

SHAPE FROM TEXTURE:
AN AGGREGATION TRANSFORM THAT MAPS
A CLASS OF TEXTURES INTO SURFACE ORIENTATION

John R. Kender
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pa. 15213

A new approach to obtaining shape information from textural information in static monocular images is outlined. Also presented is a new aggregation transform which determines (under certain conditions) vanishing points and lines. A theorem is proved which relates this Hough-like transform to the gradient space, showing that the transform also directly indicates local surface orientations. Additionally, the transform is shown to have many properties that make it an appealing substitute for some other current image transforms. An example is given of its application to a synthetic textured image.

1. INTRODUCTION

One central task of image understanding is the recovery of three-dimensional scene information from the two-dimensional perspective transformation that is the image. The recovery of the missing dimension can be achieved by the use of multiple views: either extensive in time, as in the determination of structure from motion [1], or extensive in space, as in deriving shape from binocular disparity [2, 3]. However, even a single image often contains powerful cues as to object definition and shape; for example, many properties of object surfaces can be derived from an understanding and exploitation of image intensities [4]. In restricting the input to a single image, the task necessarily becomes a heuristic one, given the vast array of scenes that can generate identical images. But such heuristic approaches, especially to the degree that they are model-free, can provide basic theories and algorithms applicable to many image tasks.

This paper begins with a very brief outline of one such relatively model-free approach to deriving shape information from a static monocular view. This method is based on the analysis of texture gradients and the application of principles of projective geometry. Just as an understanding of surface reflectivity enables shape to be derived from shading, so too can shape be derived from the textural properties of a scene.

This research was sponsored in part by an IBM Graduate Fellowship, and in part by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 3597, monitored by the Air Force Avionics Laboratory under contract F33615-78-C-1551.

The remainder of the paper is devoted to the presentation of a new image aggregation transform, in the style of the Hough transform. It is more efficient and "natural" than existing Hough-like transforms, and (under certain conditions) can determine the location of local or global vanishing points or lines. A theorem is proven which shows the intimate relation of such points and lines to the local surface gradient. It further demonstrates the natural correspondence that the transform establishes between perspectivity in the scene and the gradient space [5].

2. SHAPE-RELATED ASPECTS OF TEXTURE

"Texture" is an ill-defined term. However, in some respects, it can be considered an attribute of surfaces, like reflectance or color: its appearance is usually dependent on illumination and view angle. It is well known that blurred textures behave very much like gray scale tones. Similarly, texture gradients behave like intensity gradients.

But there are also important, exploitable differences. Intensities are usually identified one-to-one with picture elements ("pixels"); they have no shape. Under many conditions, the observed intensities of a surface's pixels are fully independent of their distance from the observer; the inverse square law of optics is cancelled by the square law of solid angle. Most discouraging of all, it is difficult to discriminate intensity differences due to illumination variations, reflectance differences, or changes in orientation.

In contrast are textures, especially those made up of identifiable texture elements ("texels"). Texel definition can be rather insensitive to illumination variation. Indeed, if individual texture elements have negligible

extent normal to the surface they define (the texture is "paint"), even shadows may not obliterate the fundamental textural pattern. Further, texture density is directly correlated with surface distance; the inverse square law holds exactly. In addition, given the necessarily larger area needed to define a texel, texels comprise a multi-dimensional family, compared to the one dimension which describes pixels. They are therefore potentially more discriminating at occlusions, and less sensitive to noise. Thus, intensity and texture are somewhat complementary; in fact, they often coexist within the same surface (e.g. the surface of a golf ball).

1.1 Physical Assumptions

At least three interrelated physical phenomena pertain to the derivation of shape information from textural information. These phenomena in turn provide the necessary assumptions needed in recovering three-dimensional surface descriptions. First, there is surface textual homogeneity, which implies that like, nearby regions are really the same region. In the intensity domain this forms the justification for region-growing or -splitting approaches. The analogue for a textured object is based on the assumptions of local texel similarity. Looked at another way, region-growing and -splitting algorithms implicitly recover near-planar surfaces based on the similarity of very small texels: pixels.

Secondly, there is the phenomenon of surface uniqueness and smoothness. It implies the assumptions directly applied in shape-from-shading: the uniqueness of the local surface ("microplane") orientation with respect to an observer, and the continuity of the global surface. Smooth changes in surface orientation usually create textural (and intensity) gradients, but the opposite is not always true; heuristic rules are necessary to reverse the implication.

