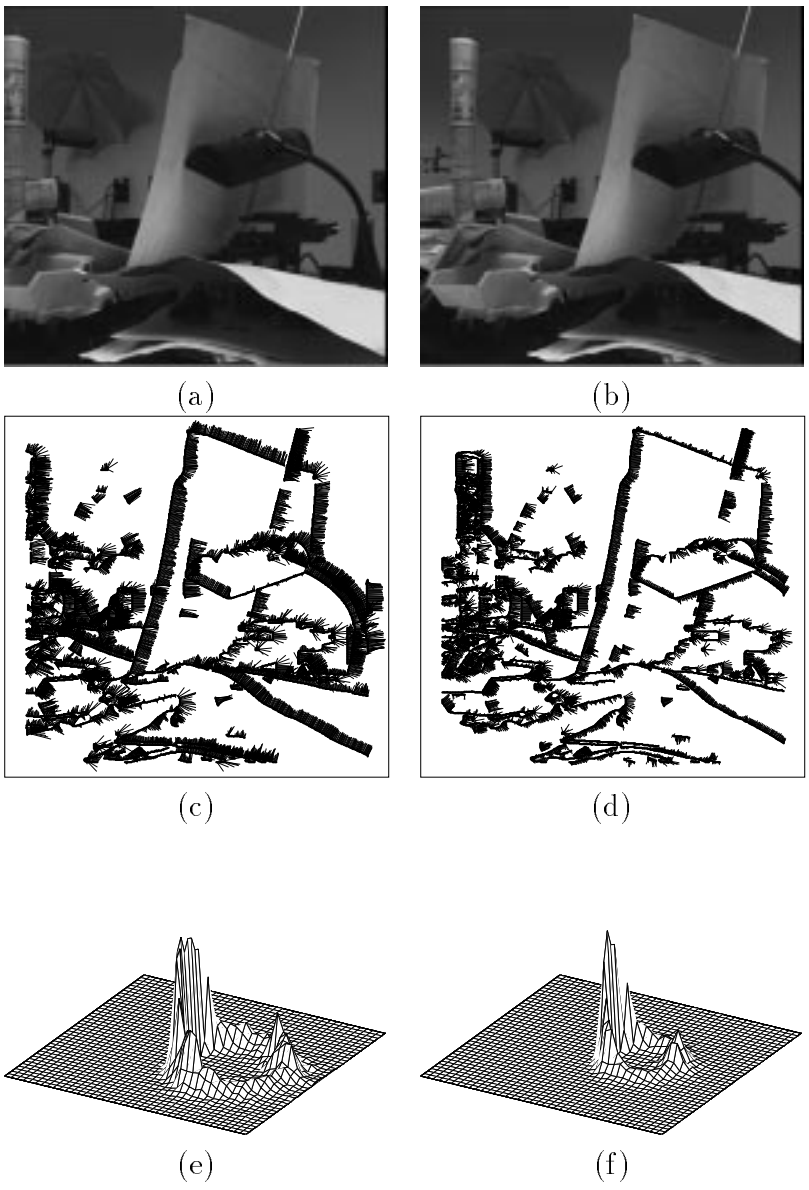


Figure 4: If the motion is panning, the image flow is approximately constant, so the normal flow at an edge of slope θ has magnitude proportional to cos θ; thus the flow histogram has a "crater". (a-b) Frames 8 and 68 of a panning sequence obtained in our laboratory. (c-d) Flow fields for these frames. (e-f) Histograms of these flow fields.

edges in many directions; for an image edge that makes an angle θ with the vertical, the magnitude of the normal flow is proportional to cos θ, and its direction is perpendicular to θ, so that the normal flow histogram displays a circular "crater" through the origin (Figures 4e-f).

