IAPR Workshop on CV - Special Hardware and Industrial Applications OCT.12-14, 1988, Tokyo

A STRATEGY FOR PROCESSING 3D RANGE IMAGES

T. Kasvand,
Department of Computer Science,
Concordia University,

1455 DeMaisonneuve Blvd W.,
Montreal, Que., H3G 1M8, Canada

ABSTRACT

A method of analyzing 3D range images Z(x,y) is described and illustrated. Prior knowledge of object models and scene content is not used. The strategy uses a mixture of analytic and image processing techniques. A pixel or surface element ("surfel") in Z(x,y) has eight degrees of freedom, of which Z, x, and y are given, and the remaining five are computed. The surfels are classified and the resultant facets are analytically "relaxed" and labelled, giving an image of facet labels Lf(x,y). Lf(x,y) is processed for edges and corners, and the facets are grouped ("conceptual generalizations"). Lists of facet surface parameters, edges and edge parameters, corners, facet edge shapes, and view-independent primitives are obtained, including various adjacency graphs. Ample information is made available for object learning, knowledge base construction, and object recognition. Considerable computing is required and the method can be practical only in a multicomputer environment or on special hardware.

INTRODUCTION

The detection of depth from pairs and sequences of gray level images is a complex problem. This complexity is very elegantly sidestepped by the 3D (laser) range finding scanner which provides distance readings directly, and also delivers gray level data and, if desired, colour information [1]. There are several other ways of detecting the range from an observer to the surfels in a scene [2]. The range finder is an "active" scanner, i.e., it provides its own light source (a laser) to illuminate the scene. Due to its "active" nature, such a scanner is not suitable when the observer wants to remain concealed, and it is dangerous to eyes. The light beam has to be scattered adequately by the nearest surfels in the scene to produce detectible return signals. Hence, the surfaces of transparent objects, reflective surfaces, and also "furry" surfaces create problems since the return signal may be absent or appear to "come" from the wrong "place". Where the range scanner can be used, it is a very useful device for computer vision. The analysis of range images is rather straight forward since range data are physically meaningful and unambiguous, namely, the distances from the camera to points on the surfaces of objects in the scene.

Differential geometry may be considered to be the theoretical foundation for the analysis of range data [3]. However, even though the required processing steps are theoretically well defined, there is a difference between theory and practice. Theoretically, the range image is a function of the form G(u1,u2) of the two surface coordinates u1 and u2. The function G(u1,u2) is assumed to be "sufficiently differentiable" to suit the theory. In practice, the range image is a spatially quantized function of the form Z(x,y). Z(x,y) is a matrix of tabulated orthogonal distances Z from the plane of the camera to some surfel (x,y) in image coordinates. Z(x,y) is neither noise free nor is it "sufficiently differentiable" since it contains discontinuities in Z values at unknown locations in the (x,y) plane. These we call the "edges" and "corners" of objects, while the "sufficiently differentiable" regions are the smooth facets in the scene. Only after considerable processing can the smooth facets be expressed as functions of the form Z = f(x,y), from which point onwards differential geometry and analytic techniques become directly applicable. Due to limited space, it is impossible to describe and list all the efforts and authors concerning range images. Summaries may be found in, for example, [4,5,6,7].

A STRATEGY

The usual approaches to image analysis are chosen by need, convenience, and convention, since there is no well formulated theory. Furthermore, it is nearly always assumed that there are only a few objects in a scene and that the objects can be found by some model matching technique, given enough constraints. Of course, since object recognition requires some prior "knowledge" of the object, the final stage of any recognition scheme consists of comparing the "knowledge" with the information extracted from the scene. However, a pure "top down" process results in a combinatorial explosion, and a pure "bottom up" procedure generates a profusion of "features" the combinations of which also "explode". The Z(x,y) image is no exception. However, the Z(x,y) information has only one unique interpretation, namely, Z is the orthogonal distance from some reference (x,y) plane to the nearest surfel in the scene. Z is independent of surface properties and colour. The author has approached this problem as follows, see Figure 1:

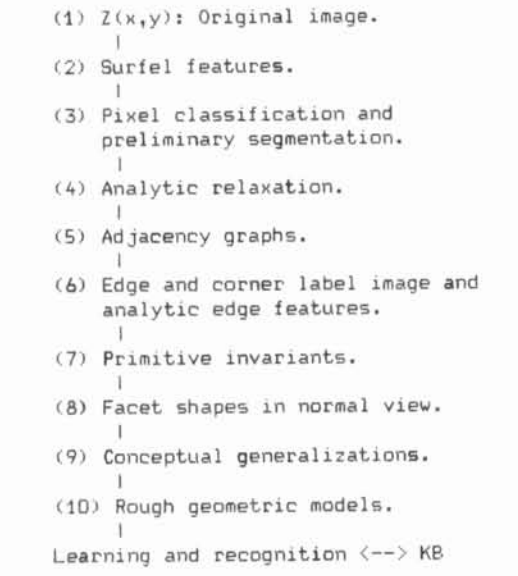


Figure 1: A brief sketch of the processing strategy. KB is the knowledge base.

1) Premises: For generality, the scene content is assumed unknown and no constraints are placed on the number, size, position, orientation, shape, and overlap of the objects. The methods must not require prior knowledge of object models. This dictates a "bottom up" or "data driven" approach but the "explosion" of primitives is to be avoided. Adequate spatial resolution and the existence of smooth and opaque surfaces in the scene are assumed.

2) Surfel features: Given the Z(x,y) image, compute the remaining five degrees of freedom (DOF) at each surfel. The basic parameters for a surfel are its position (x,y,Z) expressed as Z(x,y), its unit surface normal vector N(x,y,Z), and the maximum surface curvature k1(x,y,Z) and the minimum surface curvature k2(x,y,Z). The surface curvatures k1 and k2 are scalars with an arbitrarily defined sign. The maximum (k1) and minimum (k2) are orthogonal, and directed as indicated by the corresponding unit vectors U1(x,y,Z) and U2(x,y,Z). The vectors N(x,y,Z), U1(x,y,Z), and U2(x,y,Z) form an orthogonal triplet of unit vectors. Thus, a surfel has eight DOFs, three for the position, three for orientation in space, and two from the k1 and k2 values. When these values are available, the surfels can be considered recognized.

The numerical computations are not straight forward due to noise and discontinuities in Z(x,y), which are not considered in differential geometry. Filtering may be applied to reduce noise, but the discontinuities should not be "disturbed", since they represent edges and corners. There are basically two approaches, namely, local area fitting to obtain a local analytic approximation [8,9], or direct computations [10]. Both methods have obvious drawbacks, and the resultant "surfel features" are increasingly unreliable as a function of the amount of processing and differencing.

3) Classification: Classify the surfels by selected surfel features to obtain "homogeneous" regions (facets) in the xy-plane. However, in the absence of prior knowledge, there is no unique set or sequence of sets of surfel features for classification. An hierarchy of classifications is suggested in [11] and a single step in [12]. In any case, the "raw" facets found will depend on the features or feature sequences chosen. After classification, the "raw" facets can be considered recognized according to their surface characteristics.

4) Analytic "relaxation". If the maximum and minimum curvatures k1 and k2 are chosen in (3), the decision space H(k1,k2) segments the image at most into second order facets. Consequently, a second order analytic function is suitable for approximating the "raw" facets. The function $z(x,y) = a + bx + cy + dx2 + exy + fy2$ was chosen and the parameters found by L1 approximation [13]. The fitting is iterated to find all the surfels that can be considered to belong to a given facet. Acceptable analytic facets are found after two iterations, resulting in an image Lf(x,y) of facet labels and a list F(..) of analytic facet parameters.

5) Adjacency graphs: Given the labelled image Lf(x,y), the images of k1(x,y), k2(x,y), etc., and the analytic parameters in F(..), it is a simple matter to construct various adjacency graphs indicating which facets meet and what happens at facet contacts.

6) Edges and corners: The edges and corners in the image are found at contacts between different labels in Lf(x,y). After some processing the edges between the facets are labelled and the analytic equation for each edge is obtainable, if desired. The corners are also labelled. It should be noted that there are "true" edges and corners, and also "other types" caused by occlusion and analytic approximation of facets. The nature of the edge can be detected given the data so far, but this has not yet been confirmed experimentally.

