1. Motivation

Two basic problems in image interpretation are:

- To define an appropriate plausibility ordering on possible interpretations.
- To limit the search for interpretations to relatively few plausible choices.

Here we consider a Bayesian approach.

- Plausibility is defined in terms of posterior probabilities.
- The search is controlled by pruning low-probability interpretations.

2. Key Idea

Only the asymptotic order of the posterior probability is used.

The asymptotic order is expressed in terms of:

- a spatial resolution parameter ϵ, and
- a contrast resolution parameter δ.

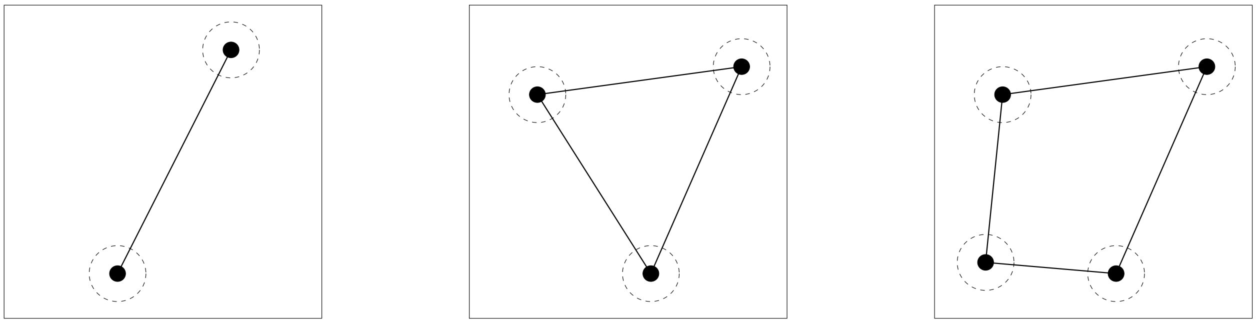
Here $\epsilon \to 0$ and $\delta \to 0$ as the resolution becomes finer.

Consequences.

1. The approach provides a mathematically rigorous notion of qualitative probability based on quantitative probability theory.
2. It is computationally simple.
3. It represents the general structure of the quantitative probabilities.
4. It is insensitive to quantitative details and assumptions about the 'true' priors.

3. Qualitative Prior Probabilities

Eg. Asymptotic order of prior for piecewise linear, 2D chains:



$\Theta(\epsilon^4)$ $\Theta(\epsilon^6)$ $\Theta(\epsilon^8)$

Notation. Here $f(\epsilon) = \Theta(\epsilon^k)$ as $\epsilon \to 0$ if and only if there exist constants $K_1,\ K_2 > 0$ such that $K_1\epsilon^k \leq |f(\epsilon)| \leq K_2\epsilon^k$, as $\epsilon \to 0$.

Derivation. Assume the prior probability, $p(\vec{x}_1, \ldots, \vec{x}_k)$, of a chain of length k with vertices $\vec{x}_i$ satisfies

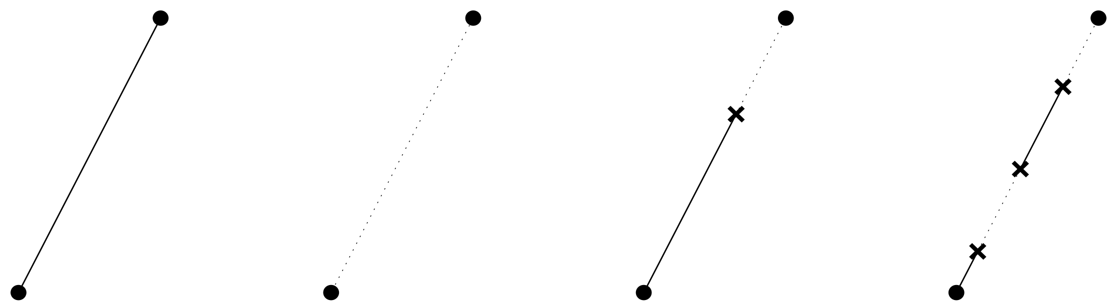
$$0 < d_0 \leq p(\vec{x}_1, \ldots, \vec{x}_k) \leq d_1$$

for some constants d_0 and d_1. Assume these vertices are resolved in the image to within a disc of radius ϵ.