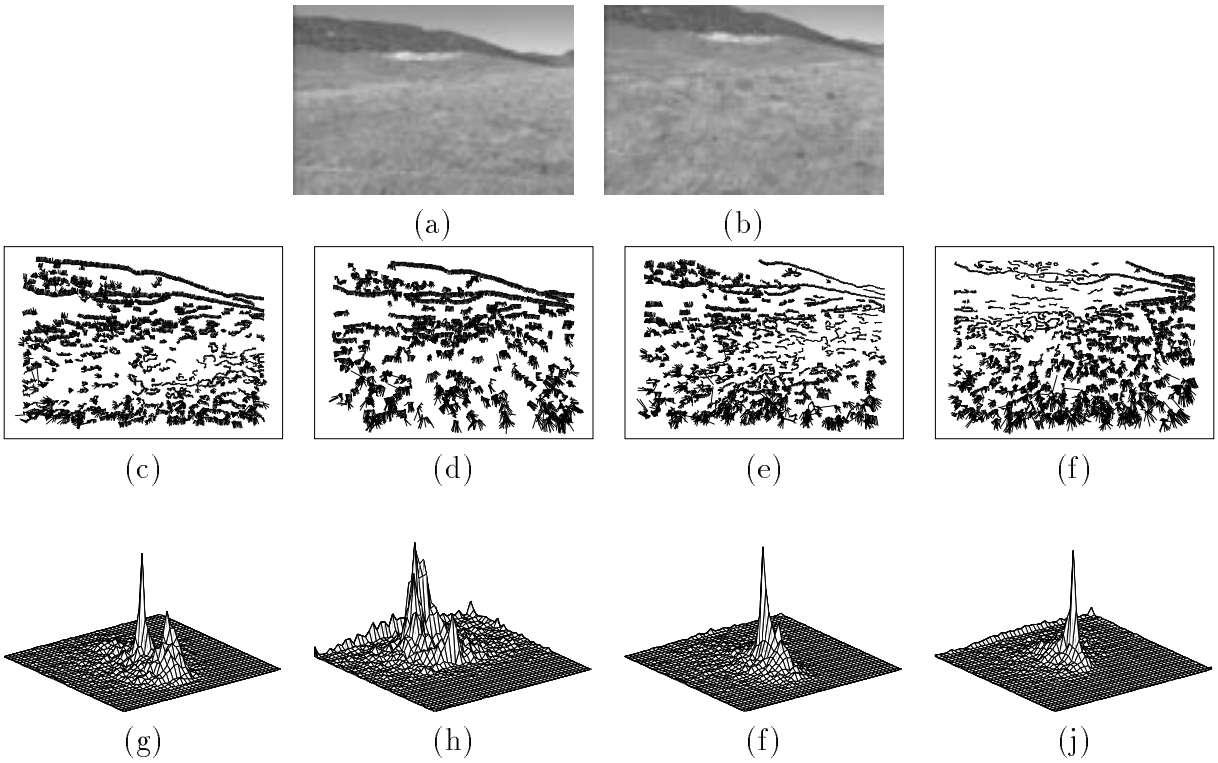


Figure 5: Mixed time-varying motions result in more complicated histograms, but stabilization can simplify them. (a-b) Two frames from an unstabilized image sequence obtained by a ground vehicle on rough terrain. (c,e,g,i) Flow fields for four frames intermediate between (a) and (b). The flow due to forward translation appears primarily in the lower (nearby) parts; the upper (distant) parts show the effects of (c) downward pitch, (e) upward pitch, (g) counterclockwise roll combined with downward pitch, and (i) counterclockwise roll only. (d,f,h,j) Histograms of these flow fields.

Figures 5a-b show two frames from an image sequence taken by a forward-pointing camera mounted on a ground vehicle moving across rough terrain. Figures 5c,e,g,i show the normal flow at four intermediate frames of this sequence. Here the backward image flow due to the forward translation appears primarily in the lower part of the image, which shows nearby parts of the terrain. In the upper part of the image, showing terrain near the horizon, the translation produces negligible flow, but the rotational effects of the bumpy motion are quite apparent. In Figure 5c the vehicle is pitching downward, resulting in upward image motion along the horizon. In Figure 5e the pitching is upward, resulting in downward image motion of the horizon. In Figure 5g the vehicle is pitching downward and rolling counterclockwise; the roll causes the left side of the horizon image

