The Combinatorics of Heuristic Search Termination for Object Recognition in Cluttered Environments

W. Eric L. Grimson [1]
MIT Artificial Intelligence Laboratory
545 Technology Square, Cambridge, Mass. 02139

Abstract

Earlier work on using constrained search to locate objects in cluttered scenes showed that the expected search is quadratic in the number of features, if all the data comes from one object, but is exponential if spurious data is included. Consequently, many methods terminate search once a "good" interpretation is found. Here, we show that correct termination procedures can reduce the exponential search to quartic. This analysis agrees with empirical data for cluttered object recognition. These results imply that one must select subsets of the data likely to have come from one object, before finding a correspondence between data and model features.

Constrained tree search [e.g. 6,7,10], which identifies data/model feature pairings consistent with a rigid coordinate transformation, is a common approach to object recognition and localization in noisy cluttered environments. Formal analysis of these methods [2] shows that if all of the data are known to have come from one object, the expected amount of search is quadratic, while if spurious data is allowed, the expected search is exponential.

Hence, a hard part of recognition is isolating, from the spurious data, a subset likely to belong to one object. While grouping methods (e.g. generalized Hough transform, or [9,8]) can reduce the search space size [4], they cannot, in general, select sets of data features all from one object, without also encurring a high false positive rate [4].

An alternative is to terminate the search [e.g. 1,6,7,9] once a measure of an interpretation's "goodness" (fraction of the object accounted for) exceeds some threshold. The threshold can be set based on scene clutter and model size [5], so that no false positive solutions are expected. Here, we show how termination reduces the expected search.

1. The constrained search model

Constrained search finds pairings of geometric data and model features, consistent with a rigid model transformation. To find feature matches, we search an interpretation tree depth-first. Nodes at the first level of the tree match the first data feature to each model feature, or to the null character indicating the data feature is not part of the object. Each node then branches to $m+1$ other nodes, where the next data feature is matched to each model feature or the null character, and so on, so that a level n node and its ancestors define a matching of the first n data features. We search the tree depth first, testing each node's consistency with unary and binary constraints [6,7] based on properties like length, relative orientation and relative separation of features. Any constraint involving the null character is always consistent. If any other constraint is false, we backtrack. If we reach a leaf, we verify the data/model pairings by solving for a rigid transformation and testing that it maps all the model features into their matched data features. If so, we save the interpretation, backtrack and continue, until all interpretations are found.

[1]Research funded in part by ONR URI grant N00014-86-K-0685, in part by NSF Grant IRI-8900267, and in part by DARPA under Army contract DACA76-85-C-0010 and ONR contract N00014-85-K-0124.

2. Previous results

Empirically [6,7], this method is very efficient when all the data features are known to come from one object. With spurious data, however, the method slows down considerably. If sets of data/model pairings consistent with similar model transformations are isolated before the search, efficiency improves. If termination is added, the method improves even more. Some of these empirical observations are supported by formal analysis [2]:

- If all s data features lie on one object with m equal size features, the noise is small, and the data is uniformly distributed, then the expected search is bounded by

$$m^2 \leq N_s \leq m^2 + ams.$$

- If only c of the s sensory features lie on an object and the other conditions above hold, then the expected search is bounded above and below by expressions of order

$$O(N_s^*) = m[1+\gamma]^s + ms2^c + \delta m^6 + m^2 s^2 [1+\mu]^c \quad \text{and} \quad o(N_s^*) = m2^c + ms.$$

Here, a, γ, δ, μ are constants that depend on the object and the sensor noise, $\gamma, \mu < 1$. Hence, constrained search is polynomial (quadratic) when all of the data is known to come from a single object, but is exponential when spurious data is included. Here, we consider the effects of heuristic search termination in reducing the exponential cost.