It follows that the prior probability of $\vec{x}_1$ appearing in a particular disc of radius ϵ is $\Theta(\epsilon^2)$ as $\epsilon \to 0$. Similarly, the order for any chain of length k is $\Theta(\epsilon^{2k})$ (see the figure above).

4. Qualitative Likelihood

Possible images of a scene line and corresponding likelihoods:



$\Theta(1)$ $\Theta(\delta)$ $\Theta(\delta\epsilon)$ $\Theta(\delta^2\epsilon^3)$

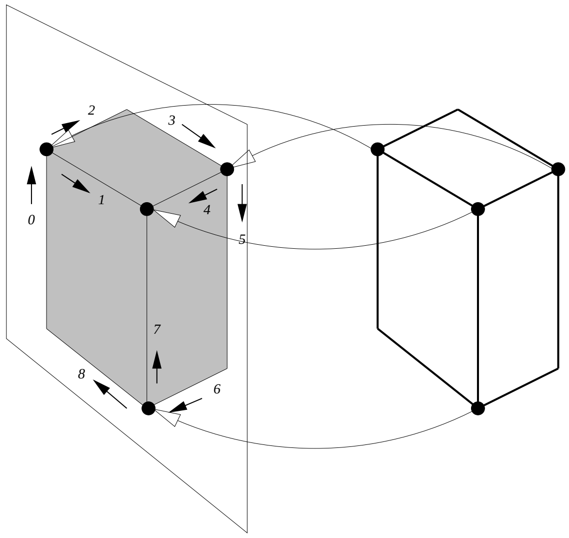
Here line-finder drop-outs are denoted by dotted lines, X's mark interior breakpoints.

Derivation. Let $p(I|L)$ denote the likelihood of observing the image line(s) I, given the scene line L. Then:

- Drop-out segments are taken to occur with probability proportional to the contrast resolution δ.
- The prior for a breakpoint at a specific position along the line (resolved to spatial resolution ϵ) is $\Theta(\epsilon)$.

Therefore, the likelihood $p(I|L) = \Theta(\delta^d\epsilon^b)$ where d is the number of drop-out segments, and b the number of interior breakpoints (see the figure above).

5. Example: Card-Blocks World



Card-Block Model Scene

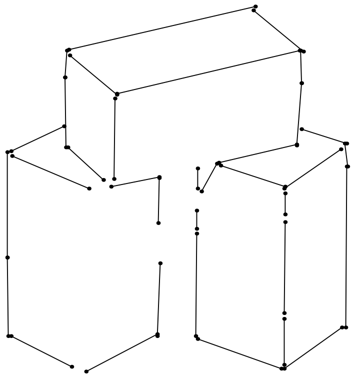
We consider scene models consisting of:

- Card-blocks arranged in depth layers.
- Sticks (isolated line segments).

The interior of a card-block is taken to be opaque, and therefore it can occlude other objects.

Image interpretations must explain all image edges in terms of card-blocks or sticks.

6. Scoring Image Interpretations



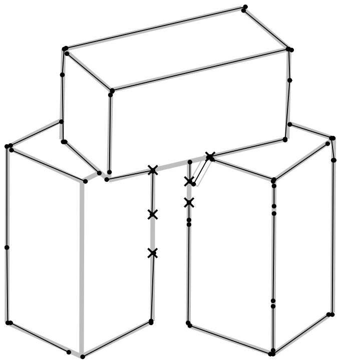


Image "9": 34 Segments Score:$\epsilon^{34}\delta^5$

The posterior probability of the model $\mathcal{M}$ given the image data I is

$$p(\mathcal{M}|I) = \frac{p(I|\mathcal{M})p(\mathcal{M})}{p(I)}.$$

We define the 'score' for an interpretation to be the asymptotic order for the unnormalized posterior, namely $p(I|\mathcal{M})p(\mathcal{M})$.

