GRID CODING: A PREPROCESSING TECHNIQUE FOR ROBOT AND MACHINE VISION

P. M. Will
K. S. Pennington

Computer Science and Applied Research Departments
IBM Thomas J. Watson Research Center
Yorktown Heights, New York, U.S.A., 10598

ABSTRACT: The problem of machine vision as evidenced In the various robot projects in existence is attacked by analogy with the supposed nature of human visual processing in that edges are enhanced, texture is examined and various heuristic approaches are studied. This paper describes a non-anthropomorphically based method of decomposing a scene subjected to a special form of illumination into elementary planar areas. The method consists in coding the various planar areas as the modulation on a spatial frequency carrier grid so that the extraction of the planar areas becomes a matter of linear frequency domain filtering. The paper also addresses the application of grid coding to other problems in recording and extracting information from 3-D images.

INTRODUCTION

The problem of machine vision is tackled essentially in the same way by each of the robot projects in existence (1, 2, 3, 4). Each uses the work of Roberts (5) as the basic starting point. An implicit conceptual analogy is made with the human visual processing system and the approach concentrates on discontinuities in the image Indicative of edges which have many anthropological connections. The procedure, which can be termed classic, is therefore to doubly-differentiate the image in two dimensions to isolate edges, next to fit straight lines to the data, extract corners and then, from these derived measurements, compute the probability that the data fits one of a set of geometric shapes used as conceptual building blocks. Great success has been obtained from this method in spite of formidable difficulties at every stage of the process. The isolation of edges in an image is fundamentally a high pass filtering operation and is noise enhancing, hence many of the problems In the conventional approach. The situation is saved in a simple picture, however, since the process of taking this string of bits and fitting straight lines to them is a task that can be reasonably accomplished in a digital computer of reasonable size.

It is significant to note that the contribution of Guzman (6) lies in the analysis of a scene of extreme complexity which has been specified to the program as a line drawing. There is a lack of a methodology for obtaining a line drawing from a three-dimensional scene in a reasonably non-noise producing manner. For example, the robot projects rely only on exterior edges or silhouettes of the objects being examined because of the uncertainty introduced by the edge enhancement operator. There have been many proposals for attacking the task of extracting planar areas directly. One such suggested approach is, starting from a random point, to examine the neighboring points to see if they fit a consistent criterion such as, for example, the hypothesis that they are derived from a uniformly reflective surface which is illuminated by a source of light of known a priori spatial distribution. The concept of analyzing scenes based on local texture dependencies is described by Gibson (7) who argues that this process is that employed by humans. If it could be automated by a digital computer this idea would not involve high pass filtering but would be a local low pass operation and therefore less noisy.

In general, the processing of visual images is performed for several different reasons; for instance, correction for geometric or radiometric distortion, enhancement of the image for improved visual acceptability and the extraction of specific features from the image. In extraction applications, the desired information should be exaggerated as much as possible to ease the decision-making problem while the undesired should be deleted. A common occurrence in extraction applications is the tendency to postpone the difficult parts of the process until later stages, when a heuristic method is the sole recourse. In robot vision, for example, the three-dimensional structure of the object is usually derived from a single projective view of the object and furthermore only edge data in that view is used. An example, Figure 1,

