Slide 1

An Initial Study of an Adaptive Hierarchical Vision System

Mark Maloof

Department of Computer Science
Georgetown University

maloof@cs.georgetown.edu
http://www.cs.georgetown.edu/~maloof

Seventeenth International Conference on Machine Learning
Stanford University

1 July 2000

Slide 2

Talk Overview

- Hierarchical Vision Systems
- Research Goals
- The Vision Task
- Characteristics of the Data Set
- Experimental Study and Results
- Related and Future Work

Slide 3

Hierarchical Vision Systems

- Start at the pixel level
- Work by grouping low-level constructs into higher-level ones (e.g., lines into corners) and by taking into account more global types of information (e.g., spatial, geometric)
- Pass grouped constructs up to the next level in the hierarchy
- Continue until forming descriptions of objects in the image
- Cannot consider all possible groupings of constructs, so select only the most promising at each level for further processing
- Early work used handcrafted methods for construct selection, such as constraint networks (Mohan & Nevatia, 1989) and linear classifiers (Lin & Nevatia, 1996)
- Recent work has used learning but only at a single level in the hierarchy (Maloof et al., 1997, 1998; Kim & Nevatia, 1999)

Slide 4

Goals of Research

- Interested in multi-level learning for object recognition
  - especially in an interactive, on-line setting
  - especially for tasks requiring low- and high-level reasoning
- Develop multiple architectures based on different computational approaches
  - Bayesian networks versus independently learned classifiers
- Apply implementations to representative vision tasks
  - human eyes, two-hand gestures, buildings
- Examine tradeoffs across tasks and architectures
  - accuracy, user involvement, instances used for learning, etc.

Slide 5

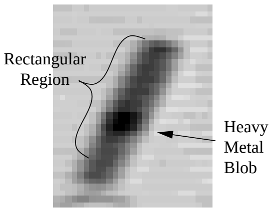
Earlier Work

- With Langley, Binford, and Nevatia
- Used BUDDS (Lin & Nevatia, 1998), a hierarchical vision system for detecting buildings in overhead imagery
- Concentrated on rooftop detection, one level of the hierarchy
- Surveyed a variety of cost-sensitive learning methods
- Used ROC Analysis for evaluation
- Findings:
  - Most learning methods were configured more quickly (informal) and achieved higher accuracy (formal) than BUDDS' handcrafted method (Maloof et al., 1998)
- **Here:** Extend work to multiple levels on a simpler task

Slide 6

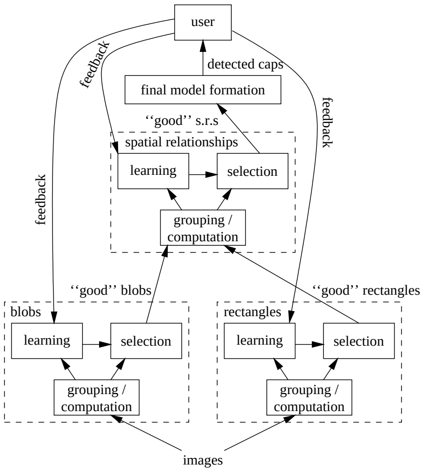
Vision Task: Detecting Blasting Caps

- A simplified task
  - requires a minimal hierarchy
  - simpler than building detection (in some respects)
- Entails detecting
  - the blobular region,
  - the rectangular region, and *then*
  - that the proper spatial relationship exists between the regions
- Similar tasks: eye detection, two-hand gesture detection



Slide 7

An Adaptive Hierarchical Vision System for this Task



Slide 8

Characteristics of the Data Set

- Derived data from 28 X-ray images of airport luggage containing blasting caps
- Identified 204 regions of interest: 102 positive, 102 negative
- Used a variety of intensity and shape features to represent blobular and rectangular regions
- Used Euclidean distance measures to capture the spatial relationship that exists between these two regions

Slide 9

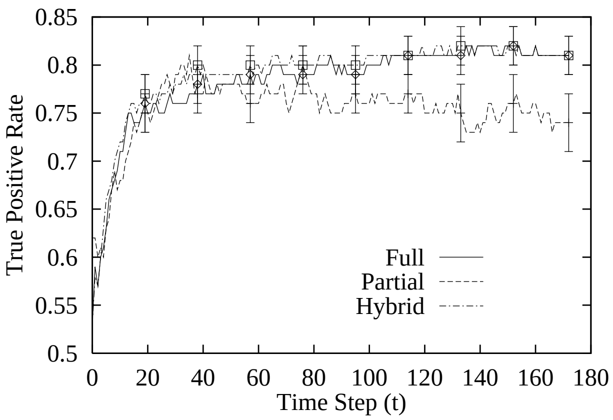
Experimental Study

- Implementation consisted of three on-line, naive Bayes classifiers organized in a hierarchy
- Evaluated three schemes of feedback:
  1. Partial: Only instances for which the user provides feedback are used for learning
  2. Full: All instances are used for learning under the assumption that no feedback means correct classification
  3. Hybrid: Full at the first level, partial at the second level
- Conducted 40 runs with 30 test cases per class per run (~30%)
- Measured (average) predictive accuracy on each task, number of instances used for learning, number of instances on which user provided feedback