Lastly, there is the phenomenon of surface position. As position with respect to the observer varies, additional texture gradients arise due to perspective deformation. Such gradients have no direct counterpart in the intensity domain. However, they can be analyzed by using the implied assumption of viewpoint uniqueness. It is with respect to this one viewer position that surfaces have direction or distance.

Each of these three phenomena can be studied in isolation by carefully selecting images that minimize the effect of the other two. In the absence of curvature and perspective, texel similarity has been explored by, among others, [6], who segmented planar objects under orthographic projection. Large, simply-curved surfaces isolate the problem of determining global shape from local clues; processing would be analogous to the intensity-based work of [7]. Large, simply-textured

planar surfaces (e.g. tilted checkerboards) single out the last aspect, perspectivity.

Perspectivity has received little attention. In fact, much research assumes orthographic projection and takes pains to compensate for, rather than utilize, perspective effects. As a consequence, model-free approaches have found it difficult to determine quantitative depth and orientation information: there is no fixed viewer position under orthographic projection. It is these issues that the remainder of this paper addresses, by exploiting the properties of a special, perspective-isolating domain.

3. A NEW AGGREGATION TRANSFORM

Suppose the task of determining shape from texture is simplified to the following very simple subtask. Texels are restricted to be one-dimensional and line-like; they are organized into textures in a regular mesh-like fashion. This is a "structural" texture, somewhat like a grid, except that the line segments can have arbitrarily long gaps, and they do not have to have a fixed spatial frequency. The scene is limited to large planes set at various distances and orientations. The planes are "painted", in the scene, with such a network of parallel lines; in the image, these lines are usually perspectively distorted. Although an abstraction, these conditions fairly well imitate many man-made surfaces: building faces, walls, floors, and so on.

The subtask has several exploitable properties. Surface planarity implies that any local surface orientation (relative to the observer) is the global one as well. Therefore, the determination of local vanishing lines is equivalent to finding global vanishing lines. The latter can be done once, in the large, with consequent improvement in accuracy. Segmentation is eased by the uniformities of the texture component directions. Individual texels are easily identified by an edge detector; no local region growing, etc., is necessary to define them.

The major problem is to aggregate the texels (in this case, edgels) into surfaces, mindful of the vanishing points. Most traditional textural transforms are of limited use here; most were developed with the implicit assumption that the image was the orthographic projection of a frontal two-dimensional scene. Using them, it would be difficult to distinguish intrinsic textural variations from perspective-induced ones.

The new aggregation transform is initially designed, then, to group texels according to the two or more vanishing points they orient towards. (This is a very strong assumption. The occurrence of two or more vanishing points is a special case of the general problem of the vanishing line. Vanishing lines exist independently of texel orientation; they occur with

statistical textures, and even with no textures at all.)

Conceptually, it ought to be sufficient to infinitely extend each line segment, and detect points of mutual intersection. Texels can then be classified into implied surfaces by their respective vanishing points. In effect, this would implement the general image understanding heuristic that converging image lines often arise from parallel lines defining a surface within the scene.

Practically, the problem is a bit more complex. Many times vanishing points are very distant, if not infinite. Further, any solution should be computationally inexpensive, and finding all mutual intersections is not. Lastly, it would be beneficial if the aggregation operator grouped together like-oriented texels in an ensemble that can be easily defined and detected. It will be seen that this, and more, can be efficiently attained.

3.1 The Hough Transform

The vector version of the rho-theta Hough transform is a likely starting point for such a transform (see Figs. 1 and 2). Under it, image points are mapped into sine waves in the parameter space; edge vectors are mapped into points [8]. Colinear edge segments are represented (and recovered) by the accumulation of their transform values at unique points in the Hough space. It is easy to see that parallel line segments are indicated by accumulation points sharing the same theta value. That

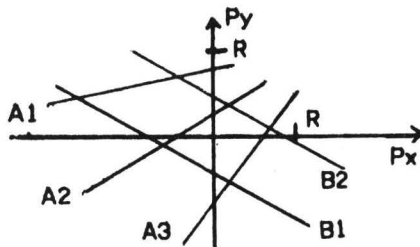


Fig. 1. An image containing parallel lines seen in perspective. Pixels are labelled (P_x, P_y).

