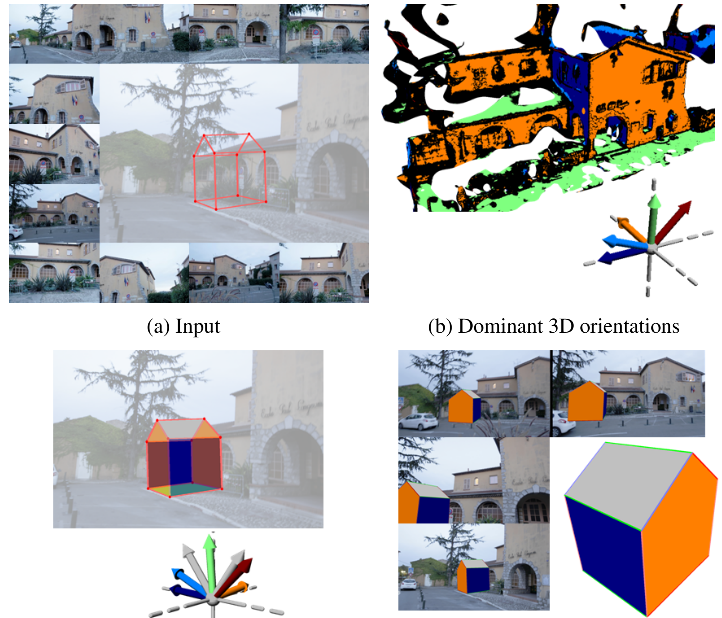
Line Drawing Interpretation in a Multi-View Context

Jean-Dominique FAVREAU, Florent LAFARGE, Adrien BOUSSEAU
Inria Sophia Antipolis, France

Many design tasks involve the creation of new objects in the context of an existing scene. Existing work in computer vision only provides partial support for such tasks. On the one hand, multi-view stereo algorithms allow the reconstruction of real-world scenes, while on the other hand algorithms for line-drawing interpretation do not take context into account. Our work combines the strength of these two domains to interpret line drawings of imaginary objects drawn over photographs of an existing scene.

Recovering a 3D object from a single line drawing is a long-standing problem in computer vision because of the infinity of shapes that can project on the same input [1]. Existing methods resolve such ambiguity by trying to enforce a variety of regularities (symmetry, parallelism, orthogonality, minimal standard deviation of angles) [2]. However, identifying regularity cues from the drawing alone is difficult because typical line configurations often have multiple concurrent interpretations.

Instead of using arbitrary regularity cues, we propose to consider the existing scene as a flexible context-driven regularizer for the new content. In particular, we leverage the fact that man-made environments are often composed of a small set of dominant planes – although not necessarily orthogonal – and that parts of the extensions drawn by designers follow a similar structure. We demonstrate the flexibility of our algorithm with several reconstruction scenarios in architecture, furniture design and archeology.



(a) Input (b) Dominant 3D orientations

(c) Orientation labeling (d) Output

Figure 1: Overview of our approach. (a) Our algorithm takes as input multiple images of a scene along with a line-drawing traced over one of these images. (b) We first compute the dominant orientations of the existing scene from its multi-view stereo reconstruction. (c) Our labeling algorithm estimates the orientation of each facet of the drawing, favoring orientations already present in the scene. We visualize each dominant orientation with a random color, gray denotes new orientations. (d) We finally solve for the 3D model corresponding to the estimated orientations.

1 Algorithm

Figure 1 illustrates the main steps of our algorithms. Starting from multiple photographs of a scene, we apply structure-from-motion and multi-view stereo to estimate a dense point cloud of the scene. In addition to the multi-view dataset, our algorithm takes as input a single line drawing traced over one of the photographs or over a rendering of the 3D reconstruction. We assume that the drawing represents a polyhedron surface, i.e. is composed of straight lines forming closed planar cycles.

We first use Mean-Shift clustering to estimate the dominant orientations of the existing scene (Figure 1(b)). Our core contribution is then to formulate the shape inference as a labeling problem that assigns one orientation to each surface component of the drawing (Figure 1(c)). The assigned orientation can either be one of the dominant orientations of the scene, or be a new orientation only present in the imaginary object. In practice, while our algorithm needs some of the cycles of the drawing to align with existing orientations of the scene, a few such cycles is sufficient to bootstrap the inference of new orientations. Given a 2D drawing and the 3D orientations of its cycles, we solve for the 3D model that best satisfies the orientation constraints while minimizing reprojection error. Figure 1(d) shows the 3D polyhedron produced by our algorithm.

2 Energy formulation

We denote by $\mathcal{G}$ the graph supporting the input drawing, where the edges represent the lines of the drawing and the nodes represent the junctions. We denote by $\mathcal{F}$ the set of simple cycles of the graph, namely facets. Our objective is to estimate the 3D normal of each facet of the drawing graph $\mathcal{G}$. The main idea behind our approach is to select part of these normals from the dominant orientations of the existing scene. We cast this selection as a labeling problem, where $l = (l_i)_{i \in \mathcal{F}} \in L$ denotes the configuration of labels that associates a normal to each facet and L is the configuration space

$$L = \{d_1, .., d_m, d_{new}\}^{card(\mathcal{F})} \tag{1}$$

where $d_1, .., d_m$ are the m dominant orientations of the scene and d_{new} is a free 3D vector. This free label is critical to allow our algorithm to capture orientations that are not present in the existing scene. We measure the quality of a configuration l by the energy

$$U(l) = U_{data}(l) + \beta U_{prior}(l) + \gamma U_{complexity}(l) \tag{2}$$

where $U_{data}(l)$ evaluates the coherence of a configuration l with respect to the input drawing, $U_{prior}(l)$ is a weak geometric prior to penalize flat interpretations, and $U_{complexity}(l)$ penalizes the use of free orientations d_{new}.

Our energy U is composed of a data term that does not respect the conditional independence hypothesis, and a complexity term that acts globally on the label configuration. We minimize this energy using the Metropolis-Hastings algorithm, a stochastic optimization known for its flexibility.

3 Conclusion

We have presented an approach to interpret line drawings of 3D objects when the drawing represents the extension of an existing scene. While little prior work has explored this application scenario, it is a common task in urban planning, furniture design and cultural heritage. At the core of our method is a labeling algorithm that combines the known dominant orientations of the scene with free orientations to offer a trade-off between regularization and discovery of new structures.

[1] H. Barrow and J. Tenenbaum. Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, 17, 1981.

[2] H. Lipson and M. Shpitalni. Optimization-based reconstruction of a 3d object from a single freehand line drawing. *Computer-Aided Design*, 28, 1996.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.