Reprinted from SPIE Vol. 829—Applications of Digital Image Processing X

Range Estimation of Parts in Bins using Camera Motion

David Vernon* and Massimo Tistarelli**

*Department of Computer Science, Trinity College, Dublin, Ireland.
**Department of Computer and Systems Science, University of Genova, Italy.

Abstract

The central requirement in the bin-of-parts problem is to direct a robot manipulator to select, grasp, and remove an arbitrarily-oriented part (or object) from a bin of many such objects. This necessitates the estimation of the pose (position and orientation) of a partially-occluded object and, in general, its 3-D structure. The solution of such a problem using passive vision requires the use of sophisticated processing incorporating multiple redundant representations, such as stereopsis, motion, and analysis of object shading. This paper describes the first step is such an approach, that of determining a depth-map of the bin-of-parts, using optical flow derived from camera motion. Since the robotics environment is naturally constrained, simple camera motion can be generated by mounting the camera on the robot end-effector and directing the effector along a known path: the simplest motion, along the optical axis, is utilised in this case. For motion along the optical axis, the rotational components of flow are nil and the direction of the translational components is effectively radial from the fixation point (on the optical axis). Hence, it remains only to determine the magnitude of the velocity vector. Optical flow is estimated by computing the time derivative of a sequence of images, i.e., by forming differences between two successive images and, in particular, of contours in images which have been generated from the zero-crossings of Laplacian of Gaussian-filtered images. Once the flow field has been determined, a depth map is computed, initially for all contour points in the image, and ultimately for all surface points by interpolation.

1. Introduction

As robot vision matures, it is becoming increasingly desirable to extend its capabilities to include 3-D sensing. A significant goal of this capability is to solve the bin-picking problem in which a robot manipulator is required to identify and grasp an object jumbled in a bin of many such objects. While active sensing (and active triangulation in the form of light-striping, in particular) has popularily been used to provide range information, it has not yet been successfully employed in bin-picking. Furthermore, future robot vision applications will require increasing robustness such as is promised by anthropomorphically-motivated image understanding systems. A central tenet of image understanding research is the necessity of inferring the 3-D structure of the imaged scene through the use of several mutually-redundant visual cues such as sterepsis and visual motion.[1,2,3] One particularly useful paradigm for the generation of these disparate cues is based on an analysis of zero-crossing contours in Laplacian of Gaussian-filtered images:[4,5,6] these contours represent the position and orientation of intensity discontinuities in the image.[7,8] While the coherent integration of information derived from such filtered images, along with other visual cues such as shading, texture, and occlusion, is still in its infancy, progress is being made and it seems sensible to begin to deploy limited versions of this technology to industrial applications now, especially as hardware becomes available to implement the computationally expensive filtering stage. The research described in this paper endeavours to do just that, while at the same time providing a pathway for future developments.

In particular, this paper describes the use of a single camera, mounted on a robot end-effector, describing a simple camera motion along the optic axis to infer the depth of objects jumbled together in bins. The optic flow field resulting from this type of ego-centric motion is very easy to compute as all flow vectors are directed radially outward from the focus of expansion (FOE), i.e. the centre of the image.[9] Knowing the direction of the flow vector, the magnitude of the visual motion is directly derived from a time-derivative of a sequence of images acquired at successive points along the camera trajectory. Such use of a constrained camera motion is ideally suited to industrial environments, as manipulator arm trajectories can be specified at will. Furthermore, the technique facilitates the future incorporation of more general camera motion and eventually the mutual integration of information derived from other passive visual sensing.

