Correlation-Based Reconstruction of a 3D Object From a Single Freehand Sketch

Hod Lipson[1] and Moshe Shpitalni[2]

[1]Cornell Computer Aided Design Lab, Mechanical & Aerospace Engineering, Cornell University, Ithaca NY 14853, USA
[2]NSF Engineering Research Center for Reconfigurable Machining Systems, University of Michigan, Ann Arbor, MI 48109, USA*
hod.lipson@cornell.edu

Abstract

We propose a new approach for reconstructing a three-dimensional object from a single two-dimensional freehand line drawing depicting it. A sketch is essentially a noisy projection of a 3D object onto an arbitrary 2D plane. Reconstruction is the inverse projection of the sketched geometry from two dimensions back into three dimensions. While humans can do this reverse-projection remarkably easily and almost without being aware of it, this process is mathematically indeterminate and is very difficult to emulate computationally. Here we propose that the ability of humans to perceive a previously unseen 3D object from a single sketch is based on simple 2D-3D geometrical correlations that are learned from visual experience. We demonstrate how a simple correlation system that is exposed to many object-sketch pairs eventually learns to perform the inverse projection successfully for unseen objects. Conversely, we show how the same correlation data can be used to gauge the understandability of synthetically generated projections of given 3D objects. Using these principles we demonstrate for the first time a completely automatic conversion of a single freehand sketch into a physical solid object. These results have implications for bi-directional human-computer communication of 3D graphic concepts, and might also shed light on the human visual system.

Introduction

In a survey of adequacy of CAD tools for conceptual design (Puttre, 1993), an industrial designer relating to an existing CAD system is quoted saying “The interface is just not for us. I can do thirty sketches on paper by the time it takes me to do two on the computer”. Indeed, it is interesting to watch how a designer, when given a 3D design problem, instinctively reaches for a pencil and paper. Despite the abundance of computerized 3D graphic software and CAD systems, raw sketching has remained one of the most useful and intuitive tools at the conceptual design stage. When designing 3D artifacts, user interfaces that deal with spatial construction are typically cumbersome to use and hamper creative flow. Freehand sketching, on the other hand, still provides a fluent method for conveying 3D information among designers, even though it uses an inherently flat medium. Humans seem to be able to understand 3D spatial concepts even when they are depicted on 2D medium in the form of simple and inaccurate line drawings. It is exactly this ability – to understand and generate sketches – that we wish to emulate.

The importance of sketching in design has been a subject of intensive study (Herbert, 1987; Larkin and Simon, 1987; Fang, 1988; Walderon and Walderon 1988; Ullman *et al*, 1990; Jenkins and Martin, 1995). These studies agree that sketching appears to be important for the following reasons:

- It is fast, suitable for the capacity of short term memory,
- It is implicit, i.e. describes form without a particular construction sequence,
- It serves for analysis, completeness check and simulation,
- It is inexact and abstract, avoiding the need to provide unnecessary details,
- It requires minimal commitment, is easy to discard and start anew

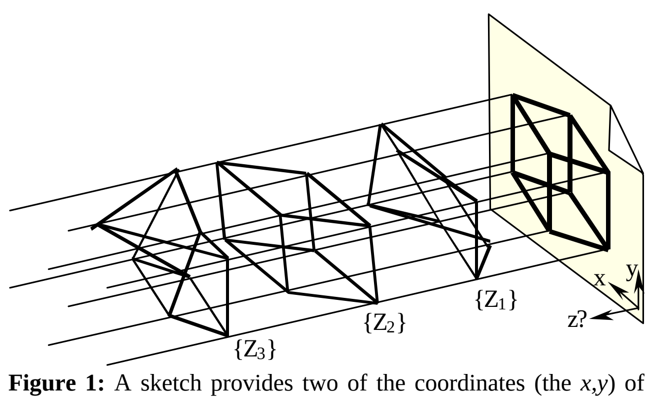


Figure 1: A sketch provides two of the coordinates (the *x*,*y*) of object vertices. A reconstruction must recover the unknown depth coordinates (*Z*). In parallel projections, these degrees of freedom are perpendicular to the sketch plane; in a perspective projection, they run along lines that meet at the viewpoint (not shown).