Full and partial schemes are similar to the Proximity and Growth algorithms (Kibler & Aha, 1987)

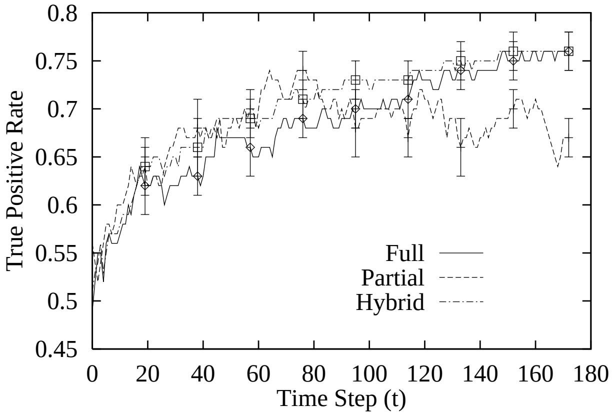
Slide 10

Accuracy for Blobular Regions



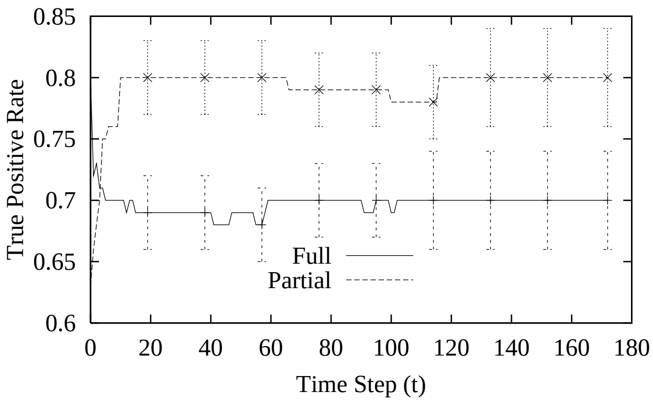
Slide 11

Accuracy for Rectangular Regions



Slide 12

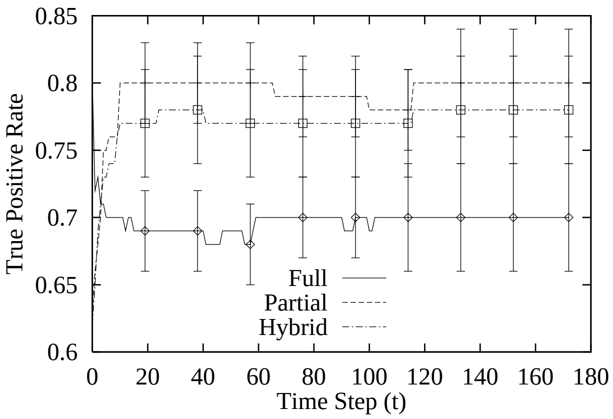
Accuracy for Spatial Relationships



Full @ 2nd level: More examples did not lead to higher accuracy

Slide 13

Accuracy for Spatial Relationships (+Hybrid)

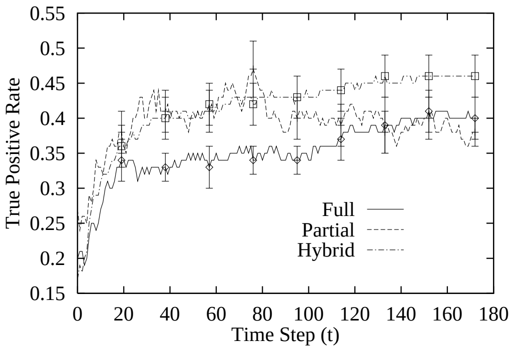


Full @ 2nd level: More examples did not lead to higher accuracy

Slide 14

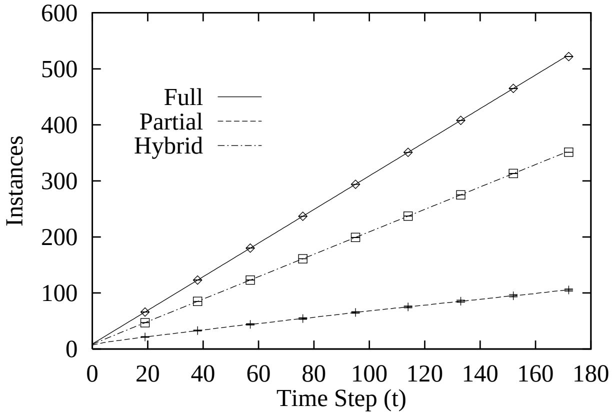
Overall Accuracy

- Generally poor because of propagation of errors
- However, true positive rate for default classifier is 0.002