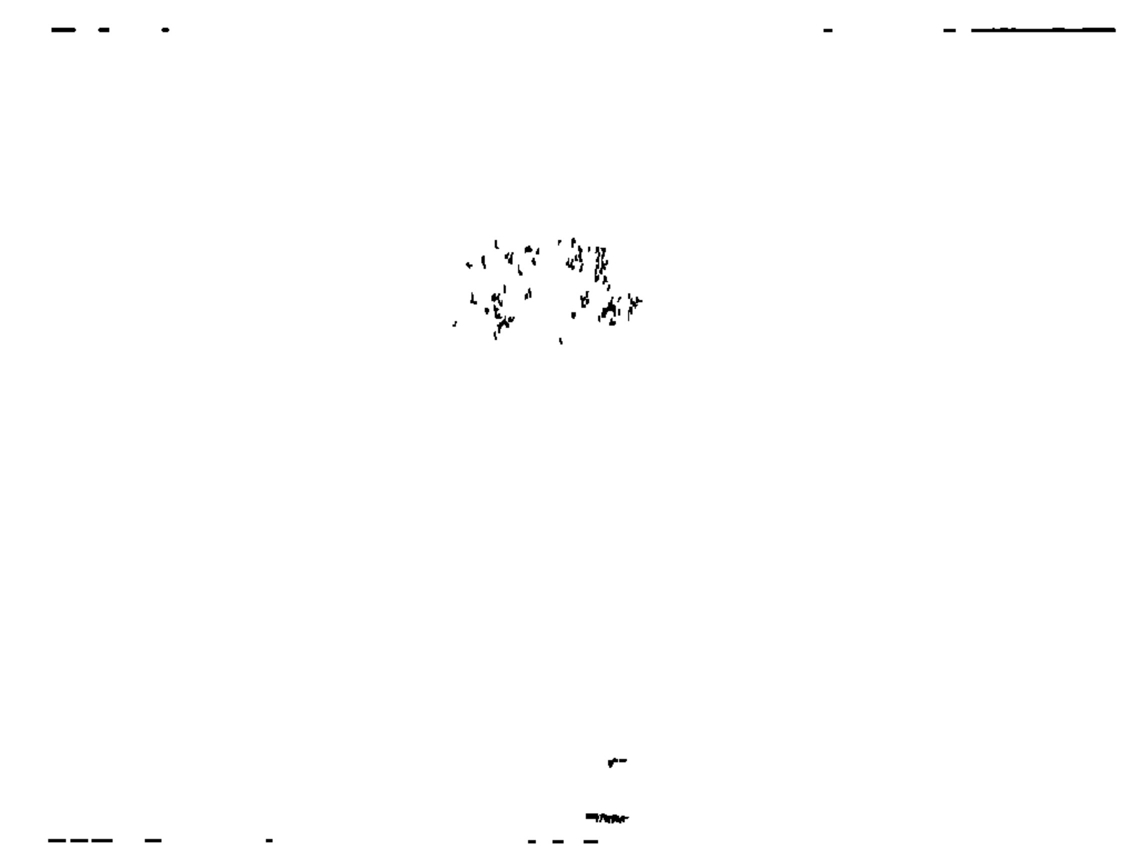

Figure 1. Normal photograph of a cube resting on a plane of support

shows a typical scene, a cube resting on a table, in which the illumination is sufficiently uniform that the photometric difference between the two front faces is so small that the problem of determining the inside edge by two dimensional double differentiation is almost impossible. This difficulty points out the importance of heuristics involving solid bodies, planes of support, continuity of planar surfaces, etc.

It would be preferable to examine the scene directly for three-dimensional properties, e.g., planar areas, and then, if necessary, define edges as the intersections of geometric planes rather than as points of photometric discontinuity.

Grid Coding

The technique described in this paper, which we call grid coding, (10), often gives an alternative to the above heuristic approach by making use of extra degrees of freedom allowed by controlling the illumination and by stereoscopy (in an unusual form) to extract three-dimensional measurements of the scene to be analyzed.

Grid coding was motivated by the problem of vision described above as well as by the desirability of using recent developments in Fourier filtering. It Is now possible to process image data extremely rapidly in the Fourier domain in a digital computer via the Fast Fourier Transform technique (8) and similarly high speed processing may be achieved optically in a coherent optical processor (9). These processes are well understood and it is tempting to try to apply them here. In the digital case, the Fast Fourier Transform may involve much computation but the computation itself can be considered to be a very simple step — the black, box approach — in that if the scene decomposition problem could be framed as a spatial frequency filtering problem, then a heuristic free method would be available as desired.

Given the two-dimensional Fourier processing is highly desirable since it is conceptually simple and easy to implement, it is appropriate to ask the question: "What is the Fourier spectrum which is most easily filtered for the extraction of any desired Information?" The simplest spectrum is that in which the desired information is contained in a pair of delta functions of specified spatial position and amplitude. Any convenient set of delta functions would do as well. Pursuing this reasoning, we note that the inverse transform of a pair of delta functions Is a sinusoidal grating. A harmonic set of delta functions gives, on inverse transforming, some other periodic structure — a special case being a Ronchi ruling which Is an equal mark/space grid. Thus it follows, albeit tenuously, that if the information to be extracted can be CODED as the modulation of a suitable chosen GRID, then the desired information in the space domain will have a Fourier spectrum which will exhibit the coding of the information as the distribution of a delta function array. The Fourier domain processing to extract this information will then be simple. The remainder of this paper involves the search for methods of achieving this form of modulation or coding of the feature to be extracted from the Image so that the standard demodulation theory of communications technology can be applied to the resultant.