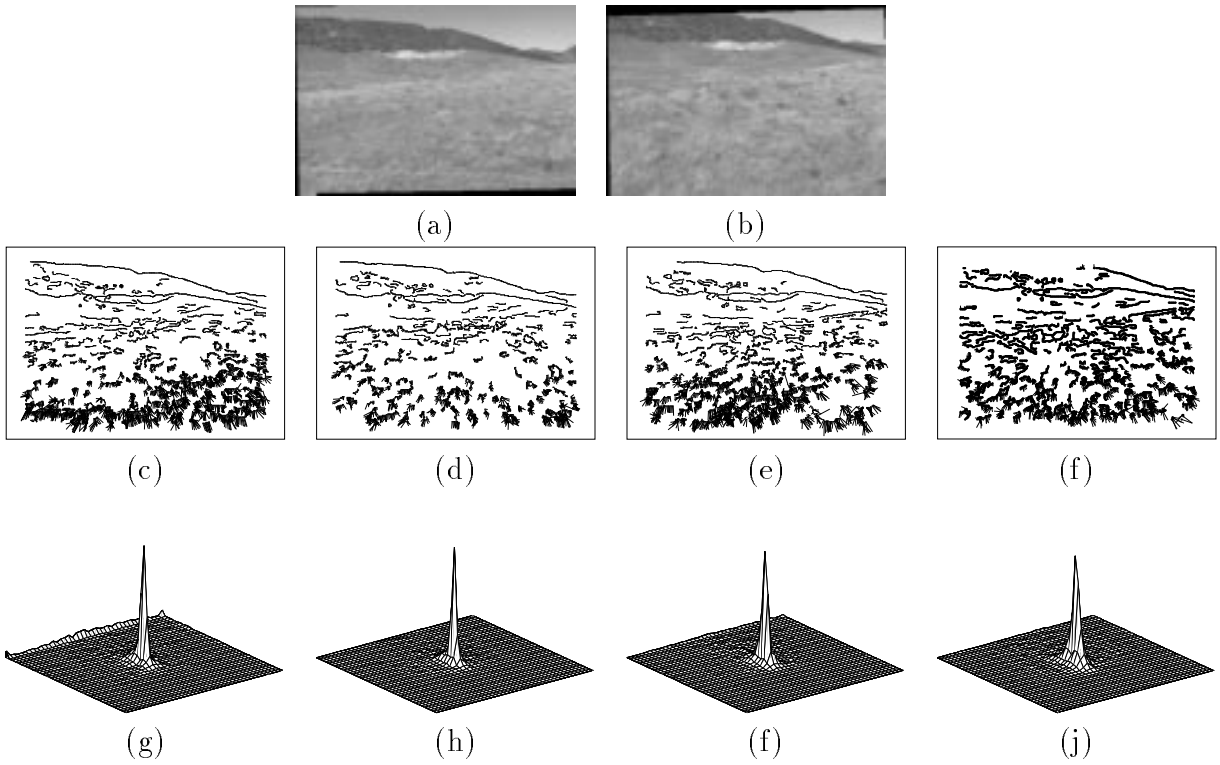


Figure 6: (a-j) Corresponding images, flow fields, and histograms after stabilization; the pitch and roll effects have been eliminated.

to move upward and the right side to more downward, while the pitch causes the entire horizon image to move upward; at the left of the image these effects add, but at the right they cancel. Finally, in Figure 5i the vehicle is rolling counterclockwise, but not pitching significantly; this results in upward motion of the horizon image on the left, and downward motion on the right. All of these effects are apparent in the normal flow and the histograms (Figures 5d,f,h,j). The image sequence can be stabilized to eliminate the rotational effects, by detecting the horizon and warping the images so that the horizon remains stationary. Figures 6a,b show the same frames as in Figures 5a,b after stabilization; note the loss of portions of the image near the edges due to the stabilization. Figures 6c-j show the normal flow and the histograms corresponding to Figures 5c-j, but for the stabilized image sequence; the rotational flow has been eliminated, leaving only the backward image motion due to the forward translation.

4 Flow histograms for translational motions and bimodal scene depth

When the camera motion is translational, qualitative information about scene depths can sometimes be obtained by examining normal flow histograms. The results in this section were obtained seven years ago (Jasinschi et al., 1991), but are reproduced here because they have not yet been widely published; they represent an early use of flow histograms