* On sabbatical leave from the Technion – Israel Inst. of Technology, Haifa, Israel

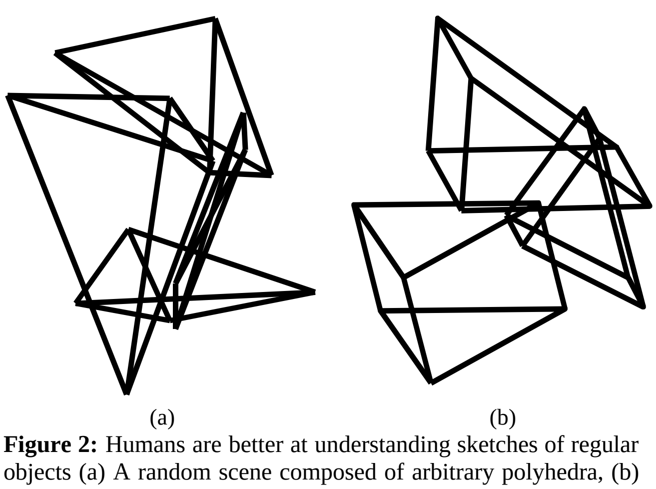
Background

A sketch is inherently a collection of lines (edges) on a flat surface (paper), representing an arbitrary projection of an arbitrary object. In this work we assume all edges of an object are sketched (wireframe) and are straight lines. We also assume an online source, so that each stroke corresponds to a single edge and edges meet at stroke endpoints. The projection transformation removes the depth information from each vertex of the edge-vertex graph. Consequently, any arbitrary set of depths {Z} that are re-assigned to the vertices of the graph constitutes a 3D configuration whose projection will match the given sketch, and so is a candidate reconstruction (Figure 1).

To recover the lost depth information, a system needs to extract spatial information from the inherently flat sketch. Although this step is mathematically indeterminate, humans seem to be able to accomplish this with little difficulty. Moreover, despite the infinitely many possible candidate objects, most observers of a sketch will agree on a particular interpretation. This consensus indicates that a sketch may contain additional information that makes observers agree on the most plausible interpretation.

There are several reports of methods used to reconstruct a 3D object from multiple views by matching features between different views. However, these approaches are not suitable for analysis of a single sketch. The computer-vision literature also deals extensively with techniques for extracting spatial information from images. These methods typically rely on various depth cues such as shading, lighting, occlusion, shadows, perspective, optical flow, stereo and motion cues. These cues are not available in our problem since we are dealing with a single non-imaging source. We are left with only the raw sketch strokes representing edges of the depicted object.

The literature contains several fundamentally different approaches to interpretation and reconstruction of objects and scenes from single-view line drawings. These are briefly described below. Many reported systems use a mixture of these approaches to enhance their performance.



(a) (b)

Figure 2: Humans are better at understanding sketches of regular objects (a) A random scene composed of arbitrary polyhedra, (b) A man-made scene composed of right-angled wedges

Line labeling is a form of interpretation of a line drawing; it provides spatial information about the scene but does not yield an explicit 3D representation. Each line in the drawing is assigned one of three meanings: convex, concave, or occluding edge. Junction dictionaries and constraint graphs are used to find consistent assignments (Huffman, 1971; Clowes, 1971, and many works since).

The gradient space approach draws a relationship between the slope of lines in the drawing plane and the gradient of faces in the depicted 3D scene. Assuming a particular type of projection, an exact mathematical relationship can be computed, and possible interpretations of the drawing can be constrained (Mackworth, 1973; Wei 1987).

The linear System approach uses a set of linear equalities and inequalities defined in terms of the vertex coordinates and plane equations of object faces, determined by whether vertices are on, in front of, or behind the polygon faces. The solvability of this linear program is a sufficient condition for the reconstructability of the object (Sugihara, 1986; Grimstead and Martin, 1995). Linear programming optimization may yield a solution.

Interactive methods gradually build up the 3D structure by attaching facets one after the other as sketched and specified by a user. The aim is to provide a practical method for constructing 3D models in an interactive CAD/CAM environment (Fukui, 1988; Lamb and Bandopadhay, 1990).