2. The Bin-of-Parts Problem and Range Estimation

The bin-of-parts, or bin-picking, problem is widely recognised as being one of the most difficult tasks in robotics. Although a considerable amount of effort has been expended by the robotics and computer vision community in attempting to solve this problem, no general solution has yet been reported in the literature. Indeed, it is worth noting that so far only two broad approaches have been documented which appear to provide any realistic bin-picking capabilities. Details of these approaches may be found in, for example, [10] and in [11]. The central requirement in the bin-of-parts problem is to be able to direct a robot manipulator to select, grasp, and remove an artibrarily-oriented part (or object) from a bin of many such objects. These objects will, in general, be jumbled together and will occlude one another significantly. Thus, in order for the robot manipulator to be able to grasp the object, it must be able to identify the position and the orientation of the object, in spite of the fact that it would probably be partially hidden by one or more other objects in the bin. Usually, the bin-of-parts problem assumes that all the objects in the bin are, in fact identical. It seems likely that until this specific problem is solved the more general problem of several different parts jumbled in a bin together will remain a very far goal. Although some work has been done in approaching the bin-of-parts problem using tactile sensing (for example, [12]), and while many systems augment their sensory capabilities with tactile sensors, by far the greatest research effort is directed at a solution using vision.

The bin-of-parts problem, in which it is assumed that the objects are three-dimensional and that they can be artibrarily oriented in three-space, requires the identification of the six degrees of freedom of the object, corresponding to three translational coordinates and three rotational coordinates. Such an identification of position and orientation, often referred to as the object pose, is an extremely daunting task and it has been suggested that, while the problem is not intractable, it is extremely difficult.[13] This very difficulty has given rise to two distinct schools of thought, one holding that it is not absolutely necessary to determine the pose of an object in the bin and that the part can be removed using more realistic techniques and the other school adhering to the pursuit of the general pose estimation problem. The former approach owes much of its success to Kelley and his co-workers at the University of Rhode Island[11,13,14,15,16,17,18,19,20] while the latter school owes much to the research of Horn and Ikeuchi.[10,21,22]

Kelley and his co-workers at the University of Rhode Island hold that general fast algorithms for pose estimation in bins are a very formidable challenge. They take the stand that it is more reasonable to view the bin-picking problem from the point of view of simply identifying, not the pose and location of an object, but rather the position and local orientation of points on an object which would be likely to facilitate successful grasping with a robot manipulator. From this view point, the bin-picking problem reduces to one of identification of hold-sites rather than identification of objects. Since a hold-site will be intrinsically related to the structure of the robot end-effector, there will effectively be as many hold-site detection algorithms as there are generic gripper types.

The alternative, and much more formidable, objective of complete determination of the position and pose of an object while it is still in the bin has been pursued by Horn, Ikeuchi, and their associates. Their approach is to combine the cues given by photometric stereo which is capable of providing information about the local orientation of the surface of the object, and stereopsis, which provides one with the absolute range information needed for the reaching action. An earlier version of this approach[21] simply used the photometric stereo approach to determine the pose of the object and approached the grasping position along the line-of-sight of the camera, using a proximity sensor in the robot end-effector to determine when the gripper had actually reached the object; no absolute range of information was supplied by the vision system.

A point which is worthy of note in the context of this papers concerns the actual acquisition of range data which can be gathered in a variety of ways. Currents trends indicate a predeliction for active systems[23] and explicit such range information has been used to guide intensity image segmentation and analysis[24,25] Since the acquisition of such range data is central to 3-D image analysis, a slight deviation to sumarise approaches to range acquisition is in order. One can characterise 3-D image acquisition systems on the basis of two criteria[26,27,28]:

a. whether they are active or passive devices, and
b. whether they are triangulation and non-triangulation devices.

Active image acquisition system explicitly utilise contrived illumination to accomplish the range estimation. Passive image acquisition systems use ambient illumination.

The underlying principle of triangulation involves the construction of two straight lines in 3-D space, one based on the projection of a ray of light upon a point on an object, and the other based on its reflection to the sensor; the intersection of these two lines corresponds to the point of reflection (on the imaged object) in 3-D space. Using an active triangulation system, the two lines are easily constructed, knowing the origin of the ray of light and its direction and detecting (in a trivial manner) the point of incidence of the reflected ray on the image plane and its direction.

