MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 536   May 1979

SEQUINS and QUILLS -- representations for surface topography.

Berthold K. P. Horn

Abstract. The shape of a continuous surface can be represented by a collection of surface normals. These normals are like a porcupine's quills. Equivalently, one can use the surface patches on which these normals rest. These in turn are like sequins sewn on a costume.

These and other representations for information which can be obtained from images and used in the recognition and description of objects in a scene will be briefly described.

Gradient Images.

The shape of a continuous surface can be represented by a collection of surface normals. These normals are like a porcupine's quills or the spines sticking out of a cactus. Equivalently, one can use the surface patches on which these normals rest. These in turn are like sequins sewn on a costume or the scales of a fish.
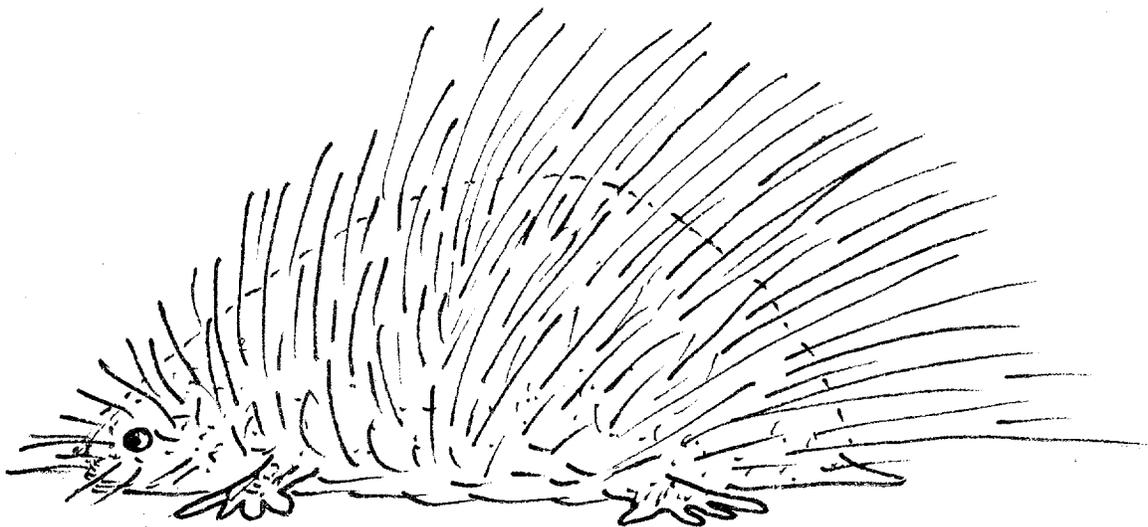
Surfaces can be approximated "in the large" using parameterized models such as generalized cylinders [Binford 1971] or "in the small" using local patches. The distinction between these two extremes is not unlike that between the fitting of parameterized standard functions and the use of chains of local approximations. Each mode of representation is uniquely suited to some applications. Also, one form often is used as an intermediate step in the derivation of the other.
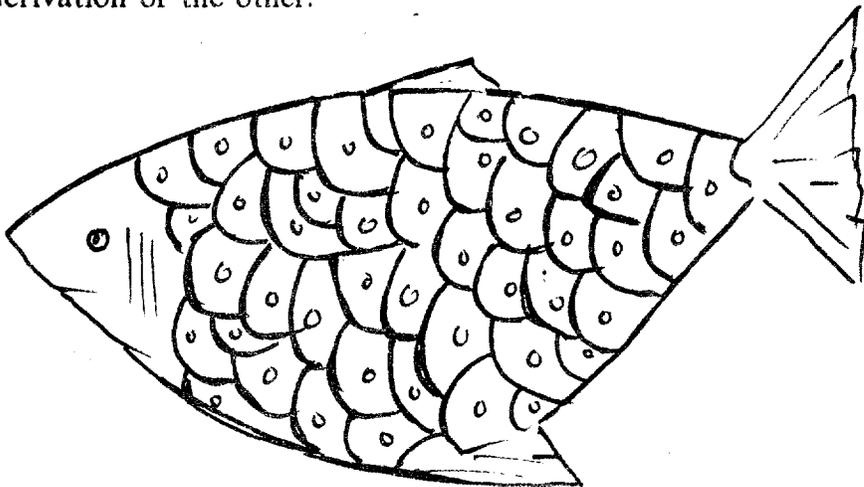
One view of machine vision is that information about the scene being imaged is to be abstracted from the raw brightness values. The estimation of scene attributes leads to results useful in the recognition and description of objects. Naturally, the desired information must be obtainable from the image and useful in some application. Surface shape has these properties and is conveniently represented in terms of surface orientation of numerous local facets. In addition one might also extract surface albedo, illumination and other properties. When this kind of information is maintained in registration with the raw image, it may be referred to as a component of the 2-½ D sketch [Marr & Nishihara 1978] or an intrinsic image [Barrow and Tenenbaum 1979]. Here I will concentrate on the surface orientation component.
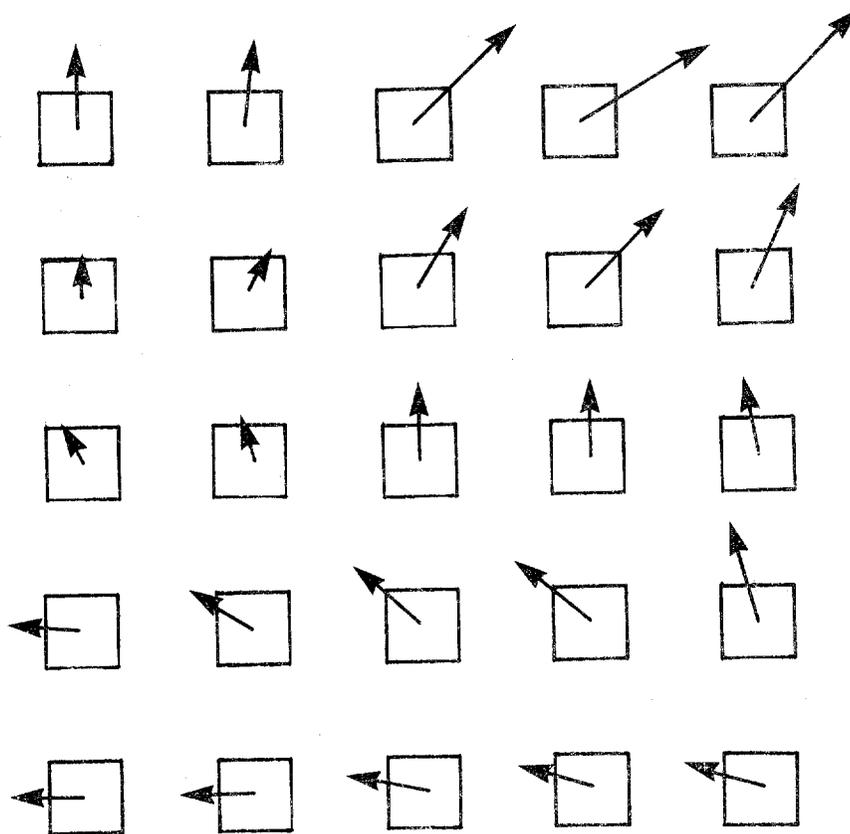
The representation described is only of interest because it arises naturally in certain computations using image data and is useful in recognition tasks. It is also valuable in the determination of the position and attitude of objects in space.

