Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision (WACV '98), 194–199.

Generalizing over Aspect and Location for Rooftop Detection

Marcus A. Maloof[*] Pat Langley[†] Thomas O. Binford[‡] Ramakant Nevatia[§]

[*]Department of Computer Science
Georgetown University, Washington, DC 20057

[†]Institute for the Study of Learning and Expertise
2164 Staunton Court, Palo Alto, CA 94306

[‡]Robotics Laboratory, Department of Computer Science
Stanford University, Stanford, CA 94305

[§]Institute for Robotics and Intelligent Systems, School of Engineering
University of Southern California, Los Angeles, CA 90089

Abstract

We present the results of an empirical study in which we evaluated cost-sensitive learning algorithms on a rooftop detection task, which is one level of processing in a building detection system. Specifically, we investigated how well machine learning methods generalized to unseen images that differed in location and in aspect. For the purpose of comparison, we included in our evaluation a handcrafted linear classifier, which is the selection heuristic currently used in the building detection system. ROC analysis showed that, when generalizing to unseen images that differed in location and aspect, a naive Bayesian classifier outperformed nearest neighbor and the handcrafted solution.

1 Introduction

Vision systems often use handcrafted knowledge to select visual constructs for further processing, configure visual operators, or choose which visual operators to apply to an image based on context. Once deployed, these systems may have to cope with unforeseen circumstances or variation due to factors such as the time of day or camera position. This suggests a natural task for machine learning: automatically acquire requisite heuristic knowledge, letting generalization and adaptation yield more robust behavior.

We have been investigating this notion using a variety of machine learning techniques and an existing hierarchical vision system that detects buildings in overhead imagery. This system, which we will describe later, uses handcrafted heuristics to select the most promising visual constructs for further processing. In general, such heuristics do not always work well because they are static and do not adapt once deployed in a vision system, and because humans can consider only a small number of images and evidential features when developing them. As a result, we, and other researchers, have begun to investigate *visual learning* approaches in the hope that automated learning techniques will let us survey more images, evaluate and combine more evidence, and compensate on-line for inevitable gaps in training, since learning need not stop: Indeed, it should continue throughout the life of the vision system.

Our goal is to design a methodology for developing, selecting, and using machine learning techniques for combining evidence and improving the behavior of a hierarchical vision system. However, to achieve this goal, we must confront several research issues, including how well various methods perform on a given recognition task, how well different classifiers generalize to unseen images, and how different imaging conditions affect performance.

In this paper, we take steps toward our goal by using machine learning techniques to acquire criteria for selecting rooftop candidates for further processing in the Building Detection and Description System (BUDDS) [1]. Using data derived from six overhead images that differed in location and in aspect, we conducted a series of experiments to determine how well different learning methods performed over a range of misclassification costs and how well the knowledge learned by each generalized to unseen images that differed in location and in aspect. ROC (Receiver Operating Characteristic) analysis [2] indicated that naive Bayes outperformed both nearest neighbor and the handcrafted linear classifier currently used in BUDDS.

Figure 1: Images of the same location taken from different aspects: nadir and oblique.

2 Building Detection

Rather than construct a new vision system, we chose to incorporate learning into a mature, robust vision system that constructs 3-D wire-frame representations of rectangular buildings detected in single, overhead, monocular images [1]. BUDDS is a hierarchical system that works in a bottom-up manner, starting at the pixel level, where it extracts edgels. It then selects linear features, which it subsequently groups into corners and then into "U-constructs." From these, BUDDS forms parallelograms that correspond to rooftops. Using the most promising candidates, BUDDS constructs buildings by matching the rooftops with walls, a process supported by shadow evidence.

A complete technical description of BUDDS is not possible here, but there are two key points to note. First, we chose to begin our study with rooftops because, at this step, BUDDS must often handle many spurious candidates. Second, at each level of processing, BUDDS generates a set of constructs (e.g., rooftops) and then uses heuristics to select the most promising for further processing, meaning that once BUDDS removes a candidate from consideration, it cannot retrieve it. We will return to implications of this for learning after describing the rooftop data.

3 Description of the Image Data

We derived the data for this study from six overhead images of Fort Hood, Texas, collected as part of the RADIUS program [8]. These images were of three regions taken from two different aspects: nadir and oblique. Since we wanted to understand how well learning methods generalized over location and aspect, we selected images that varied relatively little in terms of other factors that affect learning and recognition, such as occlusion and haze. We then used BUDDS to extract rooftop candidates from each image. Figure 1 shows two thumbnail images of a building taken from nadir and oblique aspects.

Table 1: Image and data set characteristics.

Image Number	Location	Aspect	Positive Examples	Negative Examples
1	1	Nadir	197	982
2	1	Oblique	238	1955
3	2	Nadir	71	2645
4	2	Oblique	74	3349
5	3	Nadir	87	3722
6	3	Oblique	114	4395

BUDDS uses nine continuous attributes to represent each rooftop candidate, which summarize evidence gathered from this and lower levels of processing. Positive evidence for the existence of a rooftop includes the strength of edges and corners, the degree to which opposing lines of the candidate are parallel, and support for the existence of orthogonal trihedral vertices and shadows near the corners of the candidate. Negative evidence includes the degree to which the bounding lines fail to form a well-shaped parallelogram, the existence of lines that cross the candidate, L-junctions or T-junctions adjacent to the candidate, and gaps in the edges of the candidate.

Before we could use any of the machine learning methods to acquire the selection criteria for rooftops, we had to label each extracted candidate as either a positive or negative example of this concept. To accomplish this task easily, we implemented a visualization system using Java that draws each rooftop candidate over the image from which it was extracted, letting the user click either a "Rooftop" or "Non-Rooftop" button to label the candidate