The "contrived" illumination used in active triangulation, frequently referred to as structured light, includes the projection of spots, lines and grids. Projection of spots requires 2-D scanning of thge light source, with consequent limitations in speed of image acquisition. Projection of grids requires no scanning but causes problems with identification of component lines in the imaged scene. This problem is particularly difficult when the scene being illuminated causes the image grid to be discontinuous. Single line scanning offer a compromise between these two situations.

Active image acquisition systems which do not use triangulation techniques rely on accurate measurement of time of flight of either sound (ultrasonic) or light (lasers) between the transmitter and the receiver. Since the transmitter and receiver are normally coincident, problems caused by object occlusion and "missing parts" do not arise.

Stereopsis, requiring an estimation of the shift of points on an object between two images due to displacement of their respective sensors, is a 3-D image acquisition approach based on passive triangulation. This shift, or stereo disparity, is inversely proportional to the distance between the imaging device and the points on the object. The principle problem of stereopsis lies in reliably identifying the corresponding points in images which were caused by a single point on the object.

Entirely passive techniques for 3-D imaging, which do not rely on triangulation to convey 3-D information, are numerous but are typified by an inherent ambiguity in derived conclusions regarding absolute range information. Such visual cues, e.g. shading occlusion, motion due to both observer (temporal parallax) and camera (optical flow), texture gradients, focussing, are capable, however, of providing extremely useful and reliable 3-D information when integrated in coherent manner, and especially when used with other passive range estimation techniques (such as stereopsis[10]).

In summary, 3-dimensional robot vision requires range information and measures of local surface orientation of objects to facilitate grasping by manipulators. While there are several systems capable of providing this information, most are based on active sensing. Purely passive systems, as typified by anthropomorphic visual systems, are ultimately more appealing, both on the basis of understanding perception and also on the basis of the potential ruggedness of such a system. Paradoxically, the requirement of passive systems to utilise several redundant visual cues to provide the 3-dimensional information is the source of difficulty in building such a system and the basis for its robust nature. The motivation underlying the work described in this paper is to begin is a small way to develop a purely passive system, by using one limited version of of one of the essential cues of this type of 3-D computer vision.

3. Inferring Depth from Ego-Centric Camera Motion

In the controlled environment of a robot workcell, with a camera mounted on the end-effector "looking" into the bin, a simple motion strategy can be adopted to determine the depth of the objects. The central ideal is to move the camera along the direction of the optical axis, which is the Z-axis of the camera-centred coordinate system. In this case, the direction of the optical velocities is constrained by the position of the FOE on the image, which is assumed to be at the centre of the image. Thus, the movement of the image points will simply be a contraction or expansion depending of the sense of the camera movement. The magnitude of the flow is unknown but it is computable by differentiating the image sequence with respect to time. If the luminance intensity does not change with time (i.e. there are no moving light sources in the environment) the component of the velocity vector, for each image point along the direction of the local intensity gradient, is given by:[29]

$$v^{\perp} = -(\partial I/\partial t)/|\nabla I| \quad (1)$$

where ∂ indicates the partial derivative operator and $|\nabla I|$ is the local intensity gradient.

The algorithm for computing depth can be summarised as follows:

a) Convolve the images with a Laplacian of Gaussian operator.[8]

b) Extract the zero-crossings, computing the local slope and orientation of each contour point.

c) Compute the difference between the convolution of successive frames of the sequence.

d) Compute the velocity component in the direction perpendicular to the orientation of the contour.

e) Compute the velocity along the contour, using the knowledge that (since the motion is ego-centric along the optic axis) the direction is radial from the centre of the image.

f) Search for the zero-crossings of the second frame projected from the first frame in the direction of the velocity vector.

g) Compute the depth map from this optic flow.

These steps form the body of an iterative scheme which allows one to compute the optical flow of a sequence of images.

Since all flow computations are done at image contours only (having been extracted using the Laplacian of Gaussian operator), the amount of data to be processed is limited and, futhermore, the effects of noise are less pronounced.

