Extraction of Volumetric Structures In an Illuminance Image

Marielle Mokhtari[†‡] and Robert Bergevin[‡]

[†] Distributed Synthetic Environment Group
Defence R&D Canada - Valcartier
Val-Belair, Qc, Canada, G3J 1X5
marielle.mokhtari@drdc-rddc.gc.ca

[‡] Computer Vision and Systems Laboratory
Dept of Electrical and Computer Engineering
Laval Univ., Ste-Foy, Qc, Canada, G1K 7P4
[marielle,bergevin]@gel.ulaval.ca

ABSTRACT

An original method is proposed to extract the most significant volumetric structures in an illuminance image. The method proceeds in three levels of organization managed by generic grouping principles: (i) from the illuminance image to a more compact representation of its contents by generic structural information extraction leading to a basic contour primitive map; (ii) grouping of the basic primitives in order to form intermediate primitives, the contour junctions; (iii) grouping of these junctions in order to build the high-level contour primitives, the generic volumetric structures. Experimental results for various images of cluttered scenes show an ability to properly extract the structures of volumetric objects or parts with planar and curved surfaces.

Keywords

Illuminance Image – Multi-Level Grouping - Contour Primitives (straight-line segments and circular arcs) - Contour Junctions - Volumetric Structures - Surfaces.

1. INTRODUCTION

In the context of a generic *3D* object detection and description task, high-level structures need to be extracted from basic contour primitives in an illuminance image of a cluttered scene. The scenes of interest are composed of rigid, opaque, and partly occluding man-made objects. Low-level processing of the image of a cluttered scene is to give rise to illuminance contours that are to be processed further to obtain the sought-for structural description.

Two main difficulties arise. Firstly, cluttered scenes offer a structural complexity that has to be recovered on the basis of the contours extracted. Such contours are extracted at the pixel level with no specific knowledge about the scene and the objects present. Their structure is not fully representative of the underlying structure of the scene. For instance, contours extracted at the pixel level may very well go across the borders of different nearby objects, parts and surfaces. Secondly, contours are obtained from real images and are thus very likely to suffer from image and low-level processing noise. Some contours may to be missing. Others may to be incomplete especially at surface junctions. Still others may to be spurious, resulting from various photometric effects such as shadows, highlights, and surface markings and textures. The challenge is to recover the scene structure (detect or single-out each object or part) and each object structure (single-out each of its surfaces and their structure as a description) despite these real-world difficulties.

Very few generic extraction methods for *3D* objects in an illuminance image of a cluttered scene are proposed in the literature. In fact, the description methods proposed in the literature either use images of other modalities, e.g. range data [Levine92], synthetic line drawings [Bergevin93] and feature maps [Hummel92], extract structures that are too specific for our goal [Huttenlocher92] [Lu92] [Wong92] [Yla-Jaaski96], or do not explicitly consider the volumetric nature of the objects [Denasi94] [Etemadi93] [Fuchs95] [Jacot-Descombes97] [Lu92]. The proposed methods are unlikely to properly detect each actual object in isolation in the various cluttered scenes of interest. One of the best methods proposed so far, with respect to our goal, was developed by Zerroug *et al.* [Zerroug94]. It is a combination of two methods specific to two classes of generalized cylinders: straight axes and circular sections. A major difference with our proposed method is their extraction of intermediate structures (symmetry axes) directly from local point features with no integration into generic constant curvature contour primitives. Very promising results were provided but only for a small number of cluttered scenes with close-up views of complex objects in partial occlusion.

This paper presents an original method for the extraction of generic volumetric structures in a single illuminance image of a cluttered scene. This method

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WSCG SHORT Communication papers proceedings
WSCG'2004, February 2-6, 2004, Plzen, Czech Republic.
Copyright UNION Agency – Science Press

is at the heart of the MAGNO system (Multi-level Access to Generic Notable Objects). MAGNO exploits generic knowledge available at each of its processing levels. For instance, MAGNO exploits generic knowledge about junctions of *3D* objects to both detect objects and organize constant-curvature (both straight and curved) basic contour primitives into a generic description of each visible object (or part), as inspired by human perception studies [Biederman85].

