

# Reconstruction of a 3D Object From a Single Freehand Sketch

Hod Lipson

Computational Synthesis Lab, Mechanical & Aerospace Engineering,

Cornell University, Ithaca NY 14853, USA

hod.lipson@cornell.edu

## Extended Abstract

This presentation proposes a new approach for reconstructing a three-dimensional object from a single two-dimensional freehand line drawing, as means for a CAD user interface. 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. The approach is based on two phases: In the first *training* phase, 2D-3D geometric correlations are learned from a corpus of 3D objects and their sketches. This phase is carried out offline. In the second *reconstruction* phase, given a sketch to be reconstructed, an optimization process recovers the depth coordinates of sketch vertices so that the learned correlations are maximized. The reconstruction phase is difficult because a hierarchical “Necker-cube illusion” makes the optimization landscape fractal, with an exponential number of local minima. New techniques for overcoming this difficulty will be presented.

A sketch is inherently a collection of lines on a flat surface, representing an arbitrary 2D projection of an arbitrary 3D object. The drawing can be thought of as an edge-vertex graph. The noisy projection from 3D to 2D removed the depth information from each vertex of this edge-vertex graph, and it is our goal to recover that missing depth. As shown in Figure 1, 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 exactly. Each such assignment gives, in principle, a valid candidate 3D reconstruction.

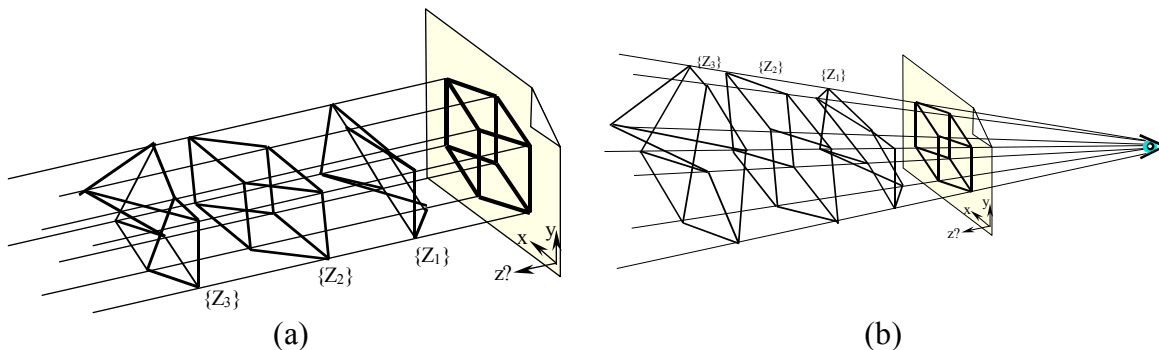


Figure 1: A sketch provides only two of the coordinates (the  $x,y$ ) of object vertices. A 3D reconstruction must recover the unknown depth coordinate ( $z$ ). (a) In parallel projections, this degree of freedom is perpendicular to the sketch plane; (b) in a perspective projection, it runs along lines that meet at the viewpoint. In either case, there are an infinite number of candidate objects – the problem is indeterminate. Each candidate object is represented by a unique set of  $Z$  coordinates, e.g. sets  $\{Z1\}$ ,  $\{Z2\}$  and  $\{Z3\}$ .

To recover the lost depth information, a reconstruction algorithm 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 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. A number of approaches have been proposed [5]. These are not surveyed in this abstract.

The proposed approach comprises three phases: (a) **The learning phase**, where a computer learns the statistical correlations between 3D objects, projections, drawing styles, and 2D drawings, and encodes these correlations in a compact form like a neural network, a Bayesian network, or a probability density function. This stage is done offline using a large corpus of training sketches and models. (b) **The inflation phase** – for a given sketch, an optimization processes tries to find the optimal depth of the vertices of the sketch such that it matches the previously learned correlations. (c) **The fleshing phase** – wraps surfaces around the wireframe and transforms it to a solid model. This presentation focuses on the first two stages.

We define a *3D-2D geometric correlation* as the probability that a certain 2D configuration represents a certain 3D configuration. For example, consider Figure 2a below. 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 correlation between  $\alpha_{3D}$  and  $\alpha_{2D}$  over many arbitrary projections of objects in a certain repertoire, we can derive the probability density function  $pdf(\alpha_{3D}, \alpha_{2D})$  for that repertoire of objects and projections. We can then use this probability density function (PDF) to define a cost function that identifies the most likely 3D object.

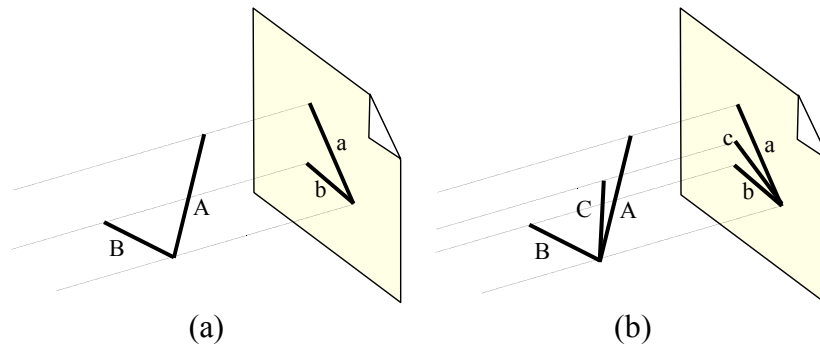


Figure 2: Measuring 2D-3D correlations. (a) second order, (b) third order.