The computation of $v^{\perp}$ (the orthogonal component of velocity) is based on a computation of the time derivative usinga five-point approximation formula.[30] Despite the greater accuracy achieved in computing $v^{\perp}$ in this manner, a few errors are still recorded in the final flow, probably due to inaccruate localisation of the Focus of Expansion on the image plane. A significant improvement can be achieved by performing a contour-to-contour matching between successive frames, along the direction of the flow vectors. This operation tunes the length of the flow vectors to the correct size. Although a small difference between successive frames is required to guarantee the accuracy in the computation of the orthogonal component $v^{\perp}$, a long baseline is required for the range measurement; in fact, the error in depth is inversely proportional to the squared modulus of the optical velocities. For this reason, many images are considered and the flow field obtained for a sequence of, say, 10 images is used for range computation.

The depth, for each contour point, is computed by:

$$Z / W_z = D_f / |V_t| \quad (2)$$

where Z is the distance of the environmental point from the camera, D_f is the distance of the image point from the image centre (the FOE), V_t is the translational component of the flow, which coincides with the flow vector, and W_z is the camera velocity along the optic axis.

It is worth noting that equation (2) holds only for the contour points with a non-null velocity; whereas the camera velocity is actively controlled so as to be different from zero.

The errors in locating the true FOE, mainly due to inaccurate movements of the robot arm and misalignments of the camera, affect the flow computation and also the depth measurement, especially if the point is close to the FOE, where small errors in the vectors and/or in FOE localisation can induce large errors in computed depth.

4. Experimental Procedure and Results.

In order to evaluate this approach to inferring the depth of objects, motion sequences of two different scenes were generated. These scenes contained, a white plane surface inclined at 45°to the horizontal with black stripes at regular intervals and a basket of fruit (see diagrams 1.a and 2.a). To generate the motion sequences, a Panasonic CCD camera was mounted on the end-effector of a low-cost revolute manipulator (a SmartArms 6R/600). The robot was programmed to position the end-effector over the work-surface, with the camera pointing directly down, and to move downwards along a vertical trajectory. An image of the scene was acquired at 20mm intervals on this trajectory: a total of 10 images in total were acquired in each motion sequence.

Each of these 10 images were then convolved with a Laplacian of Gaussian mask (standard deviation of the Gaussian function = 4.0) and the zero-crossings contours were extracted. Since the Laplacian of Gaussian operator isolates intensity discontinuities over a wide range of edge contrasts, many of the resultant zero-crossings do not correspond to perceptually-significant physical edges. An adaptive thresholding technique[30] was employed to identify these contours and to exclude them from further processing (see diagrams 1.b and 2.b).

The zero-crossings contour images and their associated convolution images were them used to generate six time derivatives; since the time derivative utilises a five point operator, the time derivative can only be estimated for images 3, 4, 5, 6, 7, and 8. The associated orthogonal component of velocity is then computed, followed by the ego-centric motion optical flow vectors (diagrams 1.c and 2.c). An extended flow field was then estimated by tracking the flow vectors from image 3 through images 4, 5, 6, 7 to image 8 on a contour-to-contour basis, i.e. tracking a total of five images (see diagram 1.d and 2.d). For the sake of comparison, two depth images (representing the distance from the camera to each point on the zero-crossing contour) were generated: one from the ego-centric motion flow vectors of image 3 (diagrams 1.e and 2.e) and one from the tracked velocity vectors (diagrams 1.f and 2.f).