The proposed method concentrates on geometric features at three levels, extracted as a three-phase process. The first phase consists in the extraction of generic structural information from a single *2D* illuminance image of a cluttered scene. The result of this phase is a primitive map made up of constant curvature segments. These segments are referred to as *basic primitives*. The extracted basic primitives are structured according to the image contours. They have a small number of defining parameters that makes them an adequate basis for the second phase of the method. At the end of this first phase, the description does not yet reveal the structure of each object. The second phase consists of grouping basic primitives according to various principles of perceptual organization [Lowe85]. The obtained groups are referred to as *intermediate primitives* or *junctions*. Junctions provide cues to the structure of the scene and its volumetric objects. In that sense, they help to reintroduce the missing aspect of a single illuminance image that is, the depth or third dimension. In the third and last phase, junctions are themselves grouped, on the basis of their forming primitives, to produce the *high-level primitives*. These are the generic volumetric structures, each corresponding to a single *3D* object or part present in the image. High-level primitives correspond to arrangements of basic primitives structured according to their junctions.

The paper presents details of each of the three grouping phases of our original high-level structure extraction method, together with a number of results from its implementation. Next section summarizes the extraction of the basic primitives from an illuminance image. Then, contour junctions extraction is addressed. This is followed by a more thorough description of the generic high-level structures extraction phase. In order to illustrate the overall behaviour of our proposed method, various results obtained using a fully automatic implementation are presented. In a concluding section, limitations are pinpointed and future improvements are proposed.

2. STRUCTURAL INFORMATION EXTRACTION

The first phase of MAGNO has three steps:

- Edge detection with the Canny operator [Canny86], at a single scale;
- Identification of open and/or closed contours that may correspond to object boundaries in the edge map. A custom contour thinning and following algorithm is used with junction and terminal edgels as starting/ending points [Mokhtari00]. Edgel P is a junction edgel[1] if and only if for $\mathcal{N}_8(P)$, the 8-neighborhood centered on P, $\Sigma val(P_j) \geq 2$, where $val(P_j)$ is the value of pixel $P_j \in \mathcal{N}_8(P)$ and $val(P_j)=1$ if P_j is an edgel and $val(P_j)=0$ otherwise. Edgel P is a terminal edgel if and only if for $\mathcal{N}_8(P)$, $\Sigma val(P_j)=1$.
- Robust multiscale segmentation and approximation of the contours leading to constant curvature segment map or *ccs* map (*ccs*: straight-line segments and circular arcs). These segments are structured according to the image contours and they are referred to as *basic primitives* [Mokhtari00] [Mokhtari01].

3. CONTOUR JUNCTION EXTRACTION

The second phase of MAGNO groups *basic primitives* into *intermediate primitives* corresponding to contour junctions. A significant innovative aspect of the method, in terms of speed and robustness, is the explicit consideration of circular arc primitives in addition to straight-line segment primitives of previous methods [Alquier98] [Etemadi91] [Fuchs95] [Havaldar96] [Horaud90] [Lu92] [Matas93].

3.1 Contour Junction Formation

The formation of contour junctions is based on planar geometrical relations between oriented versions of the extracted *ccs*. Any given *ccs* gives rise to two *oriented segments* referred as v_{ccs}. The oriented segments have complementary starting and ending extremities or endpoints. Any given oriented segment may be a *member* or *participate* to more than one contour junction. For instance, two oriented segments may give rise to a contour junction and the same two with a third segment may give rise to a three-segment contour junction.

A contour junction *J* obtained from a pair of oriented segments has an associated *junction point* in the image plane. This junction point is at the intersection of the supporting axis, line or circle, of each member segment. Besides, it is restricted to be *in front* of each oriented segment. That is, the junction point must appear nearby or after the terminating endpoint. Circular arcs spanning too large a sector (approaching a full circle) have to be processed as a particular case. It is to be noted that many two-segment junctions are directly available from the contour-structured primitives extracted during the previous phase. A contour junction

[1] It is to be noted that junction edgels are not the same as contour junctions introduced below.

previous phase. A contour junction obtained from three or more oriented segments has a junction point defined by the average position of the pairwise intersection points of its member segments.

3.2 Quality Factor

Each contour junction has a *quality factor* associated to it. This is computed from various parameters of its member segments and their structure: lengths, gaps at pairwise intersection points, relative orientation of tangents at pairwise intersection points, etc. The quality factor is a real value number normalized between 0 and 1.

3.3 Rank of Appearance

Each contour junction has also attached to it a *rank of appearance* parameter for each of its member segments. The rank of appearance of a junction for one oriented segment is computed according to *(i)* the *arc distance* between the terminating extremity and the junction point if this latter is lying on the supporting axis or *(ii)* the combination of the *shortest distance* between the junction point and the supporting axis (at point *P*) and the *arc distance* between this point *P* and the terminating extremity if the junction point is not lying on the supporting axis. The rank of appearance is a positive integer value number.