Instead of simply measuring angles, we also can measure line lengths. Here we would measure the correlation between length ratio in 3D  $\rho_{3D} = A/B$  to length ratio in 2D  $\rho_{2D} = a/b$ . Similarly, we might chose to correlate  $A/B$  with  $\angle ab$ , or  $\angle AB$  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  $A \times B \cdot C$  versus the cone angle in 2D  $\min(a \cdot b, b \cdot c, c \cdot a)$ , as shown in Figure 2b. Higher order correlations may also be recorded in the form of trivariate probability density functions (PDFs) such as  $pdf(\alpha_{3D}, \alpha_{2D}, \rho_{2D})$ . Increasing the order of the correlations is equivalent to increasing the context-dependency of the reconstruction. A bivariate PDF looks at two drawing segments, whereas trivariate looks at combinations of a larger number of segments. As the order of the learned PDFs is increased, more training data and more efficient PDF learning mechanisms are necessary (e.g. neural or Bayesian networks, instead of a simple lookup table). It is also plausible, based on neurological observations of the human visual system, that high-order correlations are combined *hierarchically* [7].

The PDFs are essentially convolutions of priors of distributions of geometrical properties of possible objects with geometrical properties of projections. For an ideal case (unbiased object geometry and pure projections) some relationships can be calculated analytically [4,3], but in practice, depicted objects are drawn from a biased repertoire (they are not uniformly random), projections are noisy, and sketching

styles vary. Figure 3 below shows some 2D PDFs collected for 100,000 randomly generated wedge-intersection scenes with noisy orthographic projections.

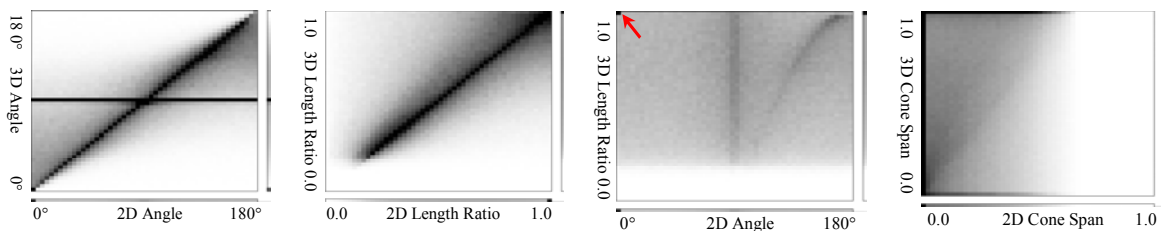


Figure 3: Measuring 2D-3D second order correlations. Dark areas show high correlation. Strips on right and bottom of each table show marginal probabilities. The observed patterns represent heuristics, e.g. the dark corner in the third plot corresponds to an abundance of parallel equal-length lines in wedges.

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

$$p(\alpha_{3d} / \alpha_{2d}) = \frac{pdf(\alpha_{3d}, \alpha_{2d}) \cdot \delta\alpha_{3d} \cdot \delta\alpha_{2d}}{pdf(\alpha_{2d}) \cdot \delta\alpha_{2d}} = \frac{pdf(\alpha_{3d}, \alpha_{2d}) \cdot \delta\alpha_{3d}}{pdf(\alpha_{2d})} \quad (1)$$

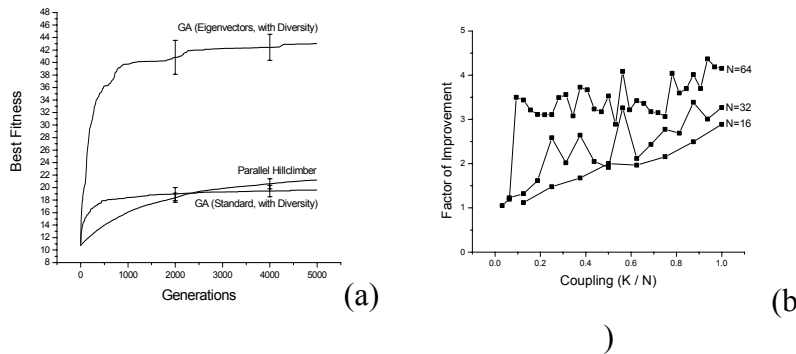
where  $\delta\alpha$ 's are uncertainties of the measurements and the sketching process. This probability is accumulated (multiplied) for all line pairs/triplets in the candidate object and the sketch using all learned correlation PDFs, to yield the overall candidate probability.

Once the likelihood of a candidate reconstruction can be calculated as above, then the reconstruction process ‘amounts’ to an optimization problem, where the objective is to find a set of depth coordinates  $\{Z\}$  that maximizes the likelihood. The high dimensionality of the search space (equal to the number of vertices in the sketch minus one) makes this a difficult combinatorial optimization process.