The discussion in this paper deals with methods of coding and extracting planes from visual scenes so that by defining an edge as an intersection of planes the processor output is directly suitable for a program such as Guzman's. Alternatively, a new program could be written in Gusman's philosophy to use these planar descriptions directly.

Grid coding dispenses with analogies to anthropology and codes the scene directly in terms of machine readable code for planes.

The simplest method of grid coding a scene is simply to shine "grid coded" light upon it. Figure 1 has shown a normally illuminated scene, Figure 2, on the other hand, shows a scene in

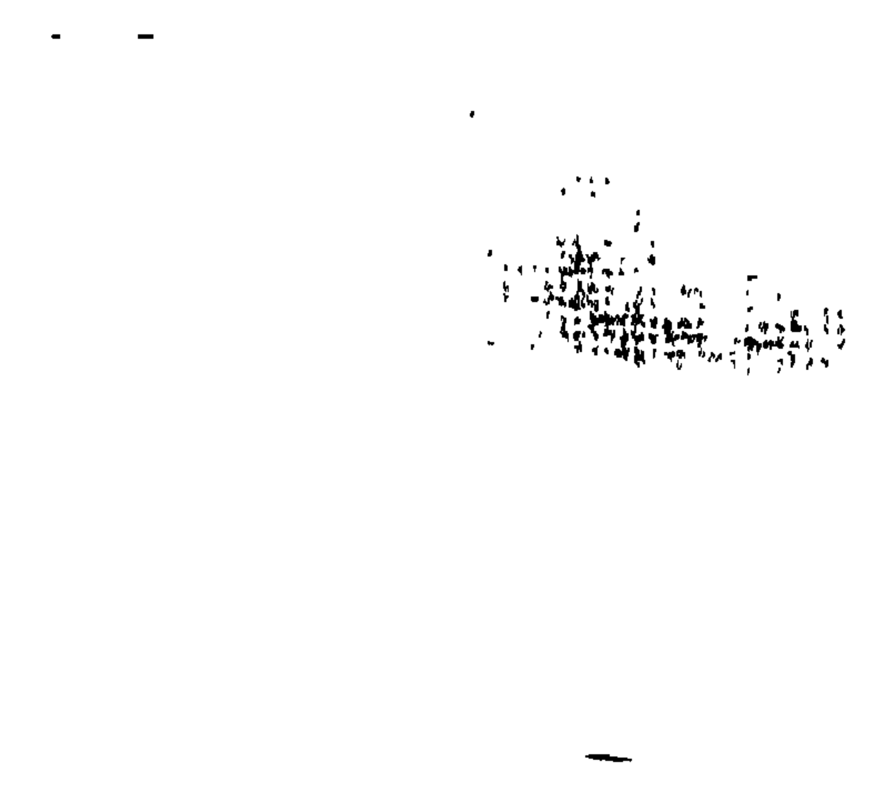

Figure 2. A photograph of the same scene as shown in Figure 1 with the addition that the illumination is grid coded.

which the illumination is derived from a projector containing a crossed grating. The scene is recorded with a camera offset from the illuminating source to an extent which can be arbitrarily chosen. The topology of the cube in Figure 2 is recognizable because of the different striping on planes of different orientation.

It can be shown that adjacent faces of a convex body (i.e., faces in J-D continuity) are uniquely coded but in general grid coding gives no unique decomposition. These details are discussed fully elsewhere (11).

PROCESSING OF GRID CODED IMAGES

The projection of crossed grids on a scene results in the image of the scene being recorded as the distortion of the grids, where the distortion is dependent on the loca] geometry of the scene. The transformation matrix necessary to restore the individual quadrilaterals in the received modulation back into the squares In the projected grid contains the direction of the loca] normal to the surface. The most general method of processing grid coded pictures is to perform the above projective mapping but the simple and well known Fourier filtering can also be applied to the processing task.