5. Discussion

One of the lessons learned while conducting this research concerns the manipulator arm articulation. The robot used in this series of experiments was a five degree of freedom anthropomorhpic-type robot with five revolute joints. Although the requisite arm trajectory involved a change in just one degree of freedom (a translation in the direction of the Z axis), this particular type of articulation necessitated the actuation of three joints (the shoulder, elbow, and wrist) to accomplish the movement. Since the robot itself was not of industrial standards, exhibiting poor accuracy and repeatability, the net error compounded from the servoing of three joints was significant. This resulted in some lateral motion of the camera and small changes in the direction of the optical axis. The effect was to introduce errors into the local estimates of the time derivative and hence into the magnitude of the flow vectors and the subsequent estimates of depth. It should be noted that since the magnitudes of the ego-centric optical flow vectors are themselves quite small, even minor lateral motion can prove to be source of significant error. A robot with a prismatic (sliding) joint in the direction of the Z axis, such as the IBM series of robots, would minimise this problem as just one joint would need to be actuated. In this case, lateral motion should be negligible.

The computation of depth is based on the expression:

$$Z = W_z D_f / |V_t|$$

Since W_z is constant and D_f is merely a spatially-varying scale factor, this implies that, neglecting scaling, the computation of depth is dependent wholly on the measurement of the flow vector magnitude. Furthermore, the quantisation (or resolution) of depth will be dependent on the quantisation of the magnitude of the velocity vector. Unfortunately, for the narrow range of depth encountered in both the bin-of-parts and the calibration of objects used in this paper, the magnitude of the velocity vectors derived from the time derivative is typically less than four pixels implying a resolution of only 1 in 4 and hence the depth measurements have similarly low resolution. To increase the resolution of the depth measurement one needs to increase the magnitude of the velocity vector and this can be achieved either by coarser temporal sampling of the image sequence (leaving larger distances between snapshots) or by tracking contours over an extended sequence of images. In the latter case, the matching procedure sometimes fails to find correspondences between

contours and large sections of the optical flow field are left undetermined (see diagrams 1.e and 2.e) resulting in total loss of information regarding depth. In the former case, the estimate of the time derivative becomes unreliable if the sampling is too coarse. Since most of the errors in tracking are caused by lateral shifts of the manipulator between images, it is expected that much improved performance, and depth resolution, will be achieved with a prismatic industrial manipulator.

The final depth images derived from a single time derivative, bearing in mind the coarse quantisation, are in fact very useful: one can immediately see that the apple in the centre of the bin is closer to the caamera than the bananas are, and that the strawberries are the furthest objects from the camera. The final range images exhibit a certain amount of local variation which is caused by errors in the estimation of some optical flow vectors and hence the corresponding depth image value. Since depth varies smoothly over most of an object surface, these errors can be reduced by the application of a local smoothing operator.

It is notable also that, as expected, the depth estimates close to the focus of expansion are unreliable; this is most obvious in diagram 1 and diagram 2, where no depth values are generated in the centre of the cone and the inclined plane. Because of this, it is important to ensure that the focus of expansion (i.e. the centre of the image) is as far as possible from the objects which are of greatest interest.

6. Conclusions

This paper has described a first attempt to utilise passive vision, in the guise of analysis of simple optical flow based on ego-centric camera motion along the optic axis, to infer depth of objects in bins. Because of the very small variations in depth exhibited by such parts and hence the relatively small magnitude of the velocity vectors derived from a single time derivative of the image sequence, the effective depth resolution is quite low for the range of distances under consideration. Nevertheless, the technique proved very successful in facilitating the discrimination between those objects which lie close to the top of the pile (and can thus be grasped without collisions between the end-effector and other objects) and those which lie deeper in the pile. The main advantage of the technique is that it can be generalised to cater for more complex camera motion and the task of integrating such motion with an analysis of stereo disparity is actively being pursued. Thus, the approach represents a useful starting point from which a robust passive robot vision system, based on the mutual integration of several visual cues, can be developed.