Computations leading to local surface description.

The shape of a surface can be computed from a single image using the imaging equation,

$$E(x,y) = R(p,q),$$

where $E(x,y)$ is the image irradiance at the point (x,y) in the image. The patch on the object imaged at (x,y) has gradient (p,q). The function $R(p,q)$, giving scene radiance as a function of surface gradient in a viewer-centered coordinate system, is called the reflectance map. The above first order non-linear partial differential equation in the two variables x and y can be solved for the depth, z, along so called characteristic strips, by finding an equivalent set of five ordinary differential equations [Horn 1970].

The resulting representation of the surface and the method of characteristics itself have disadvantages which led to the suggestion that difference methods on a uniform grid might be more suitable [page 192, Horn 1970]. The notion was that such numerical methods would be similar to ones used for solving second order partial differential equations. These appear to have been studied more intensively because of their greater interest to physicists. Some recent approaches to the shape from shading task can be viewed in this light [Woodham 1977, Brooks 1978, Strat 1979].

· Representation in terms of local surface normal also has been thought to have advantages when rotations of an object are to be studied [page 225, Horn 1977].

"Pseudo-local" computations.

Some computations on images appear to be "global" in the sense that each value depends on all or a large subset of the inputs. Many of these computations however can be carried out by iterated local operations or in a locally connected network which includes feedback elements. This is true in particular of global operators which happen to be inverses of local operators.