For the image model depicted above (right), there are:

- 3 instances of card-blocks with 8 resolved dof each (prior $\Theta(\epsilon^8)$ each).
- 1 stick (prior $\Theta(\epsilon^4)$).
- 6 interior breakpoints (likelihood $\Theta(\epsilon)$ each).
- 2 missing edges and 3 drop-out segments (likelihood $\Theta(\delta)$ each).

Together these give an unnormalized posterior of

$$p(I|\mathcal{M})p(\mathcal{M}) = \Theta(\epsilon^{24+4+6}\delta^5) = \Theta(\epsilon^{34}\delta^5).$$

7. Experiments: Image Data.

Khoros line-finder results on blocks-world images:

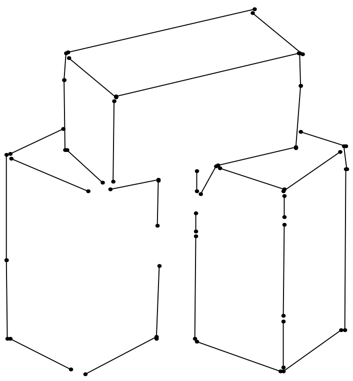


Image "9": 34 Segments

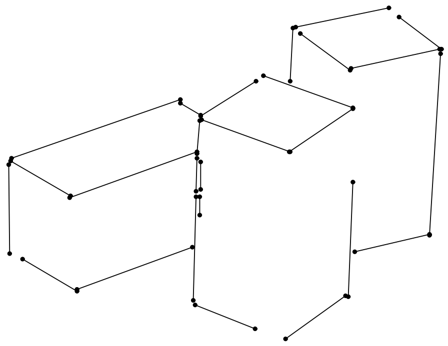


Image "5": 26 Segments

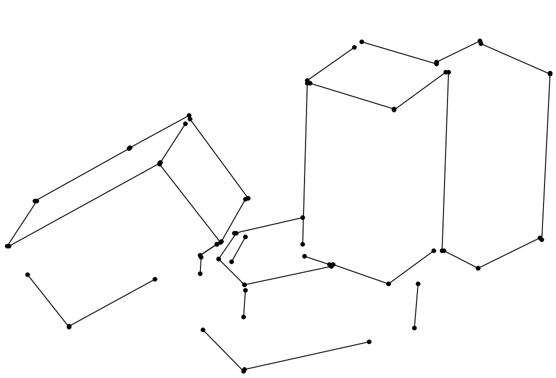


Image "E": 34 Segments

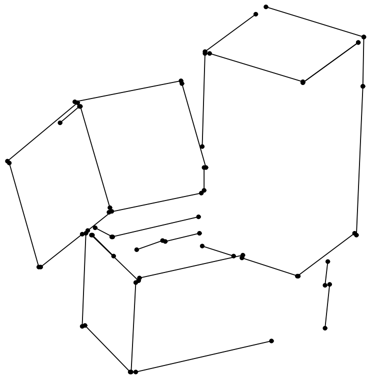


Image "D": 34 Segments

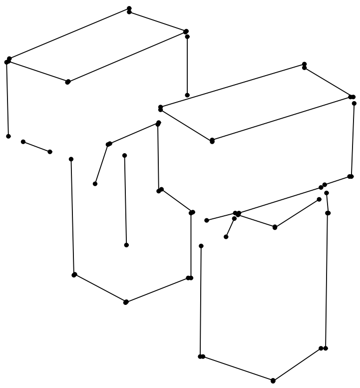


Image "7": 32 Segments

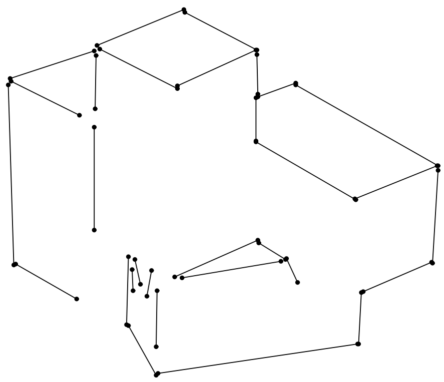


Image "4": 30 Segments

8. Search Strategy.

The search for interpretations proceeds in two stages:

1. Search for maximal subsets of edges consistent with individual blocks.
2. Search for a depth ordering involving a subset of the individual blocks found in stage 1.

Stage 1 Search: Brute Force. Similar to IT-search, Grimson and Lozano-Perez, '97. This is impractical here. It takes 4-6 CPU hours (Pentium I) even on small examples (see below), and produces several thousand individual block interpretations.

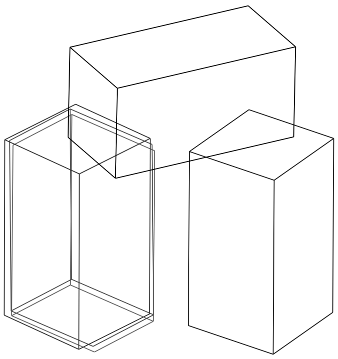
Stage 1 Search: Pruned by the Posterior Odds. Let $\mathcal{P}$ denote the current interpretation during the search, where $\mathcal{P}$ consists of a single (partial) card-block along with sticks used to account for all other image edges. Let $\mathcal{S}$ denote the 'all sticks' interpretation. Then this search branch through $\mathcal{P}$ is pruned whenever the posterior odds in favour of $\mathcal{P}$ over $\mathcal{S}$ is too small.

Note these posterior odds can be written in terms of the 'scores' for the two interpretations $\mathcal{P}$ and $\mathcal{S}$,

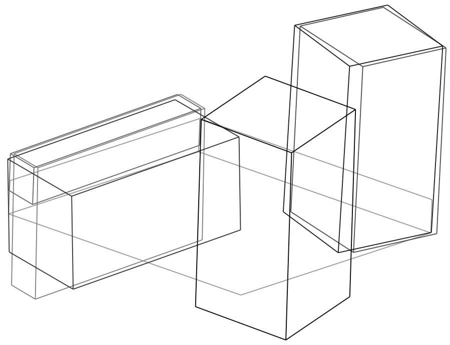
$$\frac{p(\mathcal{P}|I)}{p(\mathcal{S}|I)} = \frac{p(I|\mathcal{P})p(\mathcal{P})}{p(I|\mathcal{S})p(\mathcal{S})}.$$

In each of our test cases (see below), this pruned search required less than a CPU minute (on a Pentium I). Moreover, it produced a dozen or so plausible individual block interpretations, including the most probable blocks found by the brute-force search.