3. Setting up the formal termination model.

The probability that matching the i^{th} data and the I^{th} model feature is consistent is

$$q_{i,I} = \begin{cases} 1 & \text{if } i \mapsto I \text{ is correct, or if } I \text{ is the null character,} \\ p_1 & \text{otherwise.} \end{cases}$$

The probability that the matches $i \mapsto I, j \mapsto J$ are consistent is

$$q_{i,j;I,J} = \begin{cases} 1 & \text{if } i \mapsto I, j \mapsto J \text{ is correct, or if either } I \text{ or } J \text{ are the null character,} \\ p_2 & \text{otherwise.} \end{cases}$$

Given a partial interpretation at a search tree node, the probability of consistency is [2]:

$$\prod_i q_{i,I} \prod_{i \neq j} q_{i,j;I,J}.$$

Given these definitions, one can derive an explicit expression for the expected number of nodes in the tree [3]. For the case of terminating the search once the number of actually matched data features in a valid interpretation exceeds a predetermined threshold, some messy manipulations (which in the interest of space are deferred to [3]) give:

Proposition 1: Given a uniform distribution of correct data features among the spurious, with density $\delta = c/s$, m model features, s data features and a termination threshold t, and since $p_2 = (\kappa/m)^2$ [2], the expected amount of search is bounded by

$$N \leq t\frac{1}{\delta} + \frac{mp_1}{\delta}\left[t^{j_0+1}\mu^{j_0-1}\left(1+(t-1)\frac{\kappa^2}{m^2}\right) + \gamma^{i_0} f(t-1)\left(\frac{1-p_2^t}{1-p_2} + \beta\frac{1-p_2^{(3-\delta)t}}{1-p_2^{(3-\delta)}}\right)\right.$$

$$-\gamma^{i_0}\left[(\frac{1}{\delta}-2)(t-1)+f\right]\left(\left[\frac{p_2(1-p_2^t)}{(1-p_2)^2}-\frac{tp_2^t}{1-p_2}\right]+\beta\left[\frac{p_2^{(3-\delta)}(1-p_2^{(3-\delta)t})}{(1-p_2^{3-\delta})^2}-\frac{tp_2^{(3-\delta)t}}{1-p_2^{3-\delta}}\right]\right)\Bigg]$$

$$N \geq \frac{t}{\delta}+mp_1\left[a\frac{1-p_2^{2t}}{1-p_2^2}-b\frac{p_2^2(1-p_2^{2t})}{(1-p_2^2)^2}+\frac{btp_2^{2t}}{1-p_2^2}\right.$$
$$\left.+\alpha\left(a\frac{1-p_2^{(3-\delta)t}}{1-p_2^{(3-\delta)}}-b\frac{p_2^{(3-\delta)}(1-p_2^{(3-\delta)t})}{(1-p_2^{(3-\delta)})^2}+\frac{btp_2^{(3-\delta)t}}{1-p_2^{(3-\delta)}}\right)\right].$$

where

$$\mu=\frac{\kappa^2}{m}, \qquad \gamma=(s-3)\mu, \qquad i_0=\lfloor(s-3)\mu-1\rfloor,$$
$$j_0=\lfloor\kappa^2-1\rfloor, \quad \beta=mp_1^{1-\delta}p_2^{\frac{2-\delta^2+\delta}{2}}, \quad f=s-t-\frac{1}{2}\left(\frac{1}{\delta}+1\right)$$
$$a=(s-t+2)\left(\frac{1}{\delta}-\frac{1}{2}\right)-\frac{1}{2\delta}\left(\frac{1}{\delta}-1\right), \quad b=\left(\tfrac{1}{\delta}-1\right)\left(\tfrac{1}{\delta}-\tfrac{1}{2}\right), \quad \alpha=mp_1^{1-\delta}p_2^{\frac{3\delta-\delta^2+2}{2}}. \blacksquare$$

Corollary 1.1: The expected search is of order:

$$o(N)=ms\frac{s}{c} \text{ and } O(N)=mts\frac{s}{c}\left(1+\frac{\kappa^2}{m}\right)^2\left(\kappa^2\frac{s}{m}\right)^{\lfloor\frac{s}{m}\kappa^2-1\rfloor}.$$