Convolution with

$$(1/2\pi) \log_e(1/r), \quad \text{where} \quad r = \sqrt{x^2 + y^2}$$

for example, is the inverse to the application of the laplacian operator

$$(\partial^2/\partial x^2) + (\partial^2/\partial y^2).$$

The latter operator is clearly local in nature, while its inverse appears to be global. One can however exploit the "pseudo-local" nature of this inverse using either iterative or feedback schemes [page 290, Horn 1974]. In the application to the lightness task, the result is an "albedo image" and an "illumination image", both registered with the raw image.
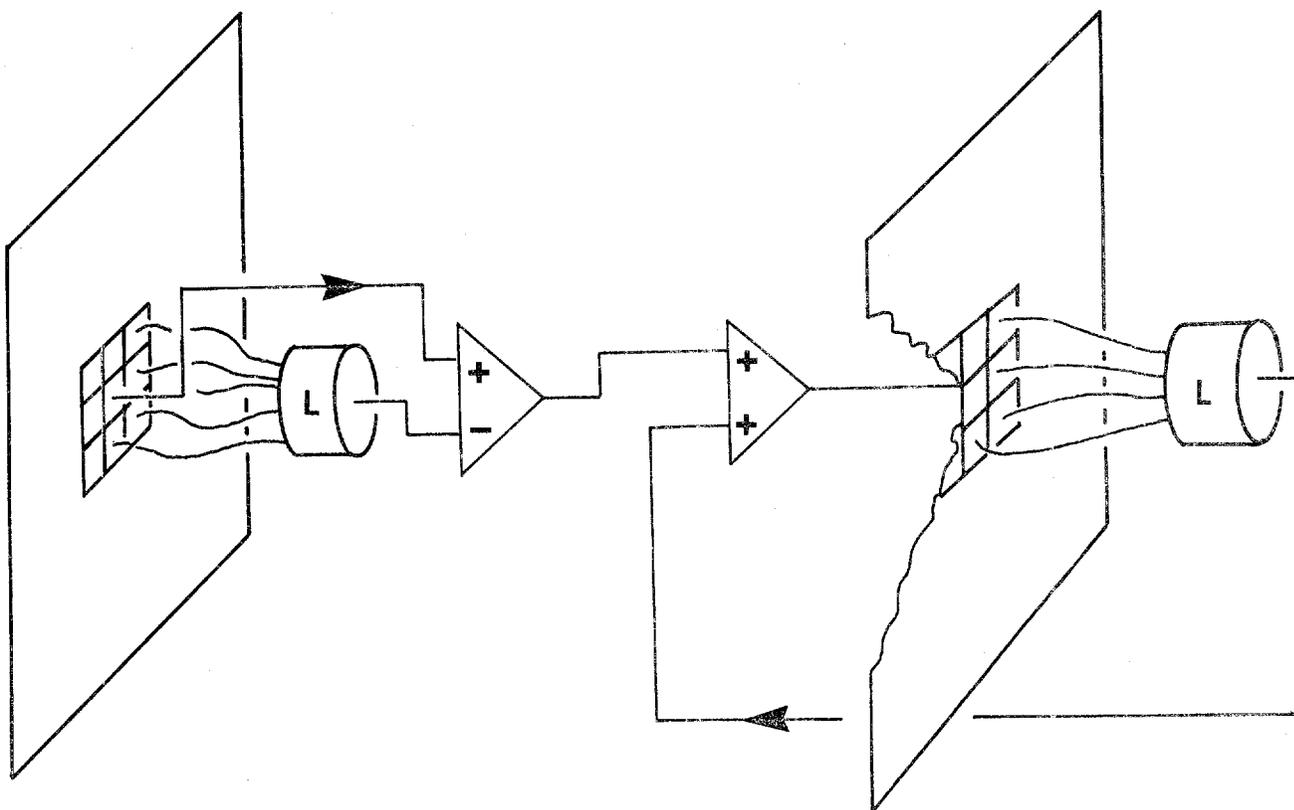
Similar methods have been called relaxation computations because of the analogy with iterative "relaxation" methods [Rosenfeld 1978] for solving large sets of simultaneous equations, often discretized versions of partial differential equations. There are also similarities with "cooperative computation", in an interconnected network of computing elements, where each node or cell is thought of as having the ability to perform simple computations based on the states of its neighbors [Marr & Poggio 1976].

In the case of interest to us here, the "forward", local operation is the image formation, which we are attempting to "invert" by solving the imaging equation.

Local Constraints.

The underlying justification for these kinds of algorithms is often stated in the form of constraints to be applied at each node. Consider for example two images of the same scene with different lighting. At each point of the image one finds two constraints:

$$E_1(x,y) = R_1(p,q),$$
$$E_2(x,y) = R_2(p,q),$$

where E_1 and E_2 are image irradiances measured in the first and second images respectively, while R_1 and R_2 are the corresponding reflectance maps. The two equations can be solved for the local gradient (p,q). Since the equations are typically non-linear there may be more than one solution.
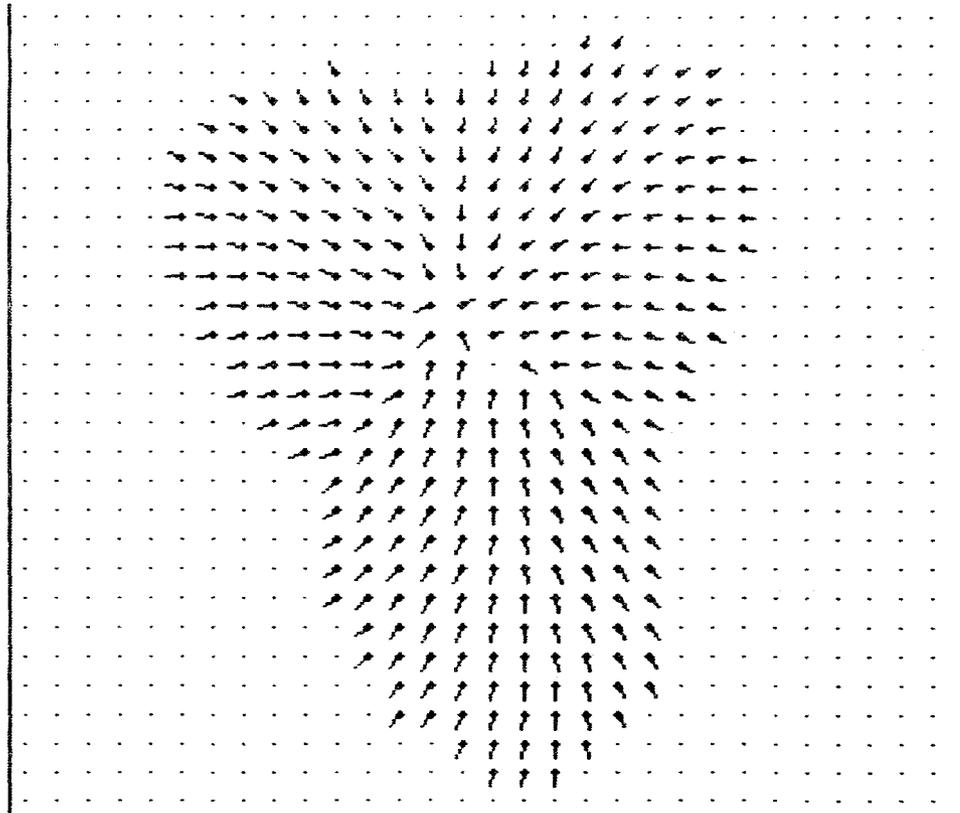
This method and similar ones using multiple images taken from the same viewpoint have been referred to as "photometric stereo" methods [Woodham 1978]. They produce descriptions of surface shape in just the form here advocated: values for p and q at each node of a mesh of points registered with the image.

Smoothness and Continuity.

Local constraints may arise from implicit assumptions rather than direct measurement on images. In one stereo algorithm, for example, it is assumed that an entity in one image matches at most one entity in the other, and that nearby image points usually have similar disparities [Marr & Poggio 1976].

Similarly, when the shape of a surface is to be computed from the shading, it is reasonable to assume that what is imaged constitutes one surface rather than a collection of disconnected patches. This adds a second constraint,

$$\partial p / \partial y = \partial q / \partial x \quad \text{or} \quad \oint (p,q) \cdot (dx, dy) = 0,$$

to the one resulting from the imaging equation. Iterated computations based on the two constraints lead to solutions for the surface shape [Brooks 1978, Strat 1979]. Here too, the end product are values of p and q on a mesh of points.
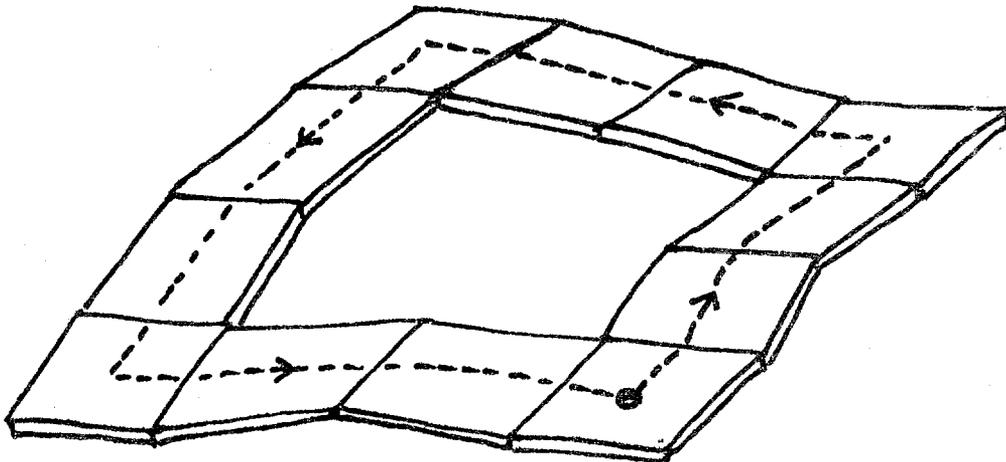
The values of p and q assigned to each point are updated during each iteration in a way designed to bring them into closer agreement with the two local constraints. One way to do this is to consider e_{NE}, e_{NW}, e_{SW}, and e_{SE}, the errors in four loop integrals along square contours passing through a given cell. We wish to minimize,