3.4 Contour Junction types

Four types of contour junctions are extracted. Each type gives rise to a list *£ (.)*. The four lists are, in the order of their extraction: *(i) £ (IJ)*, type *INTERSECT* with two v_{ccs} from the same or different contours, *(ii) £ (TJ)*, type *TANGENT* with two tangent, co-linear, or co-circular v_{ccs} from different contours, *(iii) £ (MJ)*, type *MULTIPLE* with three or more v_{ccs} from at least two different contours, and *(iv)* $£_1$ *(OJ)*, type *OCCLUSION* with one v_{ccs} and one *ccs* (on which is the junction point) or $£_2$ *(OJ)*, type *OCCLUSION* with one v_{ccs} and two tangent, co-linear or co-circular v_{ccs}, also from at least two different contours.

The junction detection algorithm builds those lists in turn, combines the last two to form *£ (OJ)*, and then sorts the four resulting lists according to the quality factors of the contour junctions. For each oriented segment, a list of the junctions in which it participates is also built. This list is sorted according to the rank of appearance of the junctions for that segment. The quality factor and rank of appearance of the contour junctions are to be used in the next phase to select a subset of best junctions for the search processes in the generic structure detection. More details of the segment-based junction extraction process appear in [Mokhtari00].

4. GENERIC VOLUMETRIC STRUCTURE EXTRACTION

The extraction of generic volumetric structures in an image of a cluttered scene is the third and final phase of the proposed method. A generic volumetric structure can be represented as an oriented graph in which the *nodes* are contour junctions and the *arcs* oriented segments. A *surface* consists of an ordered group of connected oriented segments forming a closed non-intersecting loop. A *single-surface* is a structure limited to one surface. In order to consider accidental viewpoints, single-surfaces are accepted by the method.

This final phase comprises five stages: *(i)* selection of a subset of the *MULTIPLE* junctions to initiate the search for structures, *(ii)* construction and *(iii)* validation of all potential structures, *(iv)* refinement of each validated structure, and *(v)* extraction of remaining single-surfaces. It is basically a multi-tree search process initiated by selecting the best contour junctions of type *MULTIPLE* as root nodes (so-called *potential mother-junctions*) and developing them on the basis of their member segments (so-called *father-segments*) and the junctions in which they participate.

4.1 Selection of Potential Mother-Junction

The first stage of the extraction process consists in selecting a subset of the *MULTIPLE* junctions to form an ordered list $\ell(MJ)$. The parameters of the selection process are *(i)* the threshold setting the maximum value of the rank of appearance of a junction for any of its member segments, and *(ii)* the threshold setting the minimum value of the quality factor.

Figure 1b presents the potential mother-junctions MJ_i selected for an actual scene composed by two polyhedric objects in occlusion. MJ_i has better quality factor than MJ_{i+1}, consequently *red* circle diameter associated to it is greater. Figure 1a presents constant curvature segments extracted by the first phase of MAGNO.

4.2 Construction Of One Potential Structure

Structure construction starts at a potential mother-junction. Its $n \geq 3$ member segments are considered in turn in order to construct the structure. The way to construct the structure is to extract first its envelope or silhouette. For that reason, the most angularly distant member segments are considered first in developing the search tree. One of these segments is selected as the first father-segment of the tree search at this level.

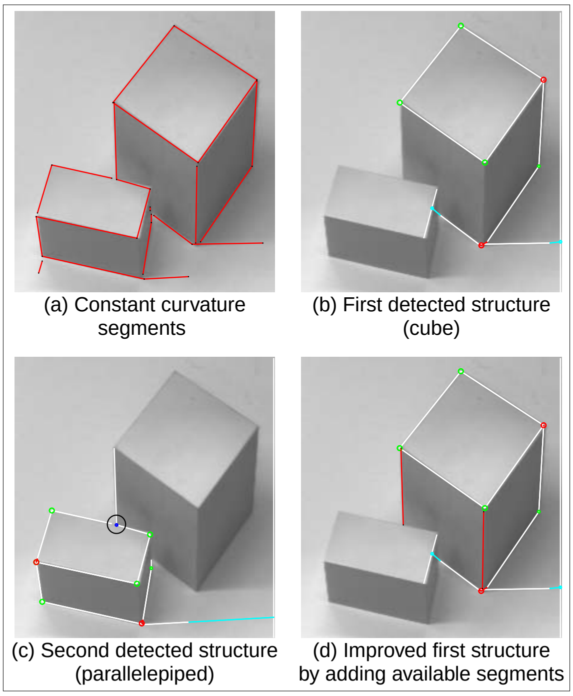
When the added available junction in which the father-segment participates is a two-segment *TANGENT* or *INTERSECT*, the father-segment at the next

5. RESULTS AND DISCUSSION

Results are presented for four real images obtained from cluttered scenes.