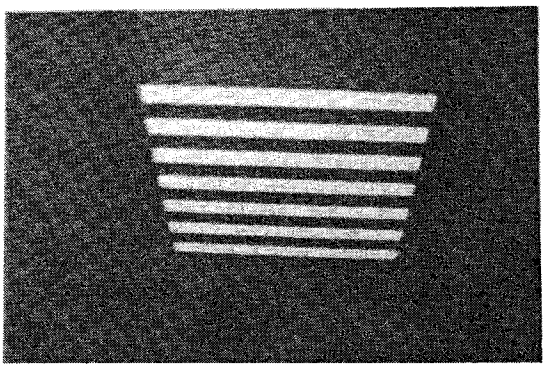
7. Acknowledgments

This research was supported by the European Strategic Program for Research and Development in Information Technology (ESPRIT) under project P419: Image and Movement Understanding.

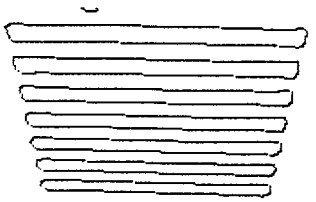
8. References

1. M. Brady "Computational Approaches to Image Understanding", ACM Computing Surveys, 14(1), 3-71 (1982).
2. T.O. Binford "Survey of Model-based Image Analysis Systems", The International Journal of Robotics Research, Vol. 1(1), 18-64 (1982).
3. J.M. Tenenbaum, H.G. Barrow, and R.C. Bolles "Prospects for Industrial Vision", SRI International Technical Note 175 (1978).
4. D. Marr Vision, W.H.Freeman and Co., San Francisco (1982).
5. W.E.L. Grimson From Images to Surfaces, The MIT Press, Cambridge, Massachusetts (1981).
6. E.C. Hildreth The Measurement of Visual Motion, The MIT Press, Cambridge, Massachusetts (1983).
7. D. Marr "Early Processing of Visual Information", Philosophical Transactions of the Royal Society of London, B275, 483-524 (1976).
8. D. Marr and E. Hildreth "Theory of Edge Detection", Proceedings of the Royal Society of London, B207, 187-217 (1980).
9. G. Sandini and M. Tistarelli "Analysis of Camera Motion through Image Sequences" in Advances in Image Processing and Pattern Recognition, V. Cappellini and R.Marconi (Editors), Elsevier Science Publishers B.V. (North-Holland), 100-106 (1986).
10. K. Ikeuchi, H.K. Nishihara, B.K. Horn, P. Sobalvarro, and S. Nagata "Determining Grasp Configurations using Photometric Stereo and the PRISM Binocular Stereo System", The International Journal of Robotics Research, 5(1), 46-65 (1986).