Proof: For both bounds, we identify, then simplify, the dominant terms:

$$o(N)=m\left(s-t+2-\frac{s}{2c}\right)\frac{s}{c}, \quad O(N)=mt\left(s-t-\frac{c+s}{2c}\right)\frac{s}{c}\left(\frac{m+\kappa^2}{m}\right)^2\left(\frac{\kappa^2 s}{m}\right)^{\lfloor\frac{s}{m}\kappa^2-1\rfloor}. \blacksquare \qquad (1)$$

Corollary 1.2: If $s\kappa^2 < 2m$ then termination has an expected search of order

$$O(N)=amts\frac{s}{c} \qquad \text{and} \qquad o(N)=ms\frac{s}{c}. \blacksquare$$

4. Implications of the results

By Cor. 1.1, terminated search need not be polynomial, although it is reduced from normal constrained search. Cor. 1.2 implies that if the scene clutter is small enough relative to the model size, we do get a polynomial algorithm. When the scene clutter increases, however, we need to select (e.g. [9,8,11]) subsets of data features of size $s < \frac{2m}{\kappa^2}$ while still having at least t features from the object in the set.

This extends earlier results on the role of selection in efficient object recognition. For pure constrained search [2], knowing that all the data features are from a given object reduces the expected search to polynomial, but general constrained search is exponential [2]. This suggests that selection must perfectly isolate relevant data subsets, since if even one spurious point is included, either an exponential search results, or the entire feature subset is rejected. With termination, however, selection can allow an amount of spurious data bounded by the conditions of Cor. 1.2 and still have an efficient search method.

The constant κ depends on properties of the object model and the sensor [2]. Since κ increases with increasing data noise, the expected search also increases, and the amount of

	th = .3 m	.4 m	.5 m	.6 m	Full Search
Predicted lower bound	1,234	1,152	1,069	987	5.4×10^6
Actual nodes, average case, in theory	1,776	1,635	1,498	1,364	
Predicted upper bound	7,017	8,675	9,992	10,969	3.2×10^8
Median, using features	2,689	2,993	2,605	2,143	
Mean, using features	6,223	6,610	9,536	15,340	10^7
Deviation, using features	9,440	9,345	30,278	47,872	
Median, using perimeter	6,627	8,834	8,977	9,479	
Mean, using perimeter	19,437	16,307	23,297	38,362	10^7
Deviation, using perimeter	50,199	34,215	50,062	104,662	

Table 1:

spurious data tolerable decreases. Typically $\kappa \approx .2\frac{P}{D}$ where P is the total object perimeter (for 2D objects) and D is the image dimension. Given this, the conditions for a polynomial search are $s \leq 50m\,(D/P)^2$ so that considerable spurious data is still tolerable.

4.1 Comparing search results

We can extend the earlier analysis [2] as follows (proof in [3]):

Proposition 2: If the data from a correct interpretation are uniformly distributed among the spurious data, then normal constrained search is bounded by

$$m\frac{s}{c}2^c \leq N_{occ} \leq m\frac{s}{c}2^c + \frac{m}{\epsilon}[1+\epsilon]^s \left[1 + \frac{p_2}{1+\epsilon}\right]^{c-1} + \frac{m^3 s}{\kappa^2}\frac{s}{c}[1+p_2]^c . \blacksquare \qquad (2)$$

From Cor. 1.2, for small scene clutter, the expected search reduces to order

$$ms\frac{s}{c} \leq N_{term} \leq mts\frac{s}{c}.$$

Comparing with Prop. 2, heuristic search termination significantly reduces the search.