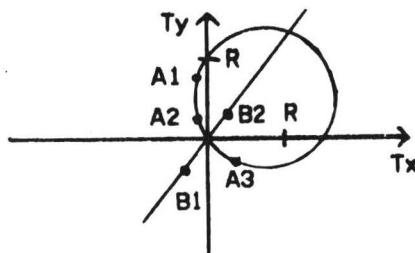


Fig. 3. The new polar form of the aggregating transform of Fig. 1. Both sets of parallel line segments are mapped into accumulation points that lie on circles.

is, parallel lines transform into a line of accumulation points in the parameter space; the line is $\theta = c$. Any mutually converging line segments (possible parallel lines in perspective) transform into accumulation points lying on a sine curve. That sine curve is the transform of their vanishing point.

Unfortunately, sines in the Hough space are difficult to detect. It is likely one would need to apply a similar Hough-like transform to do so: mapping each potential sine point into a curve in a second parameter space, and detecting accumulation points there. Still, this aggregation has the advantage of implicitly representing and aggregating like-oriented edge segments by exactly one curve.

3.2 A First Modification

The following modification of the vectored rho-theta Hough preserves its local grouping property, but represents aggregates in a form that is easier to detect. In addition, it is computationally cheaper, conceptually and visually more forthright, and can be used to replace other Hough-like transforms used for vector grouping (for example, the transform used in the gradient intensity transform method ("GITM") of [9]).

This new approach plots the rho-theta transform space on polar coordinates (see Fig. 3). This has many desirable effects. Points now map into circles which

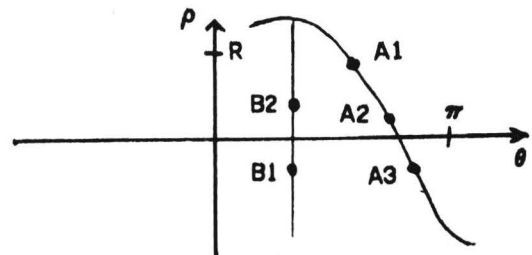


Fig. 2. The rho-theta Hough transform of Fig. 1.

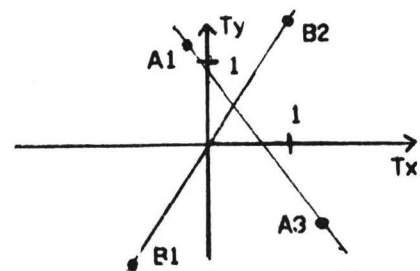


Fig. 4. The efficient, involuting form of the aggregating transform of Fig. 1. Both sets of parallel line segments are now mapped into lines.

pass through the origin. Edge vectors still map into points, but with two interesting new properties. Consider the transformed edge vector (a point) as a vector with respect to the transform origin. Now its direction is parallel to the direction of the edge vector itself. And, its length is equal to the distance (from the image origin) of the line determined by the edge itself. These two properties make the transform easier to view and imagine. As an elegant bonus, no trigonometry is required to calculate this new transform. If the edge vector is the vector $E = (E_x, E_y)$, and if its position in the image is considered the vector $P = (P_x, P_y)$, then the transformed point, represented as the vector $T = (T_x, T_y)$, is defined as:

$$T = ((E \cdot P) / \|E\|^2) E$$

where " $\cdot$ " is the dot product and " $\| \cdot \|$ " is the Euclidean norm.

Further, the transform maps mutually converging line segments into accumulation points that lie on circles passing through the origin. The vanishing point is represented by that point on the circle farthest from the origin. In the case of parallel line segments (with infinite vanishing point), the accumulation points lie on a line through the origin perpendicular to the parallels. Local grouping is preserved; but now the aggregate representation (circles through the origin) is easier to detect, as shown below. Almost all of this discussion, including the claims of computational efficiency, has been shown to apply analogously to the other uses of the vectored rho-theta Hough. As an example, in the GITM method, what once mapped into secant curves in the transform spaces now maps into straight lines.