7) Primitive invariants: View- and occlusion-independent variables that are now already available or easily computable are planar facet normal directions, surface curvatures, curvature directions, normal vector differences at edges and corners, relative sizes of facets if "fully visible", etc.

8) Normal views: The edge-shape of a facet when seen in the normal direction is easily obtained. However, the facet may be partially occluded, see (6).

9) Conceptual generalizations: The computations up to this point are lengthy but straight forward due to the uniqueness of Z(x,y). Two aspects should be noted, namely, occlusions which are "natural", and the analytic approximation which is "not natural". Occlusions can split a "natural facet" into several different facets in Lf(x,y), each of which has its own individual set of analytic parameters in F(..). Surfel classification and analytic approximation splits even a fully visible multiply curved "natural

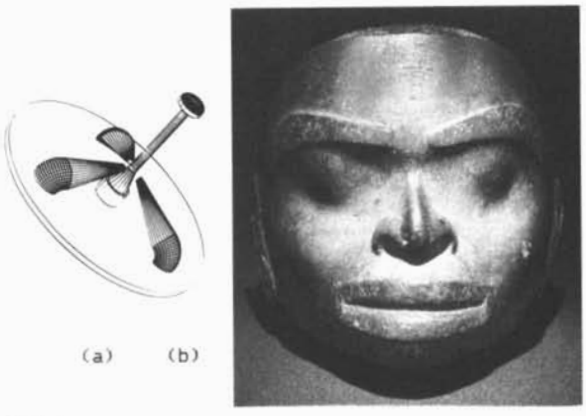
facet" into several facets, each of which has its own set of parameters in F(..). This "not natural" segmentation is caused by the analytic approach. Numerous rules may be postulated for assembling the facets in F(..) into larger and possibly "more natural" facets, see Experimental results.

10) Rough geometric and other models: The only way to satisfy the premises in (1), in the author's belief, is to have "first level" models for object recognition which are constructed from "primitives" which can be extracted from the scene without any prior knowledge of scene content. As seen from the analysis above and the experimental results, there are many such "first level primitives" (L1P's). A knowledge base constructed from L1P's need not "explode", and models (KBM's) which do not contain at least some of the L1P's cannot be candidates for further study. These aspects are under investigation.

EXPERIMENTAL RESULTS

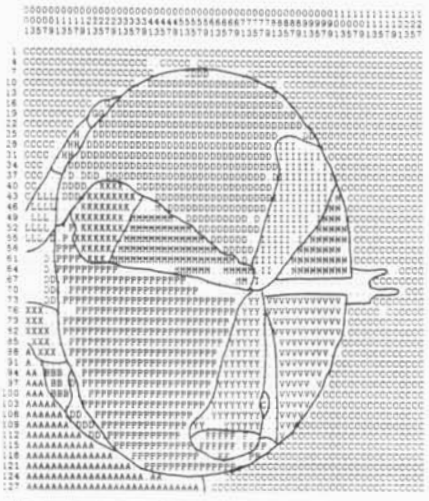
Two range images called "Grapple" and "Mask001" [14] were selected, see Figure 2. The original size of the Z(x,y) images is 256 by 256 surfels. The spatial resolution in x, y, and z is the same. The studies were carried out on reduced 128 x 128 images by selecting every second pixel on every other row.

Both images were processed for surfel features (step 2 above), the pixels were classified (3), the facets were analytically relaxed (4), and some adjacency graphs were obtained (5). The resultant facets are shown in Figure 3 and a portion of an adjacency graph is in Figure 4. Due to the rather low spatial resolution, the "probe" or the "center post" in Grapple may not resolve properly. Most of the information about the images is now available in "conventional data structures", such as lists of analytic parameters (F(..)), adjacency lists, and raster images where each pixel carries its facet label (Lf(x,y)). Edge detection (6) based on Lf(x,y) is very simple and normal view creation (8) is essentially a matter of coordinate transformation. The study was continued with conceptual generalizations (9) based on some invariants or "semi-invariants" (7).



(a) (b)

Figure 2. Displays of the scenes for Grapple and Mask001. (a) A CAD model rendering of the Grapple. (b) A photo of Mask001.