The primitive identification approach reconstructs the scene by recognizing instances or partial instances of known primitive shapes, such as blocks, cylinders etc. This approach contains a strict assumption that the depicted 3D object is composed entirely of known primitives, but has the benefit of yielding the final 3D structure in a convenient constructive solid geometry (CSG) form (e.g. Wang and Grinstein, 1989).

The minimum standard deviation approach focuses on a single and simple observation; that human interpretation of line drawings tends towards the most 'simple' interpretation. Marill (1991) defined simplicity as an interpretation in which angles created between lines at junctions are as uniform as possible across the reconstructed object, inflating the flat sketch into a regularized 3D object (Leclerc and Fiscler, 1992).

Analytical Heuristics approaches use coded soft geometrical constraints such as parallelism, skewed symmetry and others to seek the most plausible reconstruction (Kanade, 1980; Lipson and Shpitalni, 1996).

Geometric Correlations

The prevailing common assumption we abandon in this work is *generality*: The misconception that humans' remarkable ability to interpret sketches applies in general to arbitrary objects. In a simple experiment we tried a reversal of roles: We let the computer generate sketches of arbitrary 3D scenes, and asked users to interpret those sketches. We observed that when the scene contained random objects (such as tetrahedrons, Fig 2a), subjects were not able to reconstruct the scene correctly at all from a single sketch; however, when the scene depicted man-made objects (such as right-angled wedges, Fig 2b), the scenes were more readily reconstructable. Hence we concluded that the ability of humans to correctly perceive 3D scenes depicted in sketches is not general, but relies on visual experience. We thus offer the alternative reconstruction approach of acquired geometrical correlations: Humans learn correlations between 3D geometry and its corresponding 2D projected pattern. For example, human might learn to correlate 3D tactile geometric information with 2D images projected on their retina. The following section suggests one way of synthetically capturing and using this information.

At the base of our approach is the need to gather correlations between 3D geometry and its corresponding 2D projections. Whereas general geometrical relationships can be derived analytically (Ulupinar and Nevatia, 1991; Ponce and Shimshoni, 1992), these relationships become more elaborate when collected for non-general scenes. We chose to collect these relationships empirically by generating many 3D scenes (like Fig. 2b) and projecting them with noise (normal distribution σ=2% of object width).

We now define a 3D-2D geometric correlation as probability of a certain 2D configuration to represent a certain 3D configuration. For example, consider Figure 3a. The 3D line-pair *AB* creates a 3D angle $\alpha_{3D}=\angle AB$. When the line pair is projected onto the sketch plane, it produces line-pair *ab*. The projected angle is $\alpha_{2D}=\angle ab$. Measuring this correlation over many arbitrary projections of objects in a certain repertoire, we can derive the probability density function *pdf*(α_{3D}, α_{2D}) for that repertoire of objects. We can then use this probability function to determine the likelihood of a candidate reconstruction.

Instead of just measuring angles, we can measure also line lengths. Here we would measure the correlation between length ratio in 3D $\rho_{3D}=A/B$ to length ration in 2D $\rho_{2D}=a/b$. Similarly, we might chose to correlate *A/B* with *Ðab*, or *ÐAB* with *a/b*, and so forth. Moreover, we can expand these correlations to third order, by correlating various length-angle relationships among three lines, such as the cone angle of three lines in 3D *A´BxC* versus the cone angle in 2D *min(axb, bxc, cxa)*, see Fig 3b.

Higher order correlations may also be recorded in the form of trivariate probability density functions such as *pdf*(α_{3D}, α_{2D}, ρ_{2D}), and even higher orders. In our

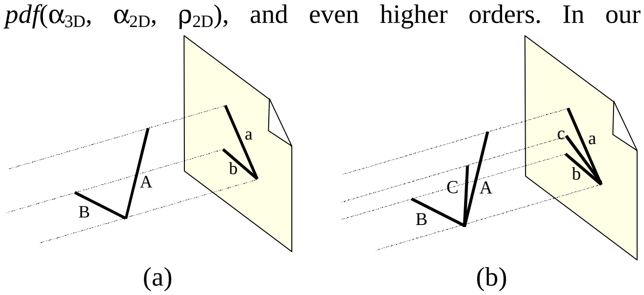


Figure 3: Measuring 2D-3D correlations. (a) second order, (b) third order.