$$e^2 = e_{NE}^2 + e_{NW}^2 + e_{SW}^2 + e_{SE}^2,$$

with respect to p and q. The values p' and q' which best "relieve the strain" in the local patch of the solution are linear combinations of the values of p and q of neighboring cells.
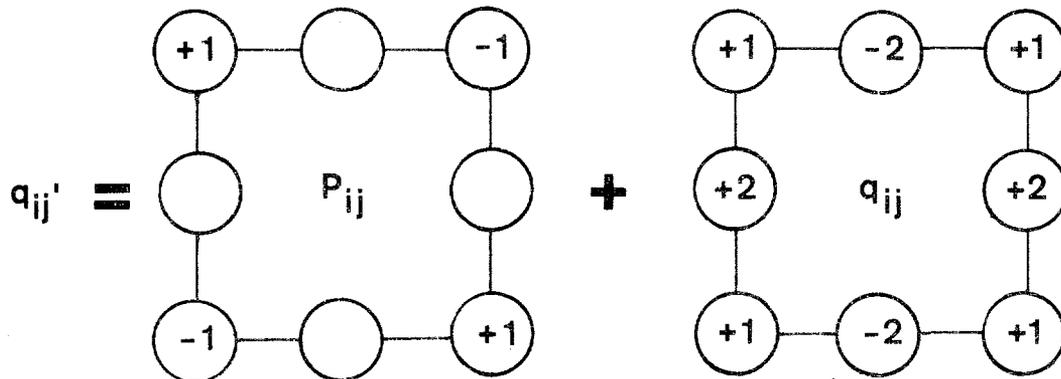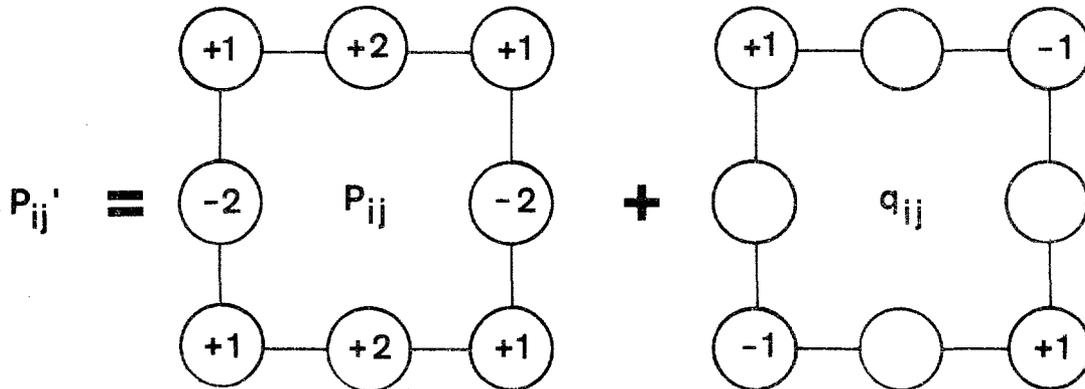
The values found this way do not in general satisfy the imaging equation however. One really wants to minimize the error subject to the constraint $R(p,q) = E(x,y)$. One can rewrite the sum of squares of errors as

$$e^2 = 4\,(p - p')^2 + 4\,(q - q')^2 + e_0{}^2.$$

Introducing the Lagrangian multiplier, λ, one can restate the problem: Minimize

$$e^2 + \lambda \, (R - E) \quad \text{subject to} \quad R = E.$$

Differentiating with respect to p and q one gets,

$$8 \, (p - p') + \lambda \, R_p = 0, \quad \text{and}$$
$$8 \, (q - q') + \lambda \, R_q = 0.$$