3.3 The Second Form

Detecting the circular arcs can be done efficiently in the following manner. Consider a second transform that involutes the result of the first polar transform (see Fig. 4). That is, all transformed edge vectors, which are represented in the first polar space as (ρ, θ) are mapped into $(K/\rho, \theta)$, for some K . This transform has even more desirable effects. Infinite vanishing points are now mapped into the origin ($K/\rho = 0$). Lines through the first polar origin (along which lie the transforms of parallel lines) are unchanged. Most importantly, all circles passing through the first polar origin (along which lie the transforms of converging lines) are mapped into straight lines. Thus, in either case, a set of scenic parallel lines, perspectively deformed in the image or not, is mapped into accumulation points that lie along a straight line. The distance of this straight line from the newest origin is inversely proportional to the distance to the vanishing point of the line set. A normal to the straight line is parallel to the vanishing point direction. Another bonus: If this transformation is composed with the first, the combined operation is even cheaper than that of the

first alone. The transform vector $T = (T_x, T_y)$ is then defined in terms of the edge vector $E = (E_x, E_y)$, the edge position $P = (P_x, P_y)$, and the scaling constant K as:

$$T = (K / (E \cdot P)) E$$

where " $\cdot$ " is again the dot product.

Lines are easy to detect, especially when compared to the original suggestion of searching for sines.

4. THE VANISHING LINE-GRADIENT SPACE RELATION

These lines, however, have yet another remarkable property. Under very natural conditions, they intersect in the transform space at (p, q) , the value of the surface gradient vector. That is, the transform simultaneously achieves three goals. It aggregates edgels into straight lines (represented by points in the transform space). It aggregates straight lines into sets of (perspectively distorted) parallel lines (represented by lines in the transform space). And it aggregates parallel line sets into planes (represented by the intersection of lines in the transform space at the unique point corresponding to the plane's orientation).

The following theorem formalizes the conditions under which the above claims are valid.

Theorem: Suppose an image contains the perspective projection of a planar surface defined by two or more coplanar sets of parallel line segments. Let the origin of the transform space correspond to the coordinates of the focal point in the image (that is, where the camera is "aimed"). Let R equal the focal distance. Then $T = R E / (E \cdot P)$ (as defined above) transforms edges so that the intersection of lines of accumulation points in the transform space is at (p, q) , the gradient vector of the surface.

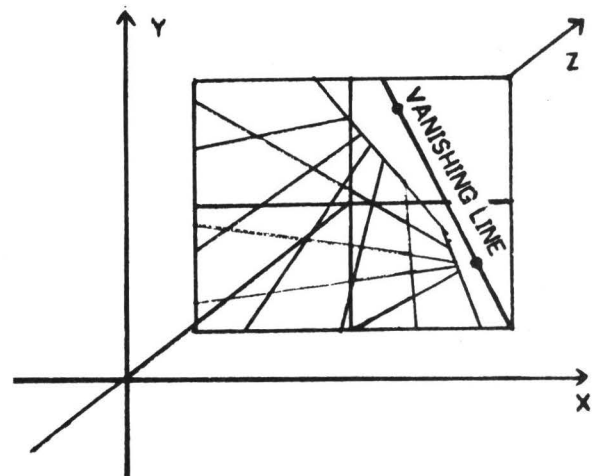


Fig. 5. The imaging model, showing a mesh-like textured plane with its vanishing points and vanishing line.

Proof: By the duality of the transform, the transform of the vanishing line (a point) lies on the intersection of the transforms of the vanishing points (lines). It is this intersection which occurs in the transform space. It is now only necessary to show the correspondence, under the transform, of the vanishing lines and surface gradients.

The imaging process can be modeled as in Fig. 5. The center of projection corresponds to the origin of the (x, y, z) coordinate system. The customarily upside-down focal surface (the "film") has been rotated through the center of projection so that it is right-side-up; it is at a focal distance of R . The projection of the focal point passes through the focal surface at $(0, 0, R)$. On the "film" is the image of a mesh-like planar surface. Also depicted in the figure, for illustrative purposes only, is the plane's vanishing line and two vanishing points. It is important to note that such lines and points are usually only theoretic constructs. They rarely appear in the image itself.

Let the surface equation be given by $px+qy+c = z$, following Huffman. Its gradient vector is thus $(\partial z/\partial x, \partial z/\partial y) = (p, q)$. Points on the surface map onto the focal surface under central projection according to the relation $(x, y, z) \rightarrow (Rx/z, Ry/z, R) = (Rx/(px+qy+c), Ry/(px+qy+c), R)$. The equation of the vanishing line on the focal surface can be found by passing a line through two different vanishing points. However, a vanishing point is defined as the limit of the projection of an infinite line lying in the surface. Let $(x, 0, R)$ and $(0, y, R)$ be two such lines; the limits of their projections, as x and y respectively tend to infinity, are $(R/p, 0, R)$ and $(0, R/q, R)$.