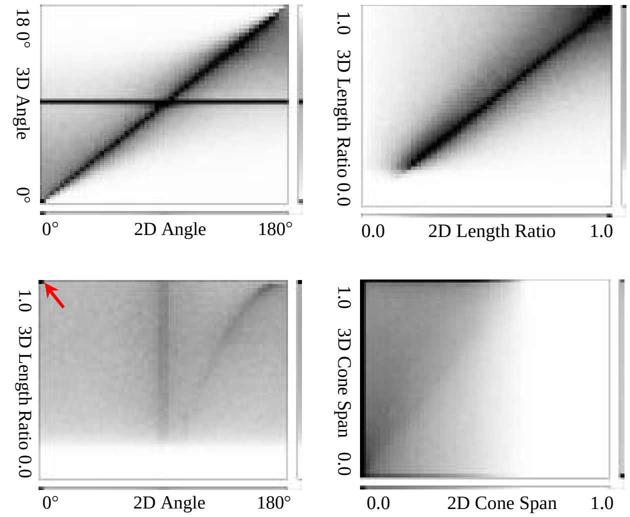


Figure 4: Measuring 2D-3D second order correlations. Dark areas show high correlation. Strips on right and bottom of each table show marginal probabilities. Note dark top-left corner of bottom-left plot.

experiments we used only bivariate probabilities. These were collected for 100,000 random scenes and stored in tables such as those shown in Fig. 4. Note that more efficient correlation memory representations could have been used, such as neural networks or Bayesian networks.

Once geometric correlation functions are known, it is possible to compute the probability of a particular 3D object being the source of a given 2D sketch. This amounts to measuring a 3D angle α_{3D} of line pairs in the candidate reconstruction, and the corresponding 2D angle α_{2D} in the sketch, and using $p(\alpha_{3D}, \alpha_{2D})$ to estimate the probability of α_{3D} given α_{2D}:

$$p(a_{3d} / a_{2d}) = \frac{pdf(a_{3d}, a_{2d}) \cdot da_{3d} \cdot da_{2d}}{pdf(a_{2d}) \cdot da_{2d}}$$

where $\delta\alpha$'s are the tolerances of the measurements. This probability is accumulated (multiplied) for all line pairs/triplets in the candidate object and sketch in question

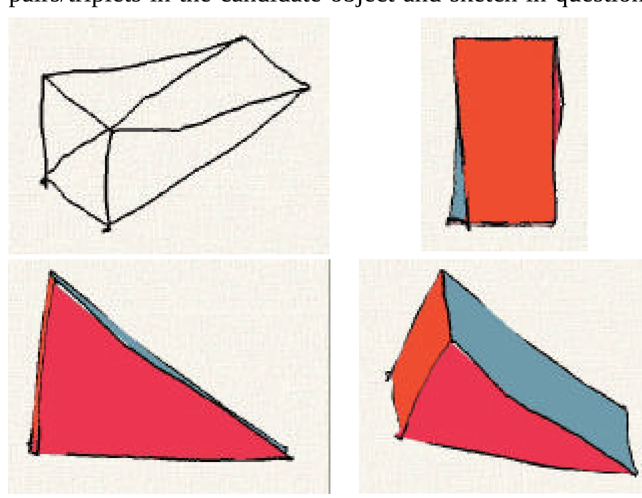


Figure 5: Wedge: Single 2D sketch input and three views of 3D output reconstruction.