Elimination of the Lagrangian multiplier λ leads to,

$$[p-p', q-q', 0] \times [R_p, R_q, 0] = 0,$$

where '$\times$' denotes the cross-product of two vectors. The desired solution then is just the point on the contour $R(p,q) = E(x,y)$ nearest to (p', q'). Many details remain to be worked out, including conditions for convergence.
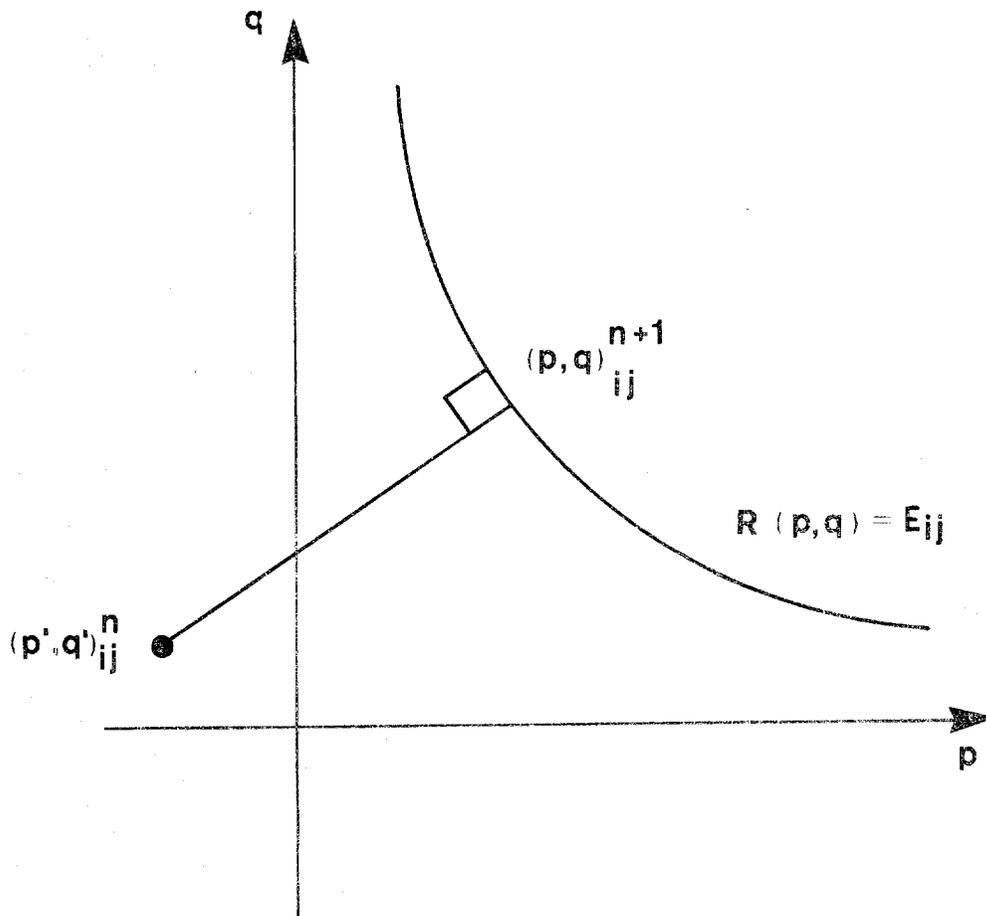
Recent work on "optical flow" [Clocksin 1978, Prazdny 1979] suggests that in the analysis of moving images one also obtains representations of surfaces in terms of local normals. Similar statements can be made about the analysis of "texture gradients" [Kender 1979].

We have seen that this representation for surfaces arises naturally in a number of machine vision algorithms. We now have to discuss the properties and applications of this representation. To start, we consider a mapping which preserves only part of the information, but greatly simplifies matters:

Extended Gaussian Images.