A planar area coded with crossed grids has a two-dimensional Fourier transform which is a crossed set of harmonically related delta functions in the spatial frequency domain. Thus the description of the plane has been compressed in the Fourier domain and is easy to extract by simple band pass filtering. The inverse transform can then reconstruct the image of the plane isolated from all other planes. A higher level program can then deduce the interrelationship of planes and objects.

Experiments were performed optically and digitally on Fourier processing of grid coded images. A coherent optical system was used in the former case and the systems described in References 12 and 13 were used for the latter.

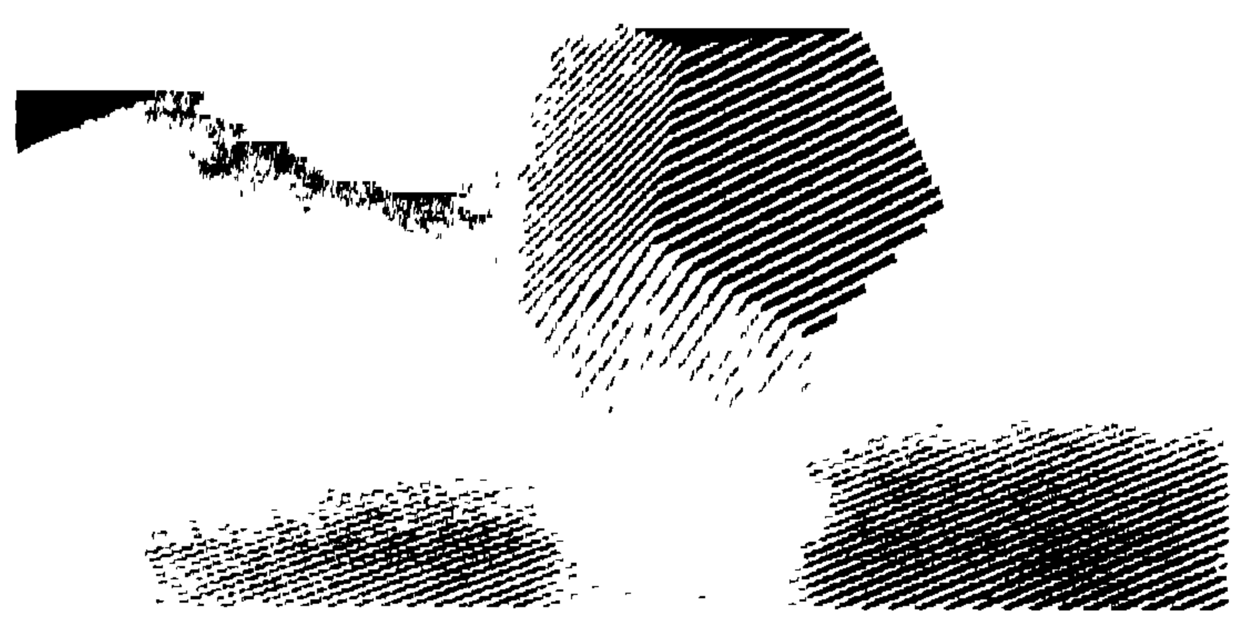

Figure 3. Original photograph of a grid coded dodecahedron. This was then scanned at 256 x 256 x 8 bit resolution and used as a test example.

As an example of the processing technique, a typical scene was taken to be the dodecahedron shown in Figure 3. The Fourier spectrum is shown in Figure 4. A filter consisting of a 1°