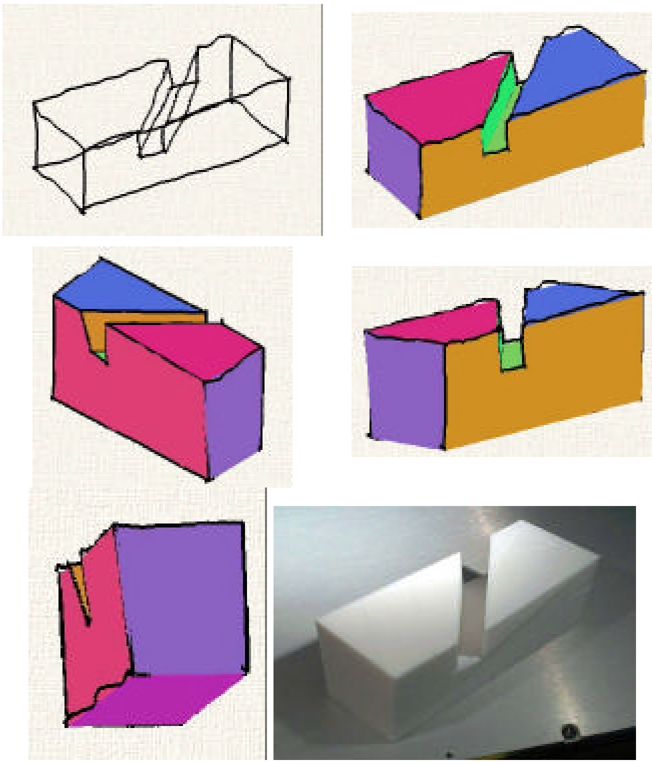


Figure 6: Object "Slot": 2D Single freehand sketch input (top left) and several views of automatically generated 3D reconstruction. Bottom right is automatically generated final physical solid object output. Reconstruction required approx. 5000 hill-climbing cycles (50 Minutes on P500)

over all correlation tables learned, to yield the overall candidate probability.

Once the probability of a candidate reconstruction can be evaluated, then the reconstruction process amounts to an optimization problem, where the objective is to find a set of depth coordinates {**Z**} that maximizes the probability.

However, the optimization process is far from easy. The relatively high degree of coupling among vertices make the optimization landscape rugged with local minima. Moreover, the high dimensionality of the search space (equal to the number of vertices in the sketch minus one) makes brute force search techniques impractical. We have been trying various techniques ranging from straightforward random search to hill-climbing, simulated annealing, and genetic algorithms. Note that for any given solution set **Z**, the inverse solution –**Z** is also equally valid (this is known as the Necker cube illusion). Similarly, the trivial solution **Z**=*constant* also has relatively high probability. This multi-modal nature creates many local optima, and good optimization techniques for this problem still require more research.

The reconstruction step outlined above generates only a 3D wireframe object. In order to complete the transition into a true solid, it is still necessary to identify which of the edge circuits constitute faces of the object, and what is the material side of each face. We use a topological face identification algorithm (Shpitalni and Lipson, 1996) to mark faces, and then chose outward-pointing normals so that joined faces are consistent and the total object volume is positive. Once the 3D solid model exists, it can be tessellated for rendering and for production using commercial 3D Printing (rapid prototyping). The automatic production of a physical model constitutes the ultimate confirmation of the rigor of the interpretation and its topology.

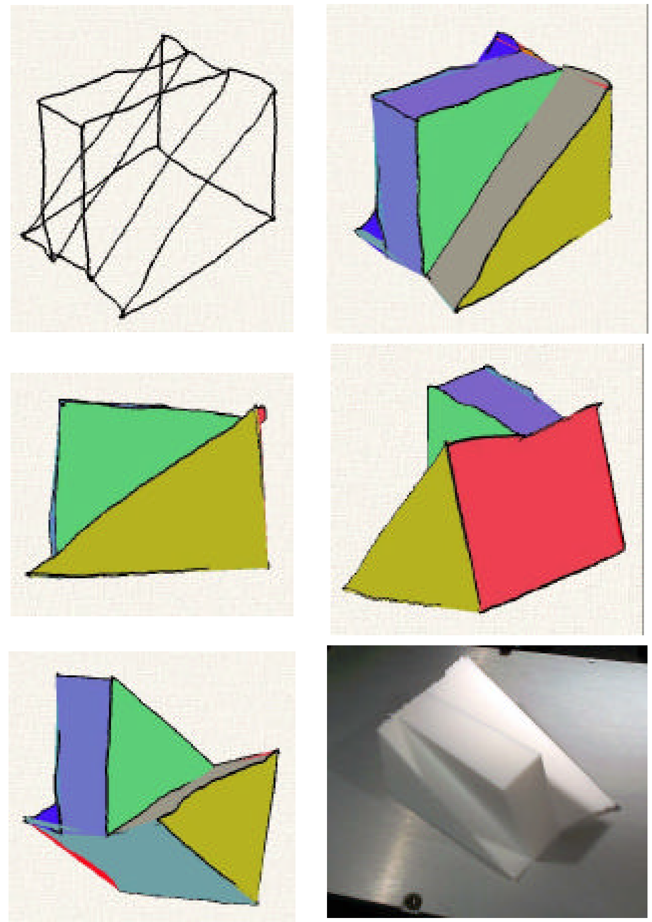


Figure 7: Object "Slide": 2D Single freehand sketch input (top left) and several views of automatically generated 3D reconstruction. Bottom right is final physical solid object output. Reconstruction required approximately 500 hill-climbing cycles (5 Minutes on P500)