5.1 Image Cube + Parallelepiped

In order to correctly detect each of the two structures in Figure 2a, the following scenario was adopted: *(i)* formation of $\ell(MJ)$ for a maximum rank of appearance of *5* and a minimum quality factor of *0.6*, *(ii)* search for structures for the same threshold on the rank of appearance and a minimum quality factor of *0.6* for *MULTIPLE* junctions and *0.75* for *INTERSECT* junctions, and *(iii)* refinement of the structures obtained by addition of available segments not considered during the construction process. The results are illustrated in Figure 2b-d. Added segments to the structure are drawn in *red* in Figure 2d. The two detected structures correspond to the two visible objects. Each structure has a spurious segment originating from a shadow contour. Let us note that the second structure has two co-linear segments belonging to a three-segment *OCCLUSION* junction (see black circle in Figure 2c).



(a) Constant curvature segments

(b) First detected structure (cube)

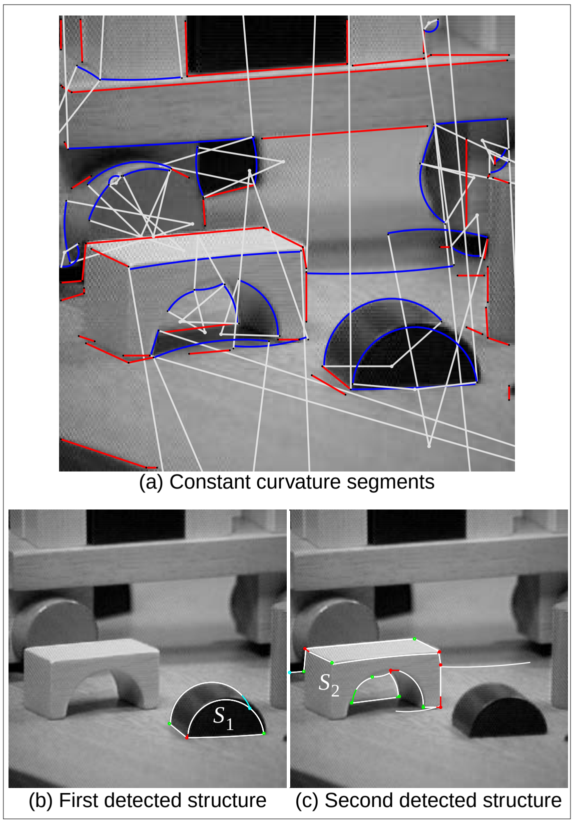
(c) Second detected structure (parallelepiped)

(d) Improved first structure by adding available segments

Figure 2: Two polyhedral objects with an occlusion

5.2 Image Wooden Blocks

This image results from a scene of wooden blocks. The scenario used is the same as above except that no minimum quality factor for *INTERSECT* junctions is considered. The two detected structures S_1 and S_2, Figure 3b-c, correspond to two frontal objects in the image. S_2 includes spurious segments.