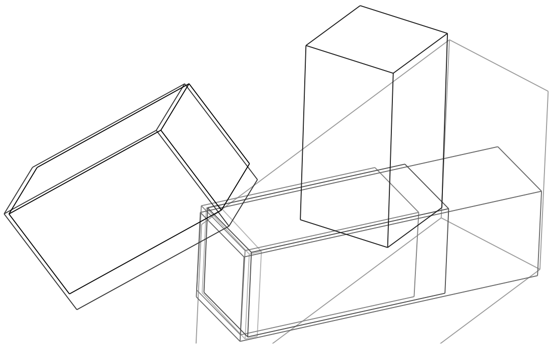
9. Results: Individual Blocks



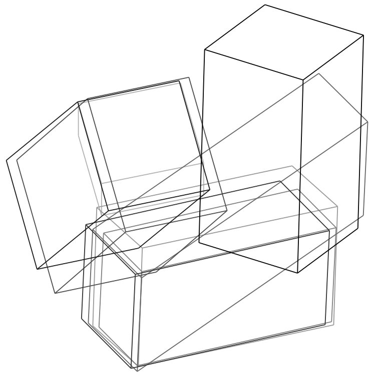
Im9: 5 blocks



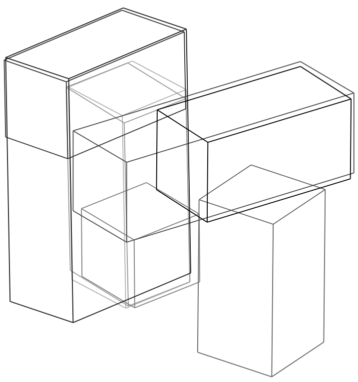
Im5: 7 blocks



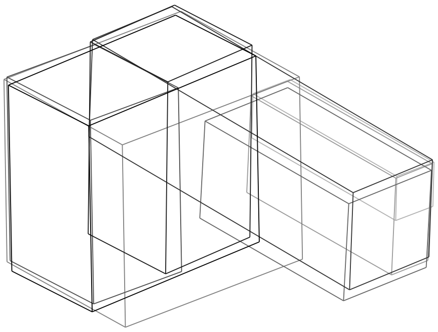
ImE: Best 8 of 12 blocks



ImD: Best 8 of 11 blocks



Im7: Best 8 of 10 blocks



Im4: Best 8 of 18 blocks

10. Search for Complete Interpretations

Given a list of individual blocks from the Stage 1 search, we then perform a **Greedy Band Search** for complete interpretations:

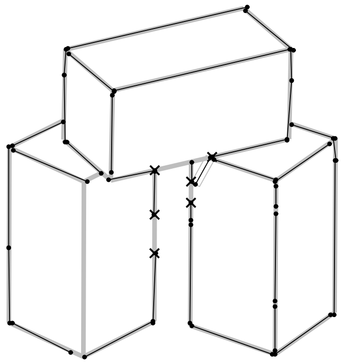
> A band of the N best interpretations found so far (i.e. with the highest unnormalized posterior probabilities) is iteratively updated. New candidate interpretations are generated from the current elements of the band by adding one block to each interpretation in the band, at each feasible depth. Blocks can also be removed, if that increases the posterior probability. The best N interpretations are then selected, and the process is repeated. The search terminates when there is no change in the band.

Stage 2 search performance for bandwidths $N = 1$, 2, 4 and 8:

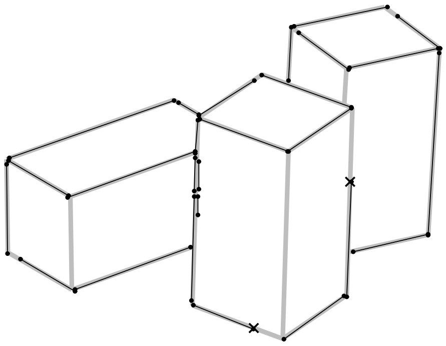
Image	Nodes Visited	CPU Time (sec.)
9	29, 44, 69, 81	6, 10, 17, 20
5	53, 67, 115, 118	12, 17, 31, 57
E	150, 210, 493, 671	56, 72, 192, 257
D	80, 143, 241, 475	20, 37, 67, 142
7	72*, 182, 351, 578	21* 66, 132, 228
4	221*, 479*, 1247, 1602	65*, 160*, 419, 518

Here * denotes the search failed to find the most preferred interpretation (the timing is for a 300MHz Pentium II).