Results

Since this article is written on a flat paper, resulting 3D reconstructions will be exhibited themselves as 2D drawing, thereby re-creating the very problem they are trying to solve. Nevertheless, we display each 3D solution rendered from multiple viewpoints to make the interpretation clear. Object faces were colored arbitrarily.

First, Fig 5 shows how a sketch of a right-angled wedge is reconstructed into a 3D wedge. This is merely a confirmation, since the geometric correlations were collected for wedge-based scenes. Figure 6 shows how a

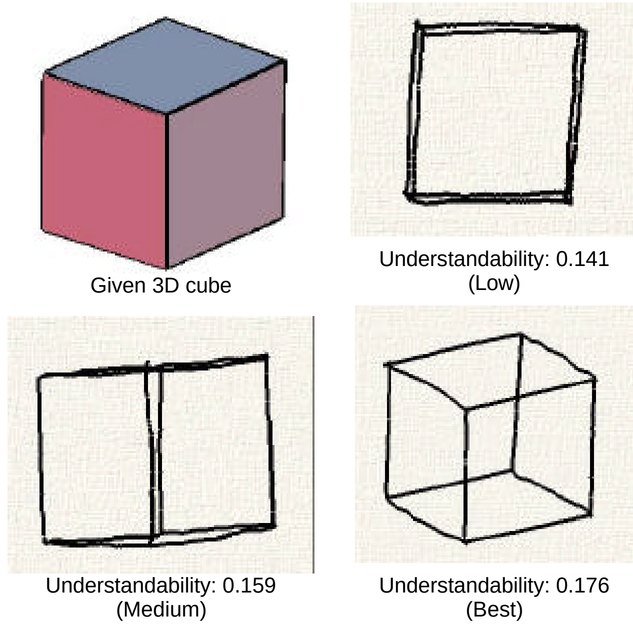


Figure 8: Measuring understandability of sketches: Given 3D Cube (top left), and three projections with varying degrees of correlations, corresponding to degree of "understandability".

more complex structure, not seen by the system during its training period, is reconstructed correctly. Note that the reconstructed object is not accurate – it is a rough 3D object, resembling the roughness of its input. While reconstruction of an accurate 3D model from a rough 2D sketch requires more information (like dimensions and specific constraints), a rough 3D model is useful for many applications. Figure 7 shows an additional example.

Assessing view quality

The correlation tables provide the probability that a 3D model is the source of a 2D sketch. This information can also be used to assess the quality, or "understandability" of a given sketch for a particular 3D scene. This is because a low sketch correlation will make the identification of the correct reconstruction harder. We can thus use the correlation factor as a figure of merit for selecting good sketch viewpoints (Figure 8).

Applications

Enabling a computer to correctly perceive hand-drawn sketches as 3D objects opens an opportunity for new forms of human-computer communication of 3D concepts, and for performing computer aided engineering (CAE) analysis at earlier preliminary design stages (Lipson and Shpitalni, 2000).

Here we investigate the possibility of combining the advantages of non-manipulable pencil-and-paper sketching on real paper, with convenience of 3D manipulation of objects on a computer. Traditionally, a paper sketch is a more fluent medium to describe a 3D object, but once the sketching has commenced, the viewpoint cannot be changed. On a CAD system, viewpoint is easily changed