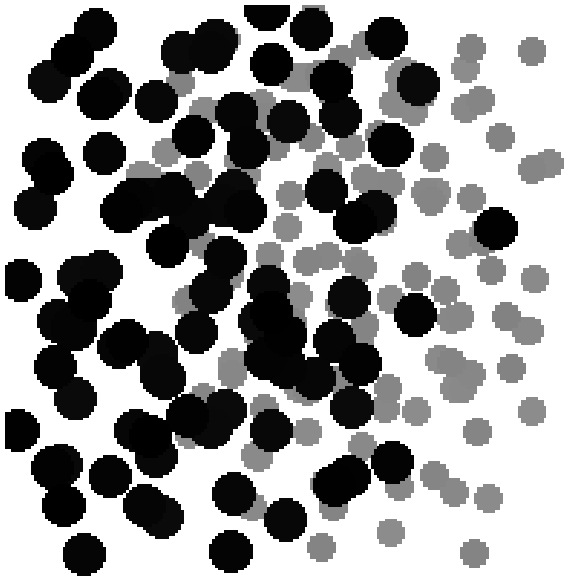
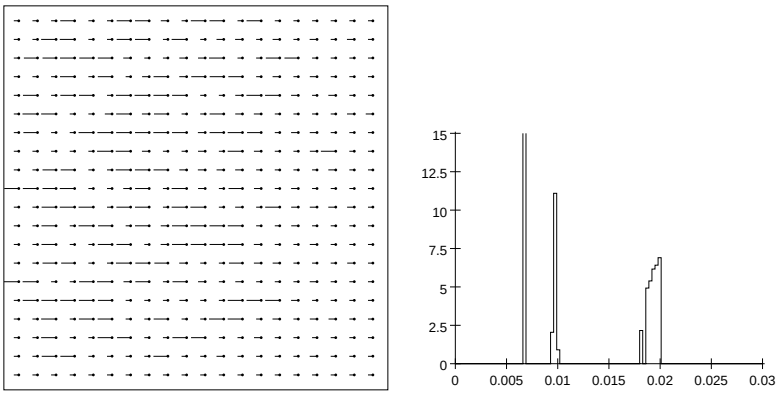


Figure 7: Depth map of an artificial scene containing two 3D textures in front of a planar backdrop. Darker patches correspond to surfaces that are closer to the observer.

for qualitative analysis of image sequences.

We first consider three examples using an artificial scene (Figure 7) containing two 3D textures at two different distances. In the first example (Figure 8) the camera is translating laterally, i.e. parallel to the scene; this yields a parallel flow field composed of vectors of three lengths (one representing each of the textures; the third representing the background). In the second example, the translation is toward the scene, which yields an expansive flow field, with the focus of expansion (FOE) located at the center of the image. The flow magnitude is proportional to both the closeness of the scene point to the camera and the distance of the image point from the FOE. To eliminate the latter effect, we divide the magnitude of each flow vector by the distance from the image point to the FOE; the result is shown in Figure 9. In the third example (see Figure 10), the motion is diagonal, but here again, dividing by distance to the FOE yields a sharply peaked flow magnitude histogram.

We next consider three real examples in which the translation is lateral. As pointed out in Section 3, leftward translation of a forward-pointing camera results in rightward image flow; the magnitude of this flow increases with closeness of the scene point to the camera. If the scene consists of layers of feature points at different distances from the camera, each layer will give rise to flow vectors having a different range of magnitudes; thus if there are two layers, the flow histogram will be bimodal. This phenomenon is illustrated in Figures 11-13. The scene in Figure 11a contains a plant in front of a patterned background, while the scenes in Figure 12a and 13a contain two plants in front of one another. In each case, the flow histogram (Figures 11b, 12b, 13b) is strongly bimodal; the peak closer to the origin arises from the more distant layer. [These histograms were constructed by a different method, described in (Jasinschi et al., 1991), but we would expect similar results using normal flow histograms.] Figures 11c-d, 11c-



(a) (b)

Figure 8: (a) The optical flow field generated by an observer translating parallel to the scene in Figure 7. (b) The velocity magnitude histogram of the flow field.

d, 11c-d show the image points that contributed flow vectors to the two peaks; they evidently correspond to scene points belonging to the two layers.

Finally, we consider a real example involving z-axis translation toward the scene (Figure 14a). The histogram of the scaled flow magnitudes (Figure 14b) is, as expected, bimodal; the peaks correspond to the two depth levels, as we see from Figures 14c-d.