(a)

(b)

Figure 3. (a) Decimated alphabetic print of facet labelled image Lf(x,y) for Grapple from analytic relaxation. The facet labels are numbered as 2, 3, 4, ..., and shown by letters (0=., 1=*, 2=C 3=D, ..., 25=Z, 26=[, 27=A, ..) For clarity, the boundaries between facets have been outlined. (b) Edge enhanced analytically reconstructed Z(x,y) based on Lf(x,y) and FF(..) for Mask001.

		B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
B	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	2	0999	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	3	0	0421	14	0	0	0	0	0	9	3	7	0	0	0	2	0	0	0	0	14
E	4	0	0	19	77	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0
F	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	6	0	2	0	0	0999	0	0	0	63	0	0	0	0	71	0	0	0	0	0	0
H	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	9	0	0	0	26	0	8	0	0389	0	8	4	0	0	0	0	0	0	0	3	0
K	10	0	0	48	0	0	0	0	0	0411	0	0	0	0	0	0	0	0	0	0	0
L	11	0	0	72	0	0	0	0	0	64	0389	5	0	0	63	56	0	0	0	0	0
M	12	0	0	0	0	0	0	0	0	26	0	21152	0	0	0	2	0	20	7	0	
N	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O	14	0	0	0	0	0	75	0	0	0	0	0	0	0143	0	0	0	0	0	0	0
P	15	0	0	64	0	0	0	0	0	0	2	0	0	0179	0	0	0	0	0	25	
Q	16	0	0	0	0	0	0	0	0	0	0	42	30	0	0	20170	0	21	0	22	
R	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0324	13	0	14	
S	18	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	33	39228	0	5	
T	19	0	0	0	0	0	0	0	0	26	0	0	31	0	0	0	0	0	0205	0	
U	20	0	0	11	0	0	0	0	0	0	0	0	0	0	0	34	2	41	24	0219	

Figure 4: The top left corner of the adjacency graph for Mask001 showing some facet contact semi-invariants. The peripheral rows and columns are facet labels and alphabetic labels. The diagonal elements give the number of pixels per facet. Below diagonal entries indicate the number of pixels making up the contact between the two facets. Above diagonal entries give average sums of 100*(1-cos(Fi,Fj)), i.e., cosinus of the difference between surface normals at contact. Numerical values >1000 are set to 999.

The Grapple image presents a rather simple problem. Given any two flat facets Fi and Fj (for which k1 and k2 are approximately zero), the generalization consists of the following, expressed as a "logical IF":

IF ((facet Fi and Fj are flat) .AND. (Fi and Fj are close, < D pixels apart) .AND. (Fi and Fj normals are parallel, within T degrees) .AND. ((Abs(Z(Fi)-Z(Fj)) at contact < Zd)) THEN join the facets.

The result (Lfcg(x,y)) is shown in Figure 5. Note that the flat background has become one facet, and the "face plate" has become another flat facet, call it Fp. Three parameters have been used, namely, a measure of "closeness" (D), angular disparity between normals (T), and how well the facets "fit together" (Zd) at the point where they would join if there were no obscuring objects in the view.

The surface normal Nfp of Fp and its center of gravity CGfp can serve as a semi-invariant reference coordinate system (x'y'z'). The "arms" on the grapple consist of a "knuckle" and a conical "bone" each. Unless the grapple is very highly tilted away from the direction of view, one or two "knuckles" and at least two "bones" remain visible and have been identified as "spherical" (k1 and k2 are nonzero) and "cylindrical" or "conical" (k1 not zero, k2 approximately zero) regions. Accurate information is available from the analytic approximation and "noisy information" may be obtained directly from the k1(x,y) and k2(x,y) images masked by Lfcg(x,y). The centers of gravity for all the facets are available from image data but, of course, they are somewhat dependent on the number of pixels seen on each facet, hence the "semi-invariance". A rotation angle for (x'y'z') may be defined with respect to the best visible "bone and knuckle" combination. Of course, Fp has to be recognized (as a circular disk).