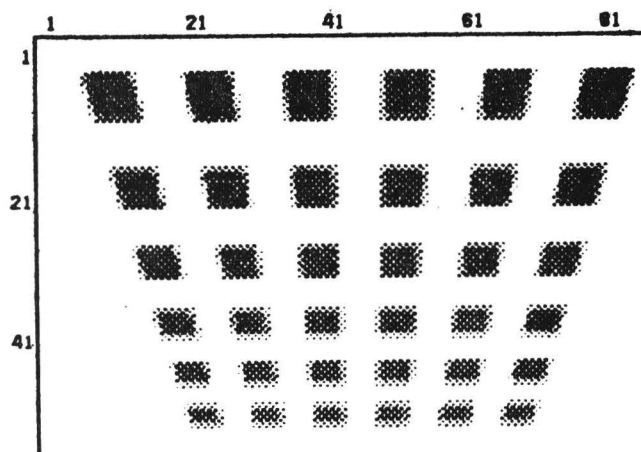


Fig. 6. A synthetic textured scene. The focal point is at $(45, 45)$, which becomes the origin, $(0, 0)$, of the inherited imaging coordinate system. The focal distance is 90, and the surface gradient vector is $(p, q) = (0, -1)$.

Now, impose a coordinate system on the focal surface in the natural way: "film" points (x, y) are the points (x, y, R) in three-space. The vanishing line in the two-dimensional coordinate system then has the equation $y = -(p/q)x + R/q$. Its normals (its "edge vectors") have everywhere a slope of (q/p) ; that is, they are in the direction of the vector (p, q) . Therefore, any point on the vanishing line is transformed into $RE / (E \cdot P) = R(p, q) / ((p, q) \cdot (x, -(p/q)x + R/q)) = R(p, q) / (px - px + R) = (p, q)$. This last is the surface gradient.

Thus, the theorem is proved: vanishing lines map into gradient vectors. Or, put another way, the transform creates a natural correspondence between the Hough transform and the gradient space.

5. EXAMPLES AND SUMMARY

The complete transform process is summarized and illustrated by the following figures, using a synthetic textured image (Fig. 6: a simulated building face). The focal point is at the center of the image; the focal distance is equal to the width of the image. The surface was created with gradient $(p, q) = (0, -1)$.

Edge vectors (Fig. 7) are first mapped into accumulation points in a polar rho-theta Hough space. Mutually converging lines are thereby mapped into accumulation points that lie on circular arcs through the origin (Fig. 8). Involuting the transform space maps these arcs into straight lines (Fig. 9). This result was actually derived directly from the edge image by means of the more efficient second form of the transform. The lines in the involuted polar space are seen to intersect at $(p, q) =$

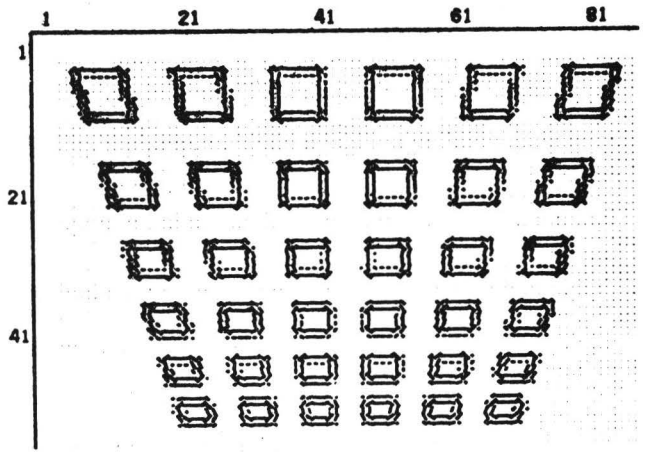


Fig. 7. Local edge elements of Fig. 6.

(0, -1), which is both the transform of the surface's vanishing line, and the value of the surface's gradient.

Determining shape from texture has many facets; the transform reported here is only one small one. Other approaches based on different classes of texels, together with an overall computational paradigm showing their common relations, will be presented in [10].