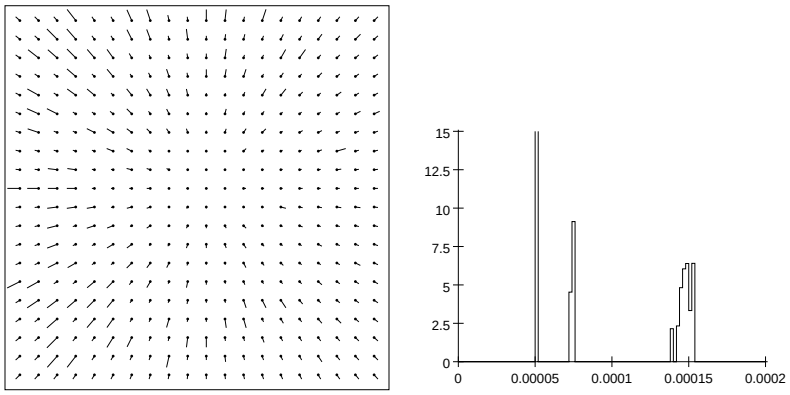
5 Concluding remarks

Normal flow histograms do not seem to have been used extensively in image sequence analysis. We have seen in this paper that they can provide useful qualitative information about the camera motion. It was shown in (Jasinschi et al., 1991) that when the motion is translational, they can sometimes also provide useful qualitative information about the scene depths. Evidently they have many other uses; for example, when a stationary camera views a scene containing many moving objects, a normal flow histogram can provide useful information about the (apparent) velocities of the objects. The authors hope that this paper will serve to call attention to the value of flow histograms, and will stimulate further study of their uses.

References

[1] Y. Aloimonos and Z. Duric, Estimating the heading direction using normal flow, *Intl. J. Computer Vision* **13**, 1994, 33–56.

[2] N. Ancona and T. Poggio, Optical flow from 1D correlation: Application to a simple time-to-crash detector, in Proc. ARPA Image Understanding Workshop, 1993, 673–682.



(a) (b)

Figure 9: (a) The optical flow field generated by an observer translating directly toward the scene in Figure 7. (b) Scaled (see text) velocity magnitude histograms of the flow field (see text).

[3] M. Bober and J. Kittler, Estimation of complex multimodal motion: An approach based on robust statistics and Hough transform, *Image Vision Computing* **12**, 1994, 661–668.

[4] P. Burlina and R. Chellappa, Time-to-x: Analysis of motion through temporal parameters, in Proc. Conf. on Computer Vision and Pattern Recognition, 1994, 461–468.

[5] R. Cipolla and A. Blake, Surface orientation and time to contact from image divergence and deformation, in Proc. European Conf. on Computer Vision, 1992, 187–202.

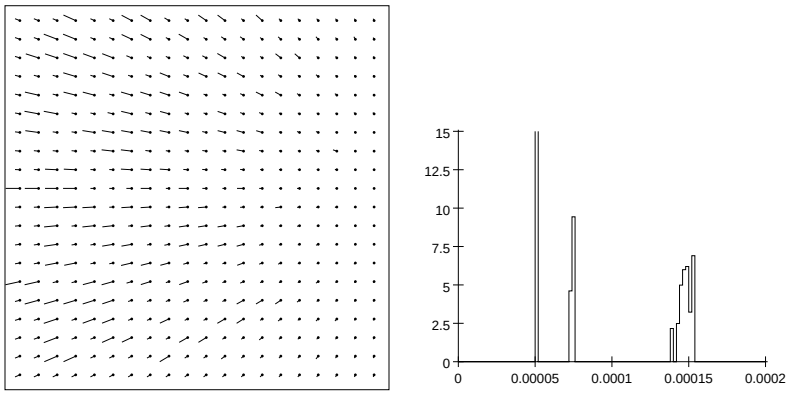
[6] R. Cipolla and A. Zisserman, Qualitative surface shape from deformation of image curves, *Intl. J. Computer Vision* **8**, 1992, 53–69.

[7] K. Daniilidis and M.E. Spetsakis, Understanding noise sensitivity in structure from motion, in Y. Aloimonos, ed., *Visual Navigation*, Erlbaum, Mahwah, NJ, 1997, 60–88.

[8] Z. Duric, A. Rosenfeld, and J. Duncan, The applicability of Green's theorem to computation of rate of approach, CS-TR-3273, Center for Automation Research, University of Maryland, College Park, MD, 1994.

[9] Z. Duric and A. Rosenfeld, Image sequence stabilization in real time, *Real Time Imaging* **2**, 1996, 271–284.

[10] C. Fermuller and Y. Aloimonos, Direct perception of three-dimensional motion from patterns of visual motion, *Science* **270**, 1995, 1973–1976.



(a) (b)

Figure 10: (a) The optical flow field generated by an observer translating diagonally toward the scene in Figure 7. (b) Scaled (see text) velocity magnitude histograms of the flow field (see text).

[11] E. Francois and P. Bouthemy, Derivation of qualitative information in motion analysis, *Image Vision Computing* **8**, 1990, 279–288.

[12] J. Heikkonen, Recovering 3-D motion parameters from optical flow field using randomized Hough transform, *Pattern Recognition Letters* **16**, 1995, 971–978.