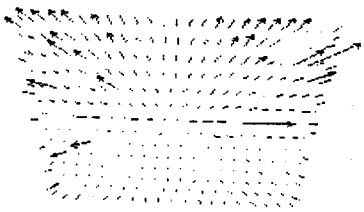
11. R.B. Kelley, H.A.S. Martins, J.R. Birk, and J-D. Dessimoz "Three Vision Algorithms for Acquiring Workpieces from Bins", Proceedings of the IEEE, 71(7), 803-821 (1983).
12. T. Sakata "An Experimental Bin-Picking Robot System", Proceedings of the 3rd. International Conference on Assembly Automation, 615-626 (1982).
13. J-D. Dessimoz, J.R. Birk, R.B. Kelley, A.S. Martins, and I. Chi Lin "Matched Filters for Bin Picking", IEEE Tranactions on Pattern Analysis and Machine Intelligence, PAMI-6(6), 686-697 (1984).
14. R.B. Kelley "Heuristic Vision Algorithms for Bin Picking", Proceedings of the 7th. Conference on Industrial Robot Technology, Gottenburg, Sweden, 599-610 (1984).
15. R.B. Kelley, J.R. Birk, H.A.S. Martins, and R. Tella "A Robot System which Acquires Cylindircal Workpieces from Bins", IEEE Transactions on Systems, Man, and Cybernetics, SMC-12(2), 204-213 (1982).
16. R. Kelley, J. Birk, J. Dessimoz, and R. Tella "Acquiring connecting Rod Castings using a Robot with Vision and Sensors", Proceedings of the 1st. International Conference on Robot Vision and Sensory Controls, IFS (Conferences) Ltd., U>K., 169-178 (1981).
17. R. Kelley, J. Birk, D. Duncan, H. Martins, and R. Tella "A Robot System which Feeds Workpieces directly from bins into machines", Proceedings of the 9th. International Symposium on Industrial Robotics, 309-355 (1979).
18. J.R. Birk, R.B. Kelley, and L. Wilson "Acquiring Workpieces: Three Approaches using Vision", Proceedings of the 8th. International Symposium on Industrial Robots, Stuttgart, West Germany, 724-733 (1978).
19. J.R. Birk, R.B. Kelley, and J-D. Dessimoz "Visual Control for Handling Unoriented Parts", Proc. SPIE, 281, 169-175 (1981).
20. J.R. Birk, R.B. Kelley, and H.A.S. Martins "An Orienting Robot for Feeding Workpieces into Stored Bins", IEEE Transactions on Systems, Man, and Cybernetics, SMC-11(2), 151-160 (1981).
21. B.K.P. Horn and K. Ikeuchi "Picking Parts out of a Bin", AI Memo 746, MIT AI Lab (1983).
22. K. Ikeuchi "Determining Attitude of Object from Needle Map using Extended Gaussian Image", AI Memo 714, MIT AI Lab (1983).
23. P.J. Besl and R. Jain. "Three-Dimensional Object Recognition", ACM Computing Surveys, 17(1), 75-145 (1985).
24. M.J. magee, B.A. Boyter, C. Chien, and J.K. Aggarwal "Experiment in Intensity Guided Range Sensing Recognition of Three-dimensional Objects", IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-7(6), 629-637 (1985).
25. R.Y. Wong and K. Hayrapetian "Image Processing with Intensity and Range Data", IEEE Computer Society Conference on Pattern Recognition and Image Processing, Las Vegas, NV, 518-520 (1982).
26. W.D.M. McFarland and R.W. McLaren "Problems in Three-Dimensional Imaging", Proc. SPIE, 449, 148-157 (1983).
27. W.D. McFarland "Three Dimensional Images for Robot Vision", Proc. SPIE 442, 108-116 (1983).
28. G. Mooney and N. Murphy "Three-Dimensional Computer Vision fro Robotic Assembly", Preliminary Report, National Institute for Higer Education, Dublin, Ireland (1986).
29. B.K.P. Horn and B.G. Schunck "Determining Optical Flow". Artifical Intelligence, 17(1), 185-204 (1981).
30. G. Sandini and M. Tistarelli "Analysis of Object Motion and Camera Motion in Real Scenes", Proc. of the IEEE International Conference of Robotics and Automation, San Francisco, 627-632 (1986).
31 D. Vernon "On the Properties of Zero-Crossing Contours in Laplacian of Gaussian-filtered Images", Technical Report No. CSC-87-04, Dept. Computer Science, Trinity College, Dublin (1987).



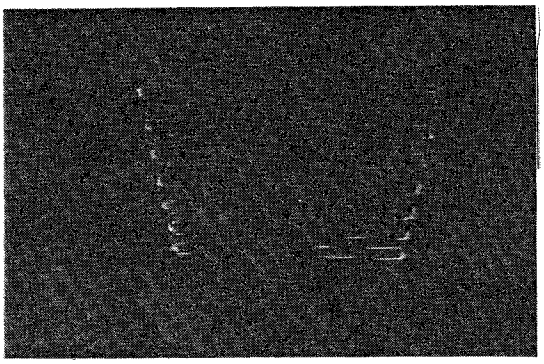
(a)



(b)

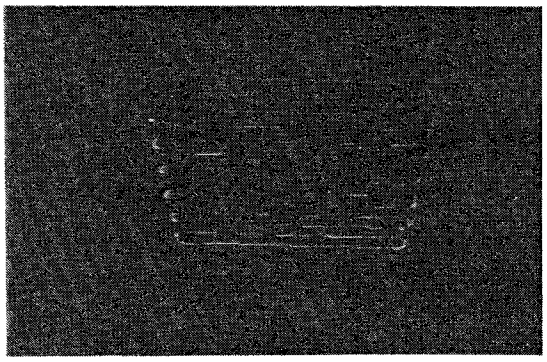


(c)