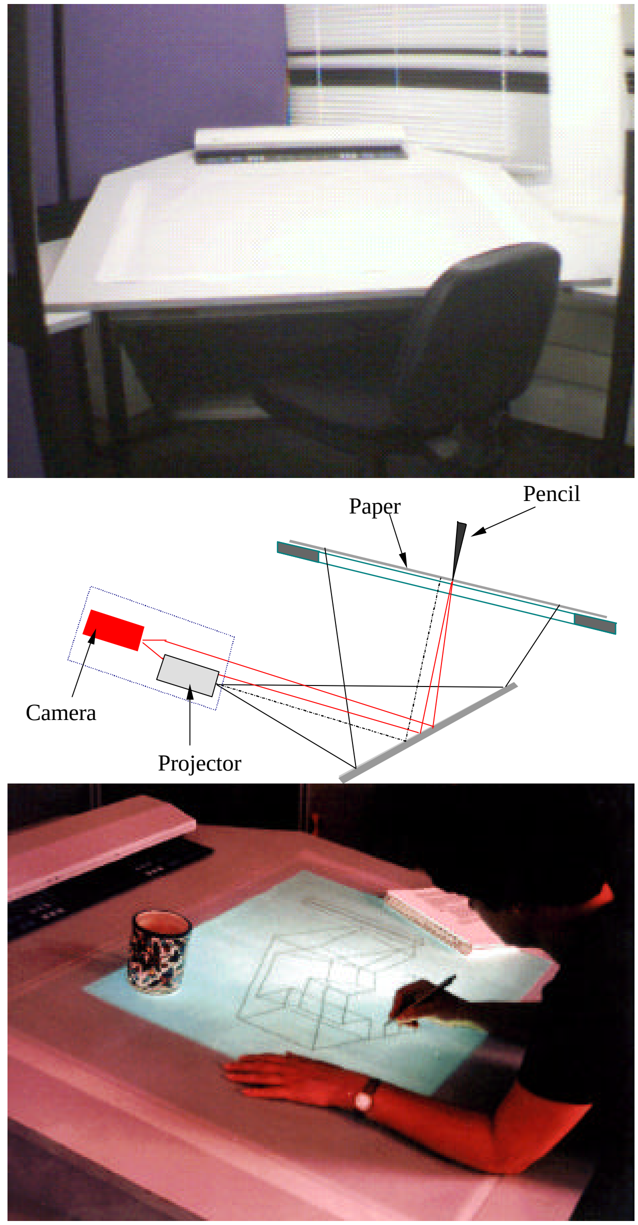


Figure 9: Sketch hardware setup combines natural pencil-and-paper sketching environment but allows changing viewpoint in midst of sketching.

but the direct pencil-and-paper input is lost. However, once a computer is capable of perception of spatial relationships in the sketch, the viewpoint can be changed in midst of sketching. In an experimental system shown in Figure 9 we set up a glass drawing board on which we mounted translucent sheet of paper. The user operated an infrared pen whose motion could be detected by a camera with an IR filter. User's pen strokes on the paper were converted into graphic strokes that were projected back onto the paper. This setup provided an emulation of real pencil-and-paper drawing, while still permitting full interactive ability. After completing partial drawing, the scene could be reconstructed and rotated, so that sketching could be resumed from a different viewpoint. Processing speed, however, still needs to be improved before this can be a truly useful system.

Conclusions

In this paper we have shown how a 2D line drawing can be reverse-projected into three dimensions based on optimizing learned 2D-3D geometric correlations. These correlations are acquired from analyzing many 3D scenes and their corresponding 2D views. Moreover, the same information can be used to judge quality of projections. We demonstrated this approach using four simple second order correlations (angle and length permutations) and one third-order correlation (2D/3D cone angles). We used matrices to store the correlations, and used hill climbing to seek the optimal reconstruction. We demonstrated how this approach could ultimately be used to automatically convert a single rough sketch into a physical solid object without external assistance.

While we appreciate that more advanced correlation representation methods could be used to store higher order correlations more efficiently (e.g. neural networks or Bayesian networks), and more specialized optimization algorithms could enhance our results, we hypothesize that this statistical approach is simpler and perhaps more biologically plausible than traditional constraint-solution techniques proposed to date in the literature. Similarly, while reconstruction time and success rate is still far from being viable for interactive applications, we hope these results may have implications for bi-directional human-computer communication of 3D graphic concepts, and might also shed light on the workings of the human visual system.

References

Clowes M.B., 1971, “On Seeing Things,” *Artificial Intelligence*, Vol. 2(1), pp. 79-112.