[13] R. Jain, Direct computation of the focus of expansion, *IEEE Trans. Pattern Analysis Machine Intelligence* **5**, 1983, 58–64.

[14] R. S. Jasinschi, A. Rosenfeld, P. Cucka, and K. Sumi, Discriminating 3D texture patterns: The velocity histogram method, CS-TR-2644, Center for Automation Research, University of Maryland, College Park, MD, 1991.

[15] R. Jasinschi, A. Rosenfeld, and K. Sumi, Perceptual motion transparency: The role of geometrical information, *J. Optical Society of America* **A9**, 1992, 1865–1879.

[16] H. Kalvainen, E. Oja, and L. Xu, Randomized Hough transform applied to translational and rotational motion analysis, in Proc. Intl. Conf. on Pattern Recognition, Vol. 1, 1992, 672–675.

[17] J.J. Koenderink and A.J. van Doorn, Invariant properties of the motion parallax field due to the motion of rigid bodies relative to the observer, *Optica Acta* **22**, 1975, 773–791.

[18] D.N. Lee, A theory of visual control of braking based on information about time to collision, *Perception* **5**, 1976, 437–459.

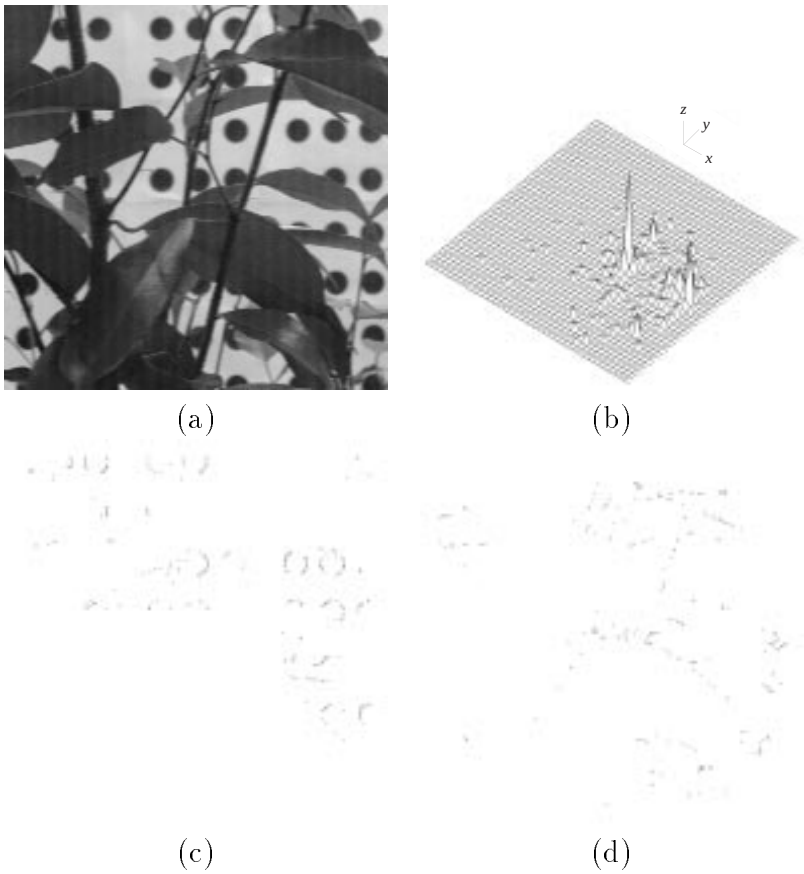


Figure 11: (a) One of a pair of images of a plant in front of a flat, patterned background, taken by a laterally translating camera. (b) The bimodal flow histogram. The taller peak (located near the origin, indicating lower image velocity) corresponds to the depth level of the background, and the shorter peak corresponds to the depth level of the plant; its position indicates higher image velocity. (c,d) Pixels that contributed to the two peaks.

[19] F. Meyer and P. Bouthemy, Estimation of time-to-collision maps from first order motion models and normal flows, in Proc. Intl. Conf. on Pattern Recognition, Vol. 1, 1992, 78–82.

[20] R.C. Nelson and J. Aloimonos, Finding motion parameters from spherical flow fields (Or the advantages of having eyes in the back of your head), *Biological Cybernetics* **58**, 1988, 261–273.

[21] R.C. Nelson and J. Aloimonos, Using flow field divergence for obstacle avoidance in visual navigation, *IEEE Trans. Pattern Analysis Machine Intelligence* **11**, 1989, 1102–1106.