ACKNOWLEDGEMENTS

The author gratefully acknowledges many stimulating discussions with Takeo Kanade.

REFERENCES

- [1] S. Ullman, "The Interpretation of Visual Motion," Ph.D. Thesis, Departments of Electrical Engineering and Computer Science, M.I.T., 1977.
- [2] D. Marr, and T. Poggio, "Cooperative Computation of Stereo Disparity," Science, Vol. 194, October 15, 1976.
- [3] D. B. Gennery, "A Stereo Vision System for an Autonomous Vehicle," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, M.I.T., 1977.
- [4] B. K. P. Horn, "Understanding Image Intensities," Artificial Intelligence, Vol. 8, 1977.
- [5] D. A. Huffman, "Impossible Objects as Nonsense Sentences," Machine Intelligence, Vol. 6, Edinburgh University Press, 1971.
- [6] F. Tomita, M. Yachida, and S. Tsuji, "Detection of Homogeneous Regions by Structural Analysis," Proceedings of the Third International Joint Conference on Artificial Intelligence, Stanford, 1973.
- [7] R. J. Woodham, "A Cooperative Algorithm for Determining Surface Orientation from a Single View," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, M.I.T., 1977.
- [8] S. A. Dudani, and A. L. Luk, "Locating Straight-Line Edge Segments on Outdoor Scenes," Proceedings of the IEEE Computer Society Conference on Pattern Recognition and Image Processing, Rensselaer Polytechnic Institute, 1977.
- [9] C. L. Fennema, and W. B. Thompson, "Velocity Determination in Scenes Containing Several Moving Objects," Technical Report, Central Research Laboratory, Minnesota Mining and Manufacturing Company, St. Paul, 1977.
- [10] J. R. Kender, "Shape from Texture," Forthcoming Ph.D. Thesis, Department of Computer Science, Carnegie-Mellon University, 1979.

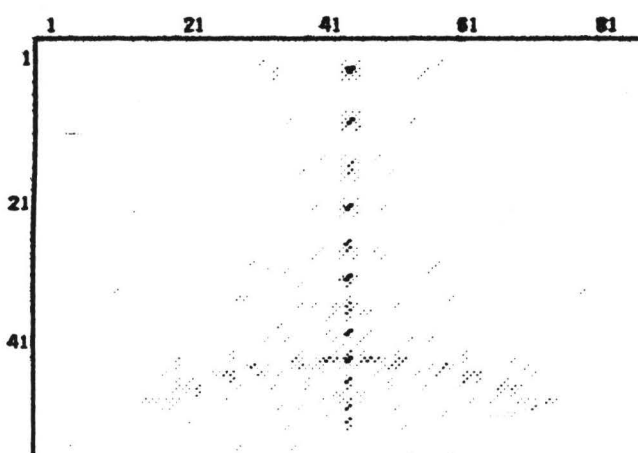


Fig. 8. The new polar form of the aggregating transform of Fig. 6. Both sets of parallel line segments are mapped into accumulation points that lie on circles through the inherited origin, which is at (45, 45).

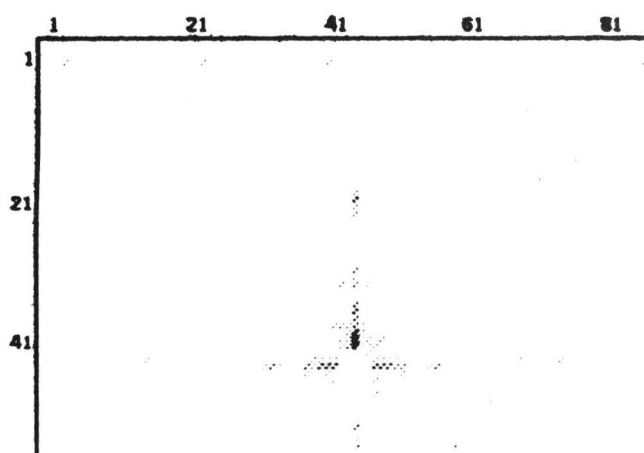


Fig. 9. The efficient, involuting form of the aggregating transform of Fig. 6. Both sets of parallel line segments are now mapped into lines. They intersect at (45, 46), which in the inherited imaging coordinate system is (0, -1). This is the value of the surface gradient vector.