4.2 Consistency with real data

We also compare this analysis with real data. Features from a cluttered image were placed in 100 random orderings, and the **RAF** [6,7] system was for thresholds of $.3m, .4m, .5m$ and $.6m$, where m is the number of model features, and thresholds of $.3P, .4P, .5P$ and $.6P$, where P is the model perimeter, with appropriate measures of an interpretation. In this example, $m = 20, s = 35, c = 17$. Table 1 lists the predicted bounds (eqn (1)) and actual number of nodes, statistics for each case, and the predicted and observed number of nodes with no termination (eqn (2)).

Note that the derived bounds on the search correctly contain the actual search. Also, the median number of nodes searched, using number of features matched as a termination procedure, lies within the predicted bounds and is in close agreement with the actual theoretical number. The mean search is higher, as expected, since the analysis assumed a uniform distribution of correct data features amoung the spurious. The increase in search when more spurious data are among the first features examined is much larger than the decrease in search when more of the correct features are among the first few features.

We also applied **RAF** to 10 real images, with threshold $.3m$. Figure 1 plots (top to bottom) the predicted upper bound, observed median, predicted actual search, and predicted

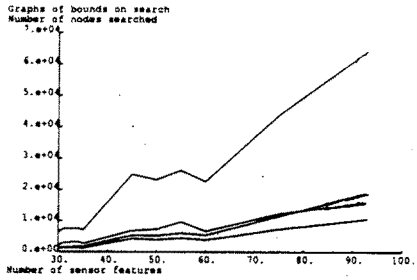

Figure 1:

lower bound, based on 100 trials, all as a function of the number of data features. While other factors can influence both the actual and predicted search, these graphs demonstrate that the predicted number is always between the bounds and is close to the lower bound, and that the observed number of nodes closely follows the prediction.

5. Conclusion

Heuristic termination of constrained search dramatically reduces the expected search in object recognition in cluttered noisy data. To obtain polynomial time algorithms, the ratio of scene clutter to object size must be small enough, and this implies that for significantly cluttered scenes, a selection method is needed to select out data subsets that are likely to include a subset arising from an instance of a known object. Moreover, such methods lead to low order polynomial performance, and to fast practical methods.

References

[1] Ayache, N. & O.D. Faugeras, 1986, "HYPER: A new approach for the recognition and positioning of two-dimensional objects," *IEEE Trans. PAMI*, **8**(1) , pp. 44–54.

[2] Grimson, W.E.L., 1989, "The combinatorics of object recognition in cluttered environments using constrained search," *Artificial Intelligence*, to appear.

[3] Grimson, W.E.L., 1989b, "The combinatoric of heuristic search termination for object recognition in cluttered environments," MIT AI Lab Memo 1111.

[4] Grimson, W.E.L. & D.P. Huttenlocher, 1988, "On the sensitivity of the Hough transform for object recognition," *Second Intl. Conf. on Computer Vision*, Tarpon Springs, FL., pp. 700–706.

[5] Grimson, W.E.L. & D.P. Huttenlocher, 1989, "On Choosing Thresholds for Terminating Search in Object Recognition," 1110, M.I.T. Artificial Intelligence Laboratory.

[6] Grimson, W.E.L. & T. Lozano-Pérez, 1984, "Model-based recognition and localization from sparse range or tactile data," *Int. Journ. Robotics Res.*, **3**(3) , pp. 3–35.

[7] Grimson, W.E.L. & T. Lozano-Pérez, 1987, "Localizing overlapping parts by searching the interpretation tree," *IEEE Trans. PAMI*, **9**(4) , pp. 469–482.

[8] Jacobs, D.W., 1988, "The Use of Grouping in Visual Object Recognition," 1023, M.I.T. Artificial Intelligence Laboratory.

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

[10] Murray, D.W. & D.B. Cook, 1988, "Using the orientation of fragmentary 3D edge segments for polyhedral object recognition," *Intern. Journ. Computer Vision*, **2**(2) , pp. 153–169.

[11] Sha'ashua, A. & S. Ullman, 1988, "Structural saliency," *Second Intl. Conf. on Computer Vision*, Tarpon Springs, FL., pp. 321–327.