[22] M. Nitzberg and D. Mumford, The 2.1-D sketch, in Proc. Third Intl. Conf. on Computer Vision, 1990, 133–137.

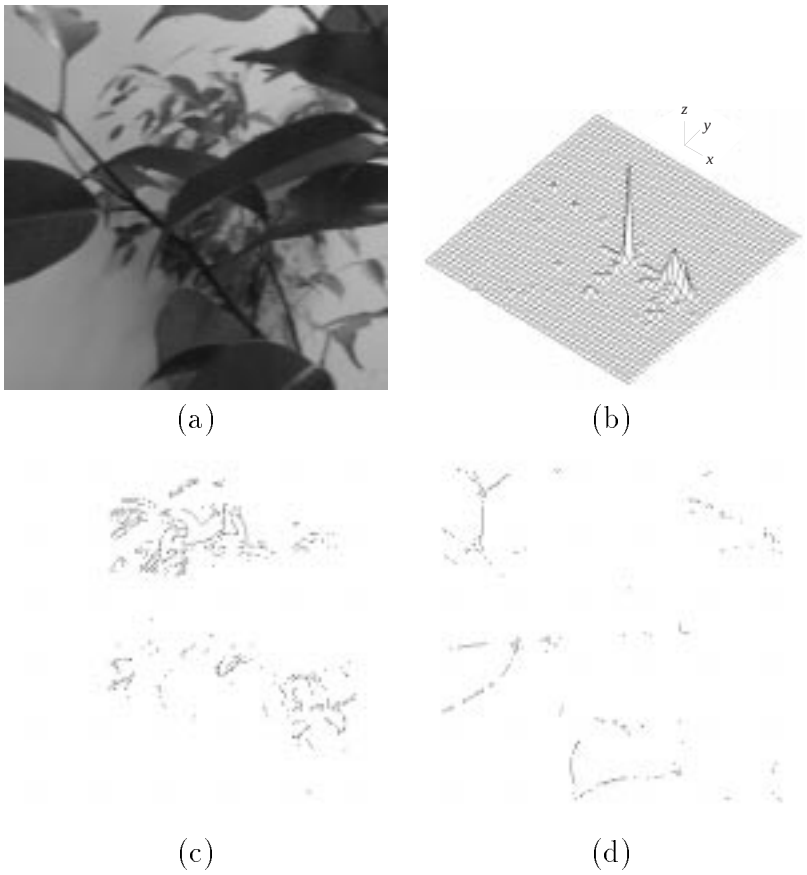


Figure 12: (a) One of a similar pair of images of two house plants, one in front of the other. (b) The bimodal flow histogram; the taller peak corresponds to the depth level of the rear plant, and the shorter peak to that of the front plant. Note the greater spread of the shorter peak due to the greater depth variability, and consequent image velocity variability, of the plant which is closer to the camera. (c,d) Pixels that contributed to the two peaks.

[23] M. Subbarao, Bounds on time-to-collision and rotational component from first-order derivatives of image flow, *Computer Vision, Graphics, Image Processing* **50**, 1990, 329–341.

[24] J.I. Thomas, A.R. Hanson, and J. Oliensis, Understanding noise: The critical role of motion error in scene reconstruction, in Proc. Fourth Intl. Conf. on Computer Vision, 1993, 325–329.

[25] W.B. Thompson and J. Kearney, Inexact vision, in Proc. IEEE Workshop on Motion, 1986, 15–22.

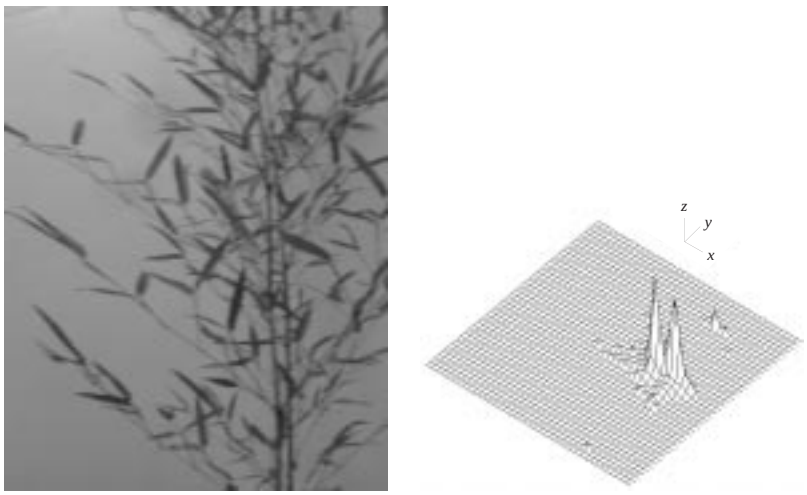
[26] W.B. Thompson and J.S. Painter, Qualitative constraints for structure from motion, *CVGIP: Image Understanding* **56**, 1992, 69–77.