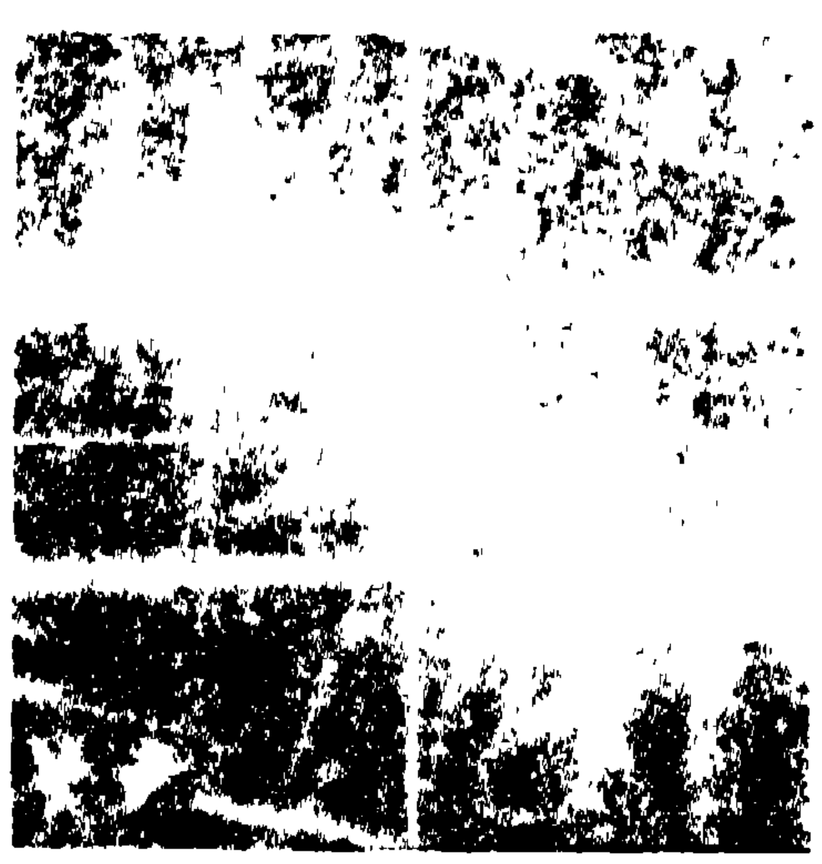

Figure 4. Fourier transform of the dodecahedron Resolution 256 x 256.