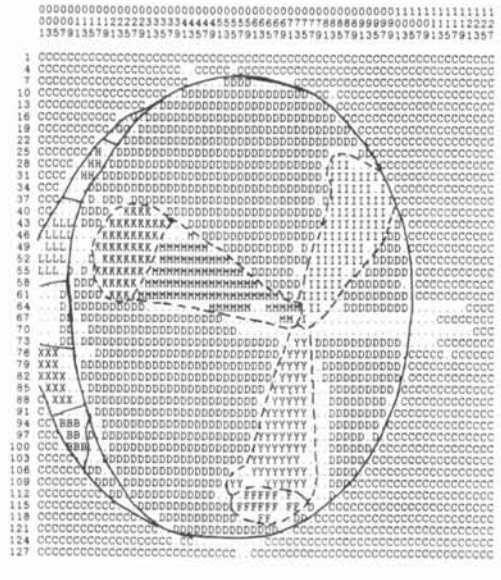


Figure 5. Decimated alphabetic print of the "conceptually generalized" facet labelled image Lfcg(x,y) for Grapple.

The "semi-invariant" recognition features in the grapple image with respect to CGfp are, among others, the spatial angles between any two "knuckles", between any two "bones", between Nfp and a "knuckle" and "bone" triplet, and that Nfp and adjacent "knuckle and bone" are approximately in the same plane. The vectors are defined with respect to the CG's. Such relationships constitute "rough geometric models" (step 10).

The Mask001 image represents a much more interesting challenge and it also point out certain weaknesses in the method. A careful study of the mask (Figure 2b) and the segmentation (Figure 3b) reveals that the computer is "most faithfully doing the best it can". Even though we can assign a meaning to most of the facets, this is insufficient for machine recognition. The facets have been forced to be of second order, and a very meticulous second order segmentation has been obtained, but there are too many such facets. To reduce the number of facets, numerous "conceptual generalizations" are possible, but to determine which facet combinations are "meaningful" in human terms and which are not, is both premature and creates the basic paradox in image segmentation. Once a set of facets have been joined, the resultant analytic approximation should correspond to the complexity of the facet. The generalizations experimented with are given below, where FO is the "absorbing" facet and Fk the facet "to be absorbed" by FO.

(a) Larger facets can absorb smaller facets (FO > Fk) if a combination of the following conditions is satisfied:
(b) The amount of contact between facets has to be adequate (>Lc), for example, expressed as P2/A, where P is the contact length between FO and Fk, and A is the area of Fk, see Figure 4.
(c) The average analytically computed absolute Z-difference at contact between FO and Fk should be less than Zd. This may be corrected for surface normal view.
(c) The average 1-cos(NfO,Nfk) at contact is less than a limit Cd, see Figure 4.
(d) The signs of k1(FO) and k1(F1), and k2(FO) and k2(Fk) are the same.
(e) The average "flatness measure" |k1|+|k2| at contact is less than a limit Kf.

It can be shown that condition (e) is not very reliable, leaving the parameters Lc, Zd, Cd, and a choice for (d), to cluster the facets. A few experimental results are in Figure 6. With adequate "fine tuning" of H(k1,k2) classification parameters and Lc, Zd, Cd, etc., rather interesting segmentations of Mask001 may be produced, but this violates the premises (1) that the scene is unknown and we are introducing our own understanding of how the scene should be segmented. The only critical requirement at this stage is consistency in segmentation for scenes of the same kind such that the knowledge base can be addressed without creating a combinatorial explosion. This argues for an interplay (feedback) between, at least, the conceptual generalizations and the knowledge base, but for the moment these are only conjectures.