Imagine moving all of the surface normals to the origin, that is, discarding the information about their position on the surface of the object. The result is a spiky ball with varying numbers of spines sticking out in different directions. Here we assume that there is a fixed number of patches per unit surface area and that a unit normal is erected on each patch. Regions on the object with small curvature lead to concentrations of normals pointing in more or less the same direction, orthogonal to the average tangent plane of the region. If the surface normals are unit vectors, their end points will lie on a unit sphere. We can replace the sea-urchin like arrangement of spines with a spherical or gaussian image, where a dot on the unit sphere marks every place a unit vector touches the surface.
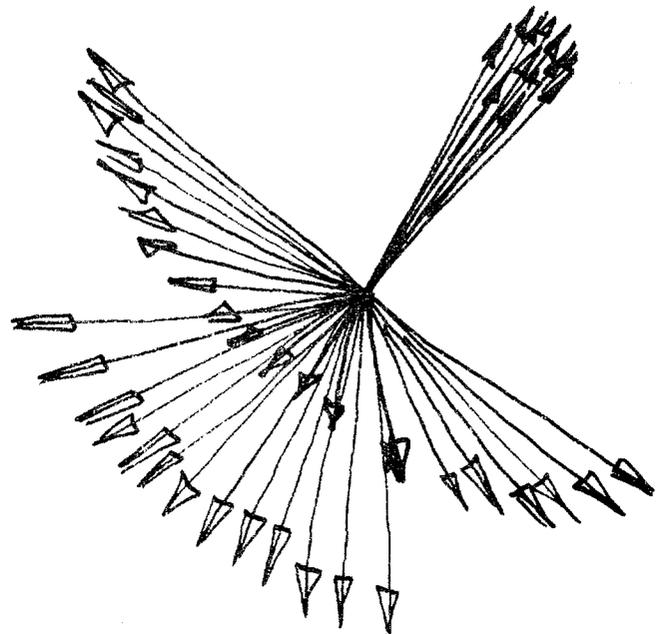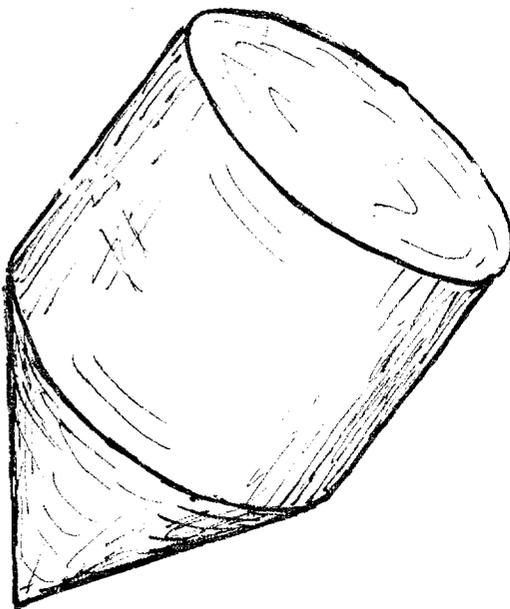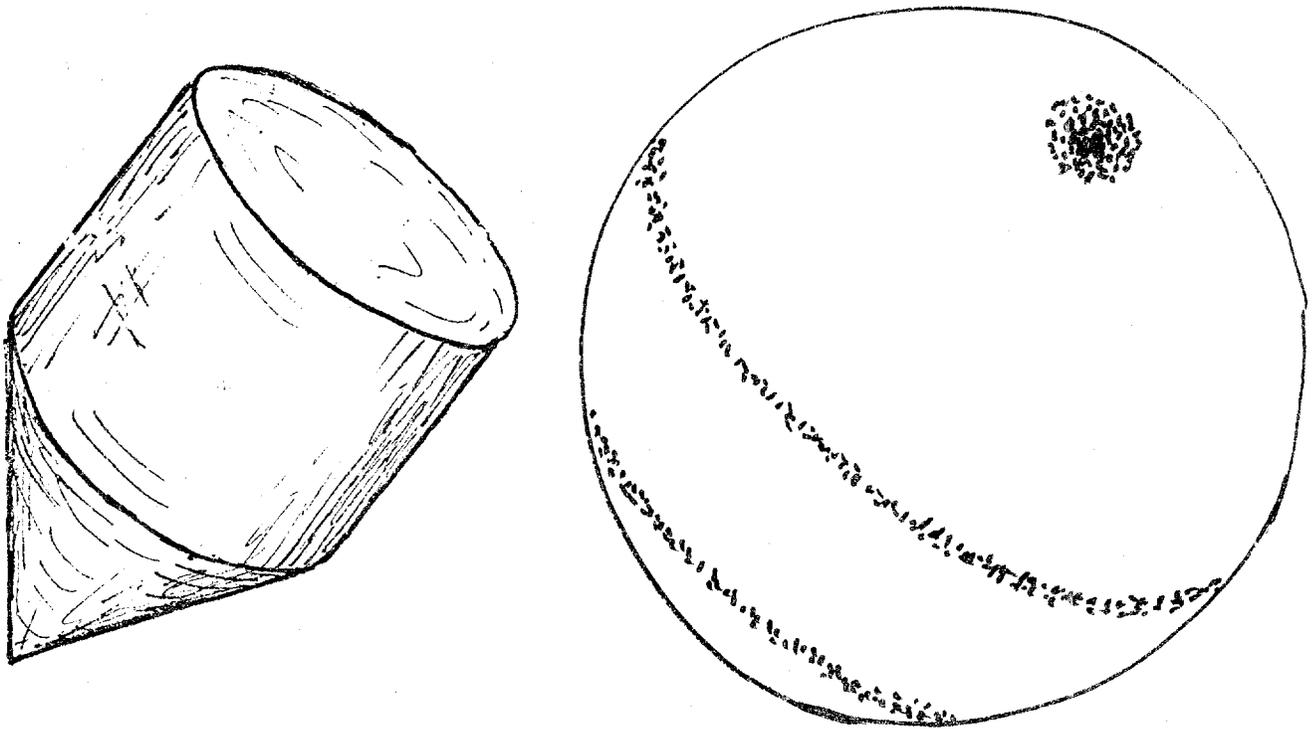
If the original data describing the surface comes from an image, one half of the sphere will be unmarked, since surfaces turned away from the viewer are invisible. Further, the representation is biased towards patches lying near normal to the viewer, since these are not foreshortened by the imaging projection. That is, if we now assume a fixed number of patches per unit *image* area, then a surface region tilted with respect to the viewer contributes fewer marks to the gaussian image since it covers fewer mesh points in the image.

Aside from this, the result obtained is a discretized version of the gaussian image of a smooth object, where the value at a particular point on the sphere is proportional to the product of the principal radii of curvature [Pogorelov 1956] at the corresponding point on the object [Hilbert & Cohn-Vossen 1952]. If the object is not convex, several distinct surface patches may contribute to a given point on the sphere.

It seems natural to use the spherical image in recognition as well as for the determination of the attitude of an object in space [Smith 1979]. Symmetries of the object are reflected as symmetries in the gaussian image and features of low gaussian curvature show up as high concentrations of marks. Segments of developable surfaces such as planes, cylinders and cones, give rise to impulsive distributions in this representation which can be used in matching of prototypes against data derived from images. Brute force techniques, such as search through the space of possible attitudes can also be used, provided that model values are multiplied by

$$\max\,[0, \cos\,e]$$

before entering in comparison, where e is the angle between the outward surface normal and the viewing direction.

For simple geometric figures the prototypes may be represented by gaussian images given as closed formulas. More complicated objects lead to discretized numerical versions. Comparing two such discretized gaussian images is not trivial and may depend on the use of high-order semi-regular tesselations of the sphere [Fejes Toth 1964, Pearce & Pearce 1978]. An alternative involves lining up the principal axes of inertia of the two gaussian images and matching the moments [Smith 1979].