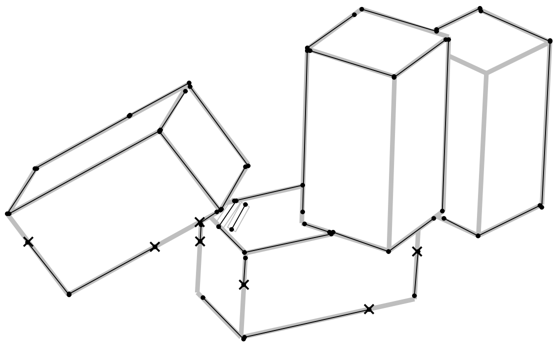
11. Results: Complete Interpretations



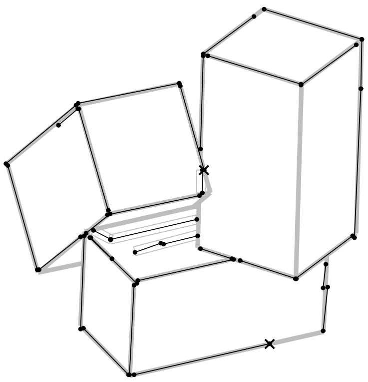
Im9: Score:$\epsilon^{34}\delta^{5}$
(next best $\epsilon^{42}\delta^{7}$)



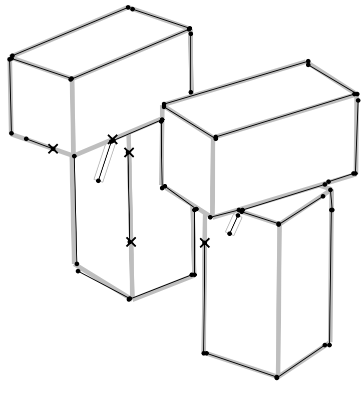
Im5: Score $\epsilon^{26}\delta^{5}$
(next best $\epsilon^{35}\delta^{7}$)



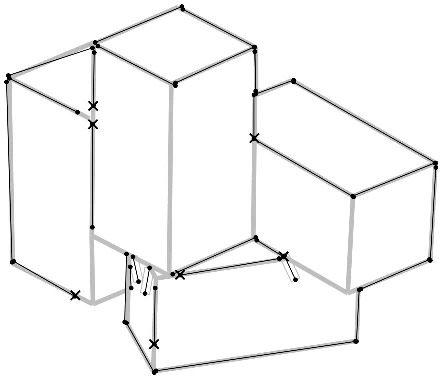
ImE: Score $\epsilon^{47}\delta^{10}$
(1 other within ϵ^{2})



ImD: Score:$\epsilon^{42}\delta^{7}$
(1 other within ϵ^{2})



Im7: Score:$\epsilon^{45}\delta^{8}$
(1 other within ϵ^{2})



Im4: Score:$\epsilon^{51}\delta^{9}$
(7 others within ϵ^{2})

12. Conclusion

- The results indicate that this qualitative probabilistic analysis provides an appropriate plausibility ordering on interpretations in blocks world.
- Qualitative probabilities motivated effective search heuristics for this domain, which involves multiple instances of occluding, parameterized objects.
- Similar approaches are currently being considered for other domains.

References

[1] T. Binford. Inferring surfaces from images. *AIJ*, 17:205–244, 1981.

[2] S. Dickinson, A. Pentland and A. Rosenfeld. From volumes to views: An approach to 3-D object recognition. *CVGIP*, 55(2):130–154, 1992.

[3] J. Feldman. Regularity-based perceptual grouping. *Computational Intelligence*, 13(4):582–623, 1997.

[4] W.E.L. Grimson and T. Lozano-Perez. Localizing overlapping parts by searching the interpretation tree. *IEEE PAMI* 9(4), 1997, pp.469-482.

[5] D. Jacobs. Matching 3-D models to 2-D images. *IJCV*, (21)1/2:123–153 ,1997.

[6] D. Lowe. *Perceptual Organization and Visual Recognition.* Kluwer Academic Publishers, Norwell, MA, 1985.

[7] L.G. Roberts. Machine perception of three-dimensional solids. TR 315, Lincoln Lab, MIT, May 1963.

[8] E. Saund. Perceptual organization of occluding contours of opaque surfaces. CVPR-98 Workshop on Perceptual Organization, Santa Barbara, CA.