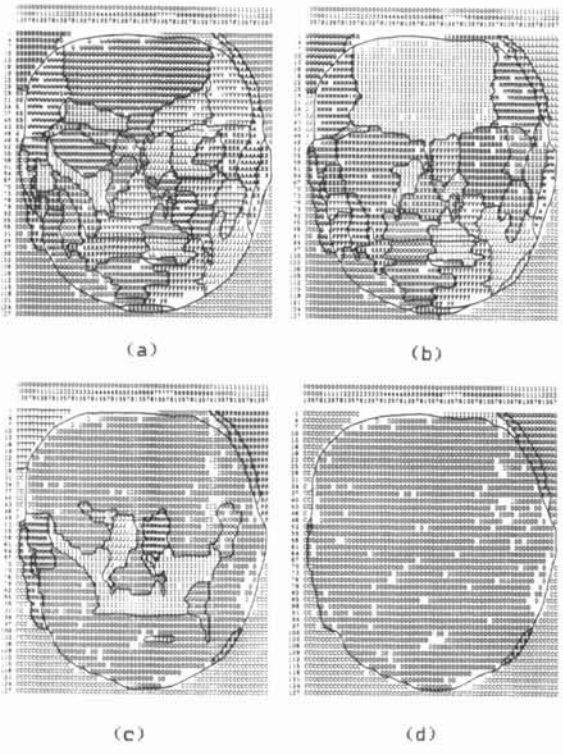


Figure 6. Decimated alphabetic prints of the "conceptually generalized" facet labelled image Lfcg(x,y) for Mask001. (a) Lc=10, Zd=25 since Z-differences are times 100, Cd=10, and condition "d" is on. (b) Lc=10, Zd=50, Cd=10, and "d" off. (c) Lc=0, Zd=300, Cd=50, and "d" on. (d) Lc=0, Zd=300, Cd=50, and "d" off.

COMMENTS

The Mask001 image illustrates that in the clustering of facets the combinatorial explosion can be avoided by predefined "rules". It is usually expected that the resultant "generalizations" have to correspond to "humanly meaningful" facets, but this expectation is premature. The only requirement at this stage of processing, based on the facets found, is to locate the most likely object model candidates in the knowledge base. In human terms, the entire processing described so far only constitutes "the first glance" (of about 0.1 seconds) at the scene!

CONCLUSIONS

A strategy has been described and demonstrated, indicating that the processing can be carried out in the absence of prior knowledge of the Z(x,y) scene. Invariant and semi-invariant descriptors are obtained which can be used to construct a knowledge base ("learn") as well as to access the knowledge base for recognition. However, only with a multicomputer configuration or with special hardware is it feasible to carry out the required computations fast enough to make this approach practical and to advance research on knowledge base structures for machine vision based on 3D range images.

ACKNOWLEDGEMENTS

The author wishes to express his sincere thanks to all colleagues and to the National Science and Engineering Research Council, the Division of Electrical Engineering of the National Research Council, Concordia University, and University of Ottawa.

REFERENCES

[1] M. Rioux, "Laser range finder based on synchronized scanners", Applied Optics, Vol. 23, No. 21, pp. 3837-3844, 1984.

[2] R.A. Jarvis, "A perspective on range finding techniques for computer vision", IEEE Trans., Vol. PAMI-5, No. 3, Mar. 1983, pp. 122-139.

[3] E. Kreyszig, "Introduction to Differential Geometry and Riemannian Geometry", University of Toronto Press, 1968.

[4] O.D. Faugeras, "Fundamentals in Computer Vision", Cambridge Univ. Press, Cambridge, England, 1983.

[5] T. Kanade, "Three-dimensional Machine Vision", Kluwer Acad. Publishers, 1987.

[6] A. Rosenfeld, Editor, "Techniques for 3-D Machine Perception", North Holland, 1986.

[7] Y. Shirai, "Three-Dimensional Computer Vision", Springer-Verlag, 1987.

[8] R.M. Haralick, "Ridges and valleys on digital images", Comp. Vision, Graphics, and Image Processing 22, 1983, pp. 28-38.

[9] P. Besl and R. Jain, Invariant surface characteristics for 3D object recognition in range images, CVGIP, Vol. 33, 1986, pp. 33-80.

[10] T. Kasvand, "Surface curvatures in 3D range images", Proc. 8'th ICPR, Paris, 1986, pp. 842-845.

[11] T. Kasvand, "The k1k2 space in range image analysis", Proc. 9'th ICPR, Rome, 1988.

[12] P. Boulanger, "Label relaxation technique applied to the stable estimation of a topographic primal sketch", Proc. Vision Interface '88, Edmonton, Alberta, June 6-10, 1988.

[13] N.N. Abdelmalek, "L1 solution of overdetermined systems of linear equations", ACM Trans. Math. Software 6, 1980, pp. 220-227.

[14] M. Rioux and L. Cournoyer, "The NRCC three-dimensional image data files", National Research Council of Canada, CNRC 29077, June 1988.