One big advantage of this representation is the ease with which arbitrary rotations can be handled. While the position of points in the image is altered in complicated ways when the object is rotated, the surface normals undergo a simple transformation that can be represented by an orthonormal 3 X 3 matrix.

Uniqueness of the Gaussian Image.

The extended gaussian image of a convex polyhedron consists of an impulse for each face of "weight" equal to the area of the face. The impulse lies on the sphere where the surface normal for that face pierces the unit sphere. Minkowski showed that two convex polyhedra are identical if corresponding faces have equal areas and the same surface normals [Lysternik 1963]. The representation is therefore unique for convex polyhedra. Unfortunately the proof of this result is not constructive and at this time no effective method exists for the construction of the convex polyhedron from its extended gaussian image.

Not all gaussian images correspond to real (closed) objects. Gaussian images must satisfy the center of mass conditions, that is, the center of mass of the "weights" associated with the marks on the surface must be at the center of the sphere. Equivalently, surface normals of length proportional to the area of the corresponding faces must form a closed chain when stuck together end to end. This can be shown by equating the cross-sectional areas of the object when viewed from two diametrically opposite directions [Smith 1978].

Less is known about gaussian images of convex objects with continuous surface normals, but it is likely that similar results apply to them too. This representation thus appears to be useful in the recognition of *convex* objects and in the determination of their attitude in space. It may also be useful in the general case, provided details are checked out after initial alignment using the methods discribed above for gaussian images of convex objects.

Summary.

Some representations for surfaces which can be obtained from images and used in the recognition and description of objects in a scene have been briefly described.

Acknowledgments.

References.

Barrow, H. G. and Tenenbaum, J. M. (1979) "Recovering Intrinsic Scene Characteristics from Images," in *Computer Vision Systems*, E. Riseman and A. Hanson (eds.) Academic Press, N. Y.

Binford, T. O. (1971) "Visual Perception by Computer," *IEEE Conf. on Systems and Control*, Miami, December 1971.

Brooks, M. J. (1978) "Surface-normals from closed paths," Department of Computer Science, University of Essex, Colchester, England.

Clocksin, W. F. (1978) "Determining the orientation of surfaces from optical flow," *Proc. AISB/GI 1978*, Hamburg, pp 93-102.

Fejes Toth, L. (1964) *Regular Figures*, Pergamon Press, N.Y.

Hilbert, D. and Cohn-Vossen, S. (1952) *Geometry and the Imagination*, Chelsea, N. Y.

Horn, B. K. P. (1970) "Shape from Shading: A method for determining the shape of a smooth, opaque object from one view," MAC TR-79, Project MAC, M.I.T., Cambridge, Mass.

Horn, B. K. P., (1974) "Determining Lightness from an Image," *Computer Graphics and Image Processing*, Vol. 3, No. 4, December, pp 277-299.

Horn, B. K. P. (1977) "Understanding Image Intensities," *Artificial Intelligence*, Vol. 8, No. 11, pp 201-231.

Kender, J. R. (1979) "Shape from Texture," *Proc. Image Understanding Workshop*, Science Applications Inc., Arlington VA, April 1979. pp. 134-138.

Lysternik, L. A. (1963) *Convex Figures and Polyhedra*, Dover, N. Y.

Marr, D. and Poggio, T. (1976) "Cooperative Computation of Stereo Disparity," *Science*, Vol. 194, pp 283-287.

Marr, D. and Nishihara, K. H. (1978) "Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes," *Proc. Roy. Soc. B.*, Vol. 200, pp 269-294.

Pearce, P. & Pearce, S. (1978) *Polyhedra Primer*, Van Nostrand Reinhold, New York.

Pogorelov, A. V. (1956) *Differential Geometry*, Noordhoff-Groningen, The Netherlands.

Prazdny, K. (1979) "Egomotion and relative depth map from optical flow," Department of Computer Science, University of Essex, Colchester, England.

Rosenfeld, A. (1978) "Iterative methods in image analysis," *Pattern Recognition*, Vol. 10, pp 181-187.

Smith, D. A. (1979) "Using Enhanced Spherical Images", M.I.T. A.I. Laboratory, May.

Strat, T. (1979) "A numerical method for shape from shading from a single image," S. M. Thesis, M.I.T., Department of Electrical Engineering & Computer Science, January.

Woodham, R. J. (1977) "A Cooperative Algorithm for Determining Surface Orientation from a Single View," *Proc. 5th Int. Joint Conf. on Artif. Intell.*, M.I.T., Cambridge, Mass. pp 635-641.

Woodham, R. J. (1978) "Photometric Stereo: A reflectance map technique for determining surface orientation from image intensity," *Image Understanding Systems and Industrial Applications*, Proc. S.P.I.E. 22nd Annual Technical Symposium, Vol. 155, August.