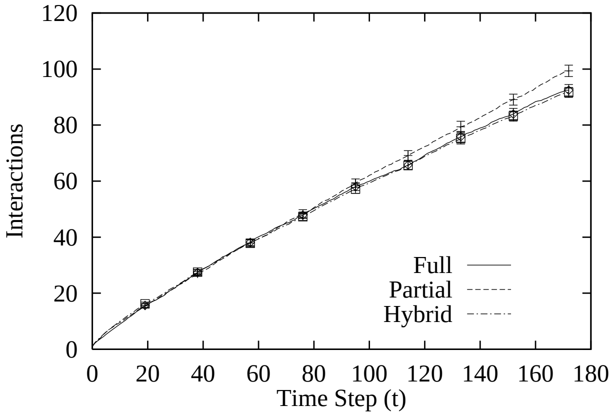
Slide 15

Instances Used for Learning



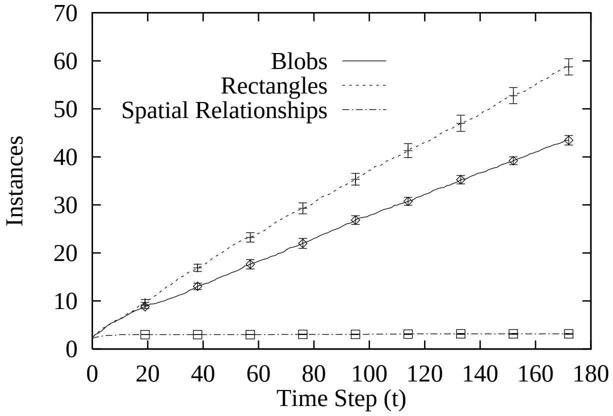
Slide 16

Instances Labeled by the User



Slide 17

Partial Feedback: Instances Used for Learning



Slide 18

Summary of Results

- In the second level of the hierarchy (the spatial-relationships class), using every instance for learning did not yield higher accuracy than the partial method
- Why? Scheme made problems associated with *data skew* more pronounced:
  - with 102 positive blobs and 102 positive rectangles, there are 102 positive spatial relationships
  - with 102 negative blobs and 102 negative rectangles, there are $102^2 = 10,404$ negative spatial relationships.[a]
- With the partial scheme, learning from only misclassified examples kept examples evenly distributed
- Performance did not suffer from few examples because classifying the spatial relationships was a relatively simple task

[a] I miscalculated this in the paper.

Slide 19

Related Work

- Draper (1997), Draper, Bins, and Baek (2000)
  - Reinforcement learning to construct hierarchies
- Colmenarez, Frey, and Huang (1999)
  - Bayesian network for face and facial expression recognition
- Fleuret and Geman (1999)
  - Multi-level learning, automated feature generation
- Kim and Nevatia (1999)
  - Learning for final hypothesis verification

Slide 20

Future Work

- Add constraints to reduce data skew: costs, spatial constraints
- Determine if filtering effect of the partial method holds for more difficult tasks
  - If so, as we add more levels to the hierarchy, will it winnow away too many instances?
- Additional data sets: two-hand gestures, human eyes, buildings
- Relax assumption: user provides feedback for every construct
  - Bayesian network modeling the object hierarchy
  - On-line algorithm: Given the correct class of an object, infer the most likely labels of the constituent constructs
- Examine tradeoffs: accuracy, memory, user interaction
- Compare to these results

References

Colmenarez, A., Frey, B., & Huang, T. (1999). A probabilistic framework for embedded face and facial recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '99)* (Vol. 1, pp. 592–597). Los Alamitos, CA: IEEE Press.

Draper, B. (1997). Learning control strategies for object recognition. In K. Ikeuchi & M. Veloso (Eds.), *Symbolic visual learning* (pp. 49–76). New York, NY: Oxford University Press.

Draper, B., Bins, J., & Baek, K. (2000). ADORE: Adaptive object recognition. *Videre*, *1*(4). (http://mitpress.mit.edu/e-journals/Videre/001/v14.html)

Fleuret, F., & Geman, D. (1999). Graded learning for object recognition. In *Proceedings of the IEEE Workshop on Statistical and Computational Theories of Vision.* Los Alamitos, CA: IEEE Press.

Kibler, D., & Aha, D. (1987). Learning representative exemplars of concepts: an initial case study. In *Proceedings of the Fourth International Conference on Machine Learning* (pp. 24–30). San Francisco, CA: Morgan Kaufmann.

Kim, Z., & Nevatia, R. (1999). Uncertain reasoning and learning for feature grouping. *Computer Vision and Image Understanding*, *76*(3), 278–288.

Lin, C., & Nevatia, R. (1996). Building detection and description from monocular aerial images. In *Proceedings of the Image Understanding Workshop* (pp. 461–468). San Francisco, CA: Morgan Kaufmann.

Lin, C., & Nevatia, R. (1998). Building detection and description from a single intensity image. *Computer Vision and Image Understanding*, *72*(2), 101–121.

Maloof, M., Langley, P., Binford, T., & Nevatia, R. (1998). Generalizing over aspect and location for rooftop detection. In *Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision (WACV '98)* (pp. 194–199). Los Alamitos, CA: IEEE Press.

Maloof, M., Langley, P., Sage, S., & Binford, T. (1997). Learning to detect rooftops in aerial images. In *Proceedings of the Image Understanding Workshop* (pp. 835–845). San Francisco, CA: Morgan Kaufmann.

Mohan, R., & Nevatia, R. (1989). Using perceptual organization to extract 3-D structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(11), 1121–1139.