(d)

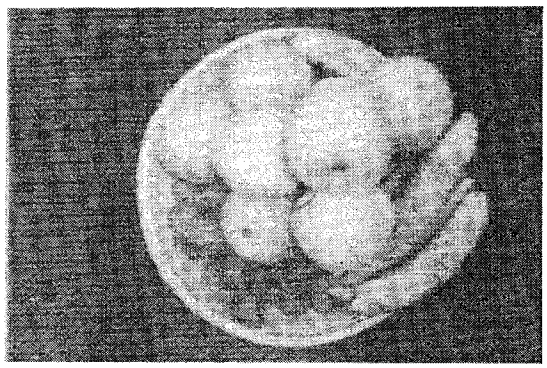
(e)



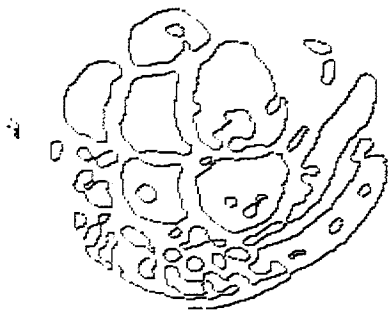
(f)

Diagram 1

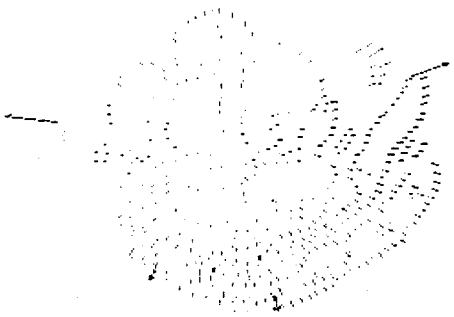
(a) Image 1 of inclined plane. (b) Selected zero-crossings. (c) Ego-motion: no images tracked. (d) Depth: no images tracked. (e) Ego-motion: five images tracked. (f) Depth: five images tracked.



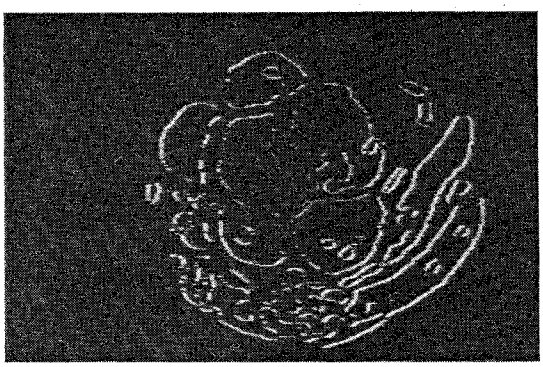
(a)



(b)

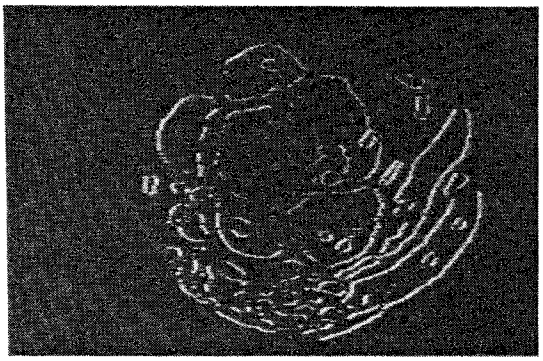


(c)



(d)

(e)



(f)

Diagram 2

(a) Image 1 of basket of fruit. (b) Selected zero-crossings. (c) Ego-motion: no images tracked. (d) Depth: no images tracked. (e) Ego-motion: five images tracked. (f) Depth: five images tracked.