We now understand properties of this optimization landscape that suggest why the 2D to 3D reconstruction process does not scale well with standard optimization techniques: It has a fractal substructure. For any given solution set  $Z$ , the inverse solution  $-Z$  is also equally valid (this is known as the Necker cube illusion). However, this is true not only for the entire drawing, but also for parts of it: For example, there are two equivalent global optima for interpreting a drawing of a block – the forward and the reverse solutions. The next closest sub-optima will be a block with one side interpreted with the forward solution, and the other side with the reverse solution – but that solution is furthest away from the optimal solutions in the search space (Hamming distance). This structure continues recursively, so that the search space has a fractal structure with an exponential number of local minima, similar to the shuffled hierarchical if-and-only-if (HIFF) problem [6].

We have recently developed new large-scale optimization methods suitable for this type of problem [8]. Our method, based on identifying coupling between degrees of freedom and decomposing the problem, has been shown to significantly outperform standard large-scale optimizers, such as simulated annealing and genetic algorithms, for problems with a high degree of coupling such as the sketch reconstruction problem. The method uses the eigenvectors of the Hessian of the cost function to dynamically transform the problem space so that linkage is tight: Linkage is the relationship between the functional dependency among parameters and their proximity in the problem representation. It is a key to the success of a decomposition-based optimizer like an evolutionary algorithm. A performance comparison using our

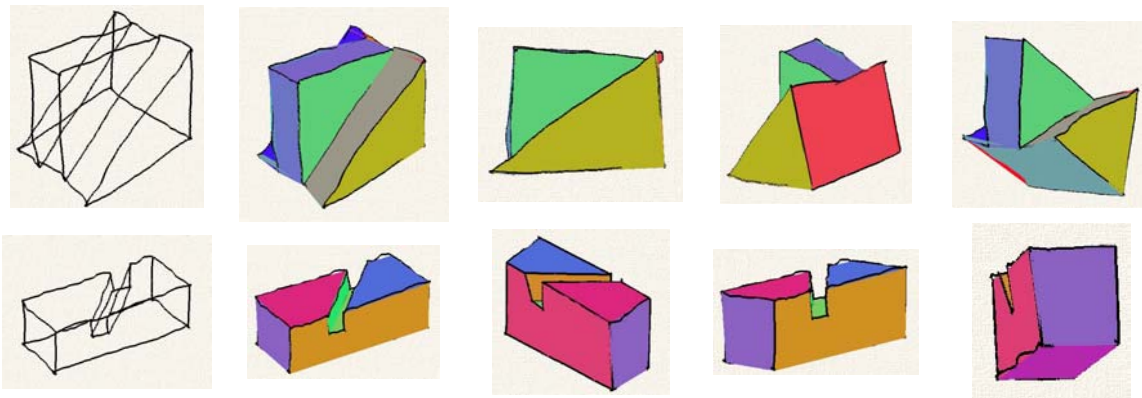
optimizer compared to a standard optimizer on a real-valued variation of Kaufmann's NK-landscapes [1] is shown in Figure 4.



**Figure 4.** Performance of new optimization algorithm [8]: (a) on a shuffled real-valued nk landscape [1] with  $n=64$  and  $k=8$ . (b) Factor of improvement on a range of large coupled problems. Statistics for 10 runs.

### Preliminary results

We tested the proposed principles on a few simple line drawings [2], using a simple lookup-table for 2D PDFs and a simple gradient optimizer. The results are displayed as an input line drawing, and the resulting 3D solution rendered from multiple viewpoints with arbitrarily colored faces. The output was generated automatically directly from the sketch input without any intermediate human intervention, and without specification of heuristics. Figure 5 shows two structures, not seen by the system during its training period, reconstructed correctly.



**Figure 5:** 2D Single freehand sketch input (left) and views of automatically generated 3D reconstruction.

### References

1. Kauffman, S., (1993) The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press.
2. Lipson H., Shpitalni M., 2002, "Correlation-based reconstruction of a 3D object from a single freehand sketch", AAAI Spring Symposium on Sketch Understanding, pp. 99-104
3. 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
4. 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.
5. Wang W., Grinstead G., 1993, "A survey of 3D solid reconstruction from 2D projection line drawings" Computer Graphics forum Vol. 12 pp. 137-158
6. Watson, R.A. and Pollack, J.B. (1999). Hierarchically-Consistent Test Problems for Genetic Algorithms, Proceedings of 1999 Congress on Evolutionary Computation (CEC 99). Angeline, Michalewicz, Schoenauer, Yao, Zalzala, eds. IEEE Press, pp.1406-1413
7. Williamson SJ, Uusitalo MA (1997) "Hierarchy of Processing Memories in the Human Visual System", The American Physical Society, 17-21 March 1997, Kansas City, MO
8. Wyatt D, Lipson H., (2003) "Finding Building Blocks Through Eigenstructure Adaptation", Genetic and Evolutionary Computation Conference (GECCO '03)