(a) Constant curvature segments

(b) First detected structure

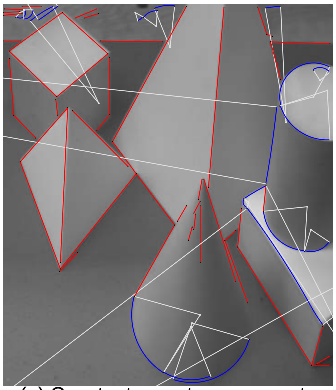
(c) Second detected structure

Figure 3: Two foreground objects with a complex background

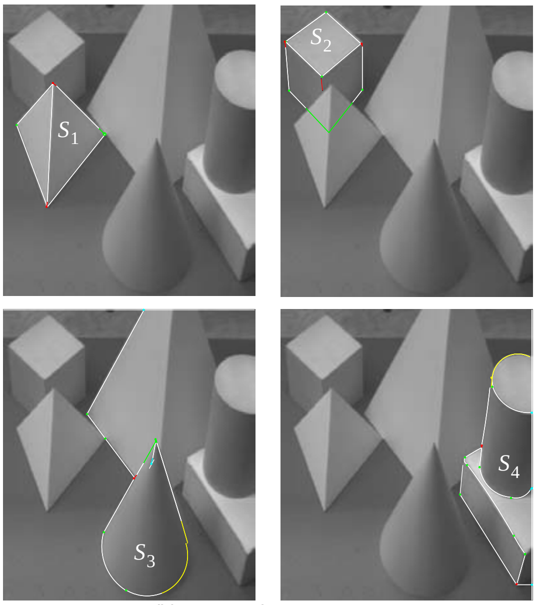
5.3 Image Six Objects

The next image represents six objects: a cube, two pyramids of different size, a cone, a cylinder, and a parallelepiped. Three objects are partly occluded: the large pyramid, the cylinder and the parallelepiped supporting the cylinder (Figure 4a).

The chosen parameters are the following: maximum rank of appearance of *10* and minimum quality factor of *0.6* for *MULTIPLE* junctions. The volumetric structures associated with the small pyramid (Figure 4b-S_1) and the cube (Figure 4b-S_2) are well detected. The cylinder and the parallelepiped (Figure 4b-S_4) are detected together. The cone (Figure 4b-S_3) is detected with spurious segments. Only the large pyramid is not associated with any structure. This could have been predicted, given the fragmentary nature of the available information.



(a) Constant curvature segments



(b) Detected structures S_i

Figure 4: Highly-cluttered scene with many occlusions

5.4 Image Nine Objects

This is our most complex example in terms of the number of objects and their structural arrangement. An additional difficulty is that objects are of different sizes.

For the purpose of the discussion, each object is numbered as indicated in Figure 5a. What should be obtained is the detection and description of nine structures, each one corresponding to an object of the scene. Many spurious primitives from the texture of the supporting table and the shadows are present. The default scenario with a maximum rank of appearance of *5* and no restriction on the quality factors is applied. It provides as output the nine structures appearing in Figure 5d-l, ordered such that S_i has better quality factor than S_{i+1}, where the quality factor of a scene is an average of the quality factors of its member junctions.

As can be seen, the eighth detected structure is a false one. It is due to an accidental arrangement of spurious segments leading to the formation of a spurious *MULTIPLE* junction in the second phase. On the other hand, the single structure extracted for O_1 and O_2 is due to a segmentation artefact in the first phase. Despite these two difficulties, this final example demonstrates the good behaviour of our method even for such a challenging scene.

6. CONCLUSION

An original method was proposed to detect and describe generic three-dimensional structures in an illuminance image. This method at the heart of the MAGNO system comprises three main grouping phases: *(i)* from image data to structural information (basic contour primitives of two types: straight-line segments and circular arcs), *(ii)* from basic primitives to junctions (planar geometrical relations between segments), and *(iii)* from junctions to generic structures corresponding to objects or parts of objects. Experimental results for various images of cluttered scenes have shown an ability to properly detect and describe the structures of volumetric objects or parts with planar and curved surfaces.

In order to focus more precisely on the best junctions in an illuminance image, it would appear judicious to combine information coming from two distinct sources. In [Mokhtari98], a hybrid method for detecting and validating junctions is proposed. This method operates by combining junctions extracted directly in the illuminance image and junctions resulting from grouping of constant curvature primitives.

By its generic nature, MAGNO should also be able to detect and describe manufactured objects in natural environments. Preliminary tests on detecting vehicles in a street scene are encouraging.

7. REFERENCES

[**Alquier98**] Alquier, L. and Montesinos, P., "Representation of Linear Structures using Perceptual Organization", *IEEE Workshop on Perceptual Organization in Computer Vision*, June 26th, Santa-Barbara, CA, USA, 1998.

[**Bergevin93**] Bergevin, R. and Levine, M.D., "Generic Object Recognition: Building and Matching Coarse Descriptions from Line Drawings", *IEEE Trans. on PAMI*, 15(1): 19-36, 1993.

[**Biederman85**] Biederman, I., "Human Image Understanding: Recent Research and A Theory", *CVGIP*, 32:29-73, 1985.

[**Canny86**] Canny, J.F., "A Computational Approach to Edge Detection", *IEEE Trans. on PAMI*, 8(6): 679-698, Nov. 1986.