Fang R. C., 1988, “2D free hand recognition system”, Master’s report, Oregon State University, Corvallis

Fukui Y., 1988, “Input method of boundary soild by sketching”, *Computer Aided Design*, Vol. 20, No. 8, pp. 434-440

Grimstead I. J., Martin R. R., 1995, “Creating solid models from single 2D sketches”, *Solid Modeling* ’95, Salt Lake City, Utah, USA, pp. 323-337

Herbert D., 1987, “Study drawings in architectural design: Applications of CAD systems”, in Proceedings of the 1987 workshop of the association for computer aided design in architecture (ACADIA)

Huffman D.A., 1971, “Impossible objects as nonsense sentences,” Machine Intelligence, pp. 295-323, Edinburgh University Press, Edinburgh, B. Meltzer and D. Michie, eds.

Jenkins D. L., Martin R. R., 1993, “The importance of free hand sketching in conceptual design: Automatic sketch input”, ASME Conference on Design theory and Methodology (DTM’93), DE-Vol 53, pp. 115-128

Kanade T., 1980, “Recovery of the three-dimensional shape of an object from a single view” *Artificial Intelligence* Vol. 17, pp. 409-460

Lamb D., Bandopadhay A., 1990, “Interpreting a 3D Object from a Rough 2D Line Drawing,” *Proceeding of Visualization* ’90, pp. 59-66.

Larkin J., Simon H., 1987, “Why a diagram is (sometimes) worth a thousand words”, *Cognitive Science*, Vol. 11, pp. 65-99

Leclerc Y. G., Fiscler M. A., 1992, “An optimization based approach to the interpretation of single line drawings as 3D wire frames” *Int. J. of Computer Vision* Vol 9 No 2 pp. 113-136

Lipson H., Shpitalni M., 1996 “Optimization-Based Reconstruction of a 3D Object From a Single Freehand Line Drawing,” *Journal of Computer Aided Design*, Vol. 28 No 8, 651-663.

Lipson H, Shpitalni M., 2000, “Conceptual design and analysis by sketching”, *Journal of Artificial Intelligence in Design and Manufacturing* (AIDAM), Vol. 14, pp. 391-401.

Mackworth A.K., 1973, “Interpreting Pictures of Polyhedral Scenes,” Artificial Intelligence, Vol. 4, pp. 121-137.

Marill T., 1991, “Emulating the human interpretation of line drawings as three-dimensional objects” *Int. J. of Computer Vision* Vol. 6 No. 2, pp. 147-161

Ponce J., Shimshoni I., 1992, “An algebraic approach to line drawing analysis in the presence of uncertainty”, Proceedings of the 1992 IEEE *Int. Conf. On Robotics and Automation*, Nice, France, pp. 1786-1791

Puttre M., 1993, “Gearing up for conceptual design”, *Mechnical Engineering*, March 93, pp. 46-50

Shpitalni, M. and Lipson, H., 1996, "Identification of Faces in a 2D Line Drawing Projection of a Wire frame Object", *IEEE Transactions on Pattern Analysis and Machine Intelligence* (PAMI), Vol. 18, No. 10, pp. 1000-1012

Sugihara K., 1986, *Interpretation of Line Drawings*, The MIT Press.

Ullman, D.G., Wood, S., Craig, D., 1990, “The Importance of Drawing in the Mechanical Design Process,” *Computers & Graphics*, Vol. 14 No. 2, pp. 263-274

Ulupinar F., Nevatia R., 1991, “Constraints for interpretation of line drawings under perspective projections”, *Computer Vision Graphics Image Processing* (CVGIP): Image Understanding, Vol. 53, No. 1, pp. 88-96.

Walderon M. B., Walderon K. J., 1988, “Conceptual CAD tools for mechnical engineers”, in Patton E. M. (Ed.), *Proceedings of Computers in Engineering Conference*, Vol. 2, pp. 203-209, Computer and Graphics, 1988

Wang W., Grinstein G., 1989, “A polyhedral object’s CSG-Rep reconstruction from a single 2D line drawing,” Proc. Of 1989 *SPIE Intelligent Robots and Computer Vision III: Algorithms and Techniques*, Vol. 1192, pp. 230-238.

Wei X., 1987, “Computer Vision Method for 3D Quantitative Reconstruction from a Single Line Drawing,” PhD Thesis, Department of Mathematics, Beijing University, China (in Chinese; for a review in English see Wang and Grinstein, 1993).