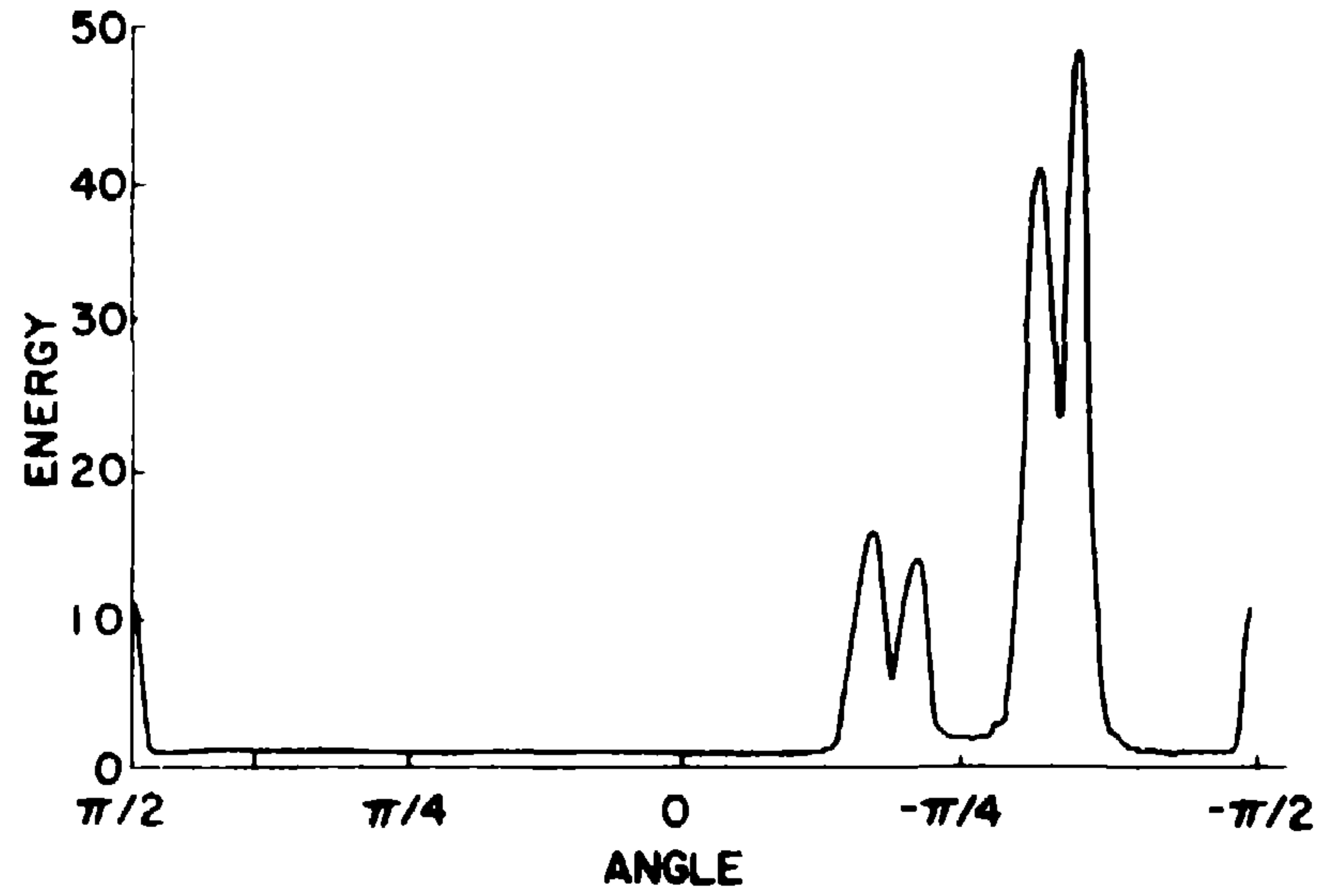

Figure 5. Graph of energy versus angle for the spectrum of Figure 3. The energy was computed by passing the spectrum of Figure 3 through a set of 1° wide sector filters. This curve was used automatically to generate filters for the extraction of planes.

wide sector of a circle with an all-pass response in the radial direction was applied to the spectrum and the energy transmitted at various rotations was computed. The graph of energy vs angle is shown in Figure 5. A set of data dependent sector filters was designed with the angular width and center angular spatial frequency chosen to pass the peaks of the spectrum. Each peak corresponds to a plane (or a set of planes at the same angle). The results of applying these filters are given in Figures 6-9, showing that a decomposition of the scene, including the background, into its elementary planar components, is possible.

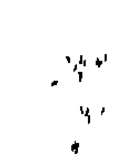

Figure 6. Plane - extracted from the dodecahed ron of Figure 3. The filter was 8° wide centered at -38°.

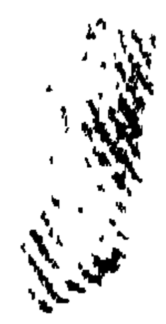

Figure 7. Plane extracted by a filter of 8 at -47°.

Figure 8. Plane extracted by a filter of 8° at -64°.

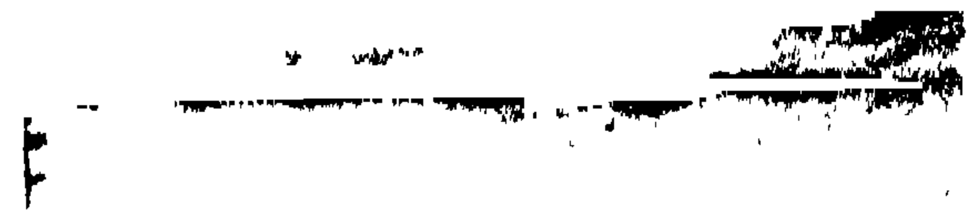

Figure 9. Plane of support extracted by a filter of 8° at -68°.

NEW FILTERING APPROACHES

As has been mentioned above, a plane surface in a general scene is not uniquely coded in angle or spatial frequency carrier shift. It is possible to remove this restriction by passing the illumination through a grid structure in which the code carries positional information, for example by using a shift register derived code plate or the grid known in optics as a linear zone plate. With these grids the local area carries its own position information and it is possible to determine three-dimensional position by triangulation (11).