[**Denasi94**] Denasi, S., Magistris, P. and Quaglia, G., "Saliency Based Line Grouping for Structure Detection", *Intelligent Robots and Computer Vision XIII: Algorithms and Techniques*, Oct. 31-Nov. 2, Boston, MA, USA, pp. 246-257, 1994.

[**Etemadi91**] Etemadi, A., Schmidt, J.P., Matas, G., Illingworth, J. and Kittler, J., “Low-Level Grouping of Straight-Line Segments”, *BMVC*, Glasgow, UK, 1991.

[**Etemadi93**] Etemadi, A., Object Recognition Toolkit (ORT) Version 2.3.1, 1993.

[**Fuchs95**] Fuchs, C. and Förstner, W., “Polymorphic Grouping for Image Segmentation”, *Fifth ICCV*, June 20-23, Cambridge, MA, USA, pp. 175-182, 1995.

[**Havaldar96**] Havaldar, P. and Medioni, G., “Perceptual Grouping for Generic Recognition”, *IJCV*, 20(1): 59-80, 1996.

[**Horaud90**] Horaud, R., Veillon, F. and Skordas, T., “Finding Geometric and Relational Structures in an Image”, *First ECCV*, April 23-27, Antibes, France, pp. 374-384, 1990.

[**Hummel92**] Hummel, J.E. and Biederman, I., “Dynamic Binding in a Neural Network for Shape Recognition”, *Psychological Review*, 99:480-517, 1992.

[**Huttenlocher92**] Huttenlocher, D. and Wayner, C., “Finding Convex Edge Groupings in an Image”, *IJCV*, 8(1): 7-27, 1992.

[**Jacot-Descombes97**] Jacot-Descombes, A. and Pun, T., “Asynchronous Perceptual Grouping: From Contours to Relevant 2D Structures”, *CVIU*, 66(1): 1-24, 1997.

[**Levine92**] Levine, M.D., Bergevin, R. and Nguyen, Q.L., “Shape Description using Geons as 3D Primitives”, *Visual Form: Analysis and Recognition*, C. Arcelli, L.P. Cordella, and G. Sanniti di Baja Editors, Plenum Press, New York, 1992.

[**Lowe85**] Lowe, D.G., *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, 1985.

[**Lu92**] Lu, H.Q. and Aggarwal, J.K., “Applying Perceptual Organization to the Detection of Man-Made Objects in Non-Urban Scenes”, *Pattern Recognition*, 25(8): 835-853, 1992.

[**Matas93**] Matas, J. and Kittler, J., “Junction Detection using Probabilistic Relaxation”, *IVC*, 11(4): 197-202, 1993.

[**Mokhtari98**] Mokhtari, M., Bubel, A. and Bergevin, R., “Robust Extraction of 3D Structures by Fusion of Intensity-Based and Contour-Based Junction Features”, *IAPR Workshop on MVA*, pp. 335-338, Nov. 17-19, Chiba, Japan, 1998.

[**Mokhtari00**] Mokhtari, M., *Multiscale Segmentation and Approximation of Planar Curves: Application to the Extraction of Generic High-Level 3D Structures in an Illuminance Image*, Ph.D Thesis, Dept. of Electrical and Computer Eng., Laval University, Qc, Canada, 2000.

[**Mokhtari01**] Mokhtari, M. and Bergevin, R., “Generic Multiscale Curve Segmentation and Approximation Approach”, *IEEE Workshop on Multiscale and Morphology* (before ICCV 2001), July 7-8, Vancouver, BC, Canada, 2001.

[**Wong92**] Wong, A.K.C. and Gao, Q.G., “A Corner Detector for 3D Object Recognition”, *Intelligent Robots and Computer Vision XI: Algorithms, Techniques and Active Vision*, Nov. 16-18, Boston, MA, USA, pp. 222-229 (vol. 1825), 1992.

[**YlaJaaski96**] Ylä-Jääski, A.S. and Ade, F., “Grouping Symmetrical Structures for Object Segmentation and Description”, *CVIU*, 63(3): 399-417, 1996.

[**Zerroug94**] Zerroug, M. and Medioni, G., “The Challenge of Generic Object Recognition”, *Object Representation in Computer Vision*, M. Hebert Editor, Lecture Notes in Computer Science, LNCS 994, pp. 217-232, 1994.

Figure 5: Many objects of largely varying sizes on a textured plane