[27] M. Tistarelli and G. Sandini, On the advantages of polar and logpolar mappings for direct estimation of time-to-impact from optical flow, *IEEE Trans. Pattern Analysis Machine Intelligence* **15**, 1993, 401–410.

[28] S. Ullman, The interpretation of structure from motion, *Proc. Royal Society London B* **203**, 1979, 405–426.

[29] D. Weinshall, Direct computation of qualitative 3D shape and motion invariants, *IEEE Trans. Pattern Analysis Machine Intelligence* **13**, 1991, 1236–1240.

[30] Y.S. Yao and R. Chellappa, Selective stabilization of images acquired by unmanned ground vehicles, *IEEE Trans. Robotics Automation* **13**, 1997, 693–708.



(a) (b)

(c) (d)

Figure 13: (a) One of a similar pair of images of two bamboo plants, one in front of the other. (b) The bimodal flow histogram; the taller peak corresponds to the depth level of the rear plant, and the shorter peak to that of the front plant. (c,d) Pixels that contributed to the two peaks.

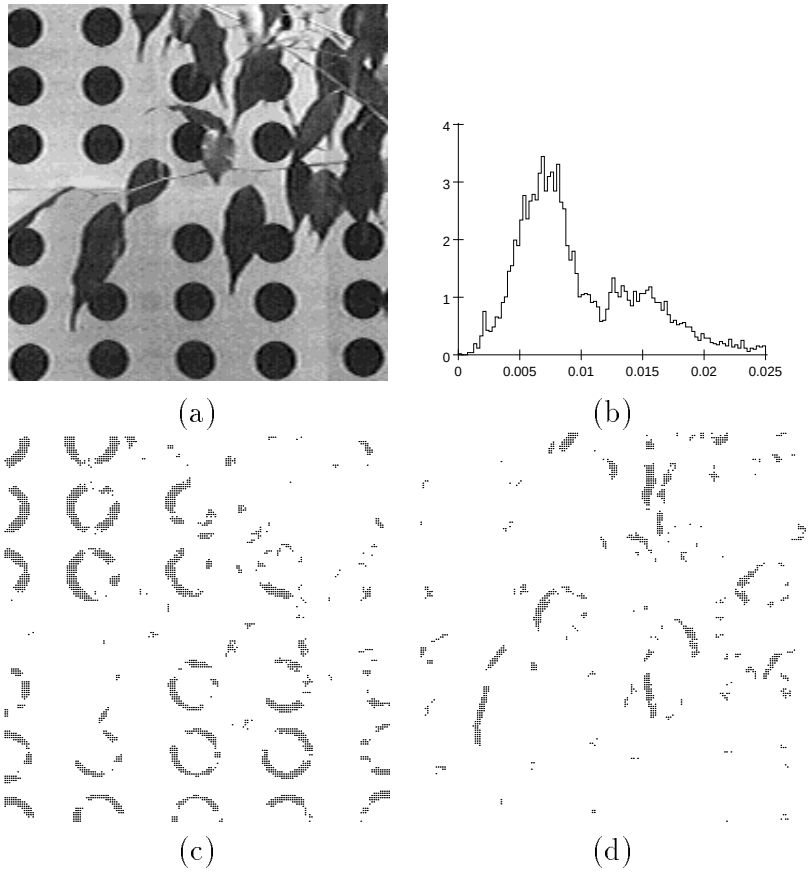


Figure 14: (a) One of a pair of images of a plant in front of a flat, patterned background, taken by a camera translating toward the scene. (b) The histogram of the flow magnitudes, scaled by the distances of the image points from the center of the image (where the focus of expansion is located). The scaled magnitude should be proportional to the closeness of the scene point to the camera. As we see, the histogram is in fact bimodal; the taller peak corresponds to the depth level of the backdrop, and the shorter peak to that of the plant. (c,d) Pixels that contributed to the two peaks.