OTHER APPLICATIONS

a) Stereo

Grid coded pictures have enough Information in them that a corresponding view can be automatically generated to give a stereoscopic pair. A simple support for this statement can be seen by considering that the grid structure is known a priori and the received image is a warping of the initial structure. All that is necessary to create at least one stereo pair is for the individual planar areas to be extracted from the received picture, warped back into the image of the input by a stretching operation and placed in proper position. This produced image when taken with the received image form a stereo related pair. These operations, which are described in more detail in (11) , have not all been reduced to practice but are clearly possible and give rise to interesting possibilities for further study.

b) Differences

It is also possible to apply grid coding to computing differences in images. This operation, non-trivial optically, can be achieved by applying grid coding to form a composite Image from a montage of pieces from each of the two input images. If the first input image is multiplied by a binary mask, say a set of stripes, the second image multiplied by the complement of the mask and these images (which are now geometrically mutually exclusive) are added, then this composite image will contain stripes where differences exist and will be smooth otherwise. The filtering to extract differences is exactly similar to the previous discussion. The interesting points here are that non-uniform but complementary masks can be used to emphasize differences in any given area independent of other areas in the scene and that the technique may facilitate frame to frame encoding for a video phone.

Difference extraction has been discussed fully elsewhere (14).

c) 3-D Recording

It is also possible to consider recording the full 3-D information of a scene by grid coding the image obtained by moving the camera. In this case each image point gives rise to a streak whose length is proportional to the range of the corresponding object point. This streak can thei. be grid coded by suitably shuttering the camera at the image plane. Processing to extract range is simple, particularly if the shuttering is periodic. In this case each streak is encoded as a linear array of points whose period is proportional to range (15).

CONCLUSION

This paper has described some experiments in coding a scene by controlling the illumination in a spatial sense. It has been shown that the coded scene can easily be processed to extract planes by a heuristic-free method. The computations required are not complex and the method is particularly attractive since it is tailored to machine computation and therefore conceptually satisfying in robot vision applications.

The heuristic interface is therefore claimed to be pushed back a little deeper into the recognition process where it probably belongs.

REFERENCES

1. M. Minsky and S. L. Papert, "Research on Intelligent Automata" status Report II, Project MAC, MIT, September 1967.
2. J. McCarthy, "Plans for the Stanford Artificial Intelligence Project", Stanford University Report, April 1965.
3. N. Nilsson and B. Raphael, "Preliminary Design of an Intelligent Robot," Computers and Information Sciences-II, Academic Press Inc., New York, N. Y., 1967.
4. R. J. Popplestone, "Freddy in Toyland," Machine Intelligence 4, B. Meltzer and D. Michie (eds.) pp. 455-462 American Elsevier (1969).
5. L. G. Roberts, "Machine Perception of Three Dimensional Solids," in Optical and Electro-optical Information Processing, J. T. Tippet, et al (eds.) MIT Press, Cambridge, Mass. 1965.
6. A. Guzman, "Decomposition of a Visual Scene into Three Dimensional Bodies," AF1PS Proc. Fall Joint Comp. Conf. Vol. 33 pp. 291-304 (1968).
7. J. J. Gibson, The Perception of the Visual World, Houghton Mifflin, Boston 1950.
8. J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," J. Math of Comp., 19, No. 90, pp. 297-301, 1965.
9. A. Vander Lugt, "A Review of Optical Data Processing Techniques," Optica Acta, 15, pp. 1-33, 1968.
10. K. S. Pennington, G. L. Shelton, P. M. Will, "Grid Coding — A Technique for Image Analysis," presented at IEEE International Symposium on Information Theory, Noordwik, The Netherlands, June 15-19, 1970.
11. P. M. Will, K. S. Pennington, G. L. Shelton, "Grid Coding". To be submitted for publication.
12. N. Herbst and P. M. Will, "Design of an Experimental Laboratory for Pattern Recognition and Signal Processing," IBM Research Report RC 2619, Sept. 1969; also published in Proceedings of Computer Graphics - 70 April 1970, Brunei University, London.
13. P. M. Will and G. M. Koppelman, "MFIPS — A Multifunctional Digital Image Processing System," IBM Research Report, April 1970, RC-3313.
14. K. S. Pennington, P. M. Will and G. L. Shelton, "Grid Coding: A Technique for the Extraction of Differences from Scenes," Optics Communications, Vol. 2, No. 3, pp. 113-119, August 1970.
15. K. S. Pennington and P. M. Will, "A Grid-coded Technique for Recording 3-dimensional Scenes Illuminated with Ambient Light," Optics Communications, Vol. 2, No. 4, pp. 167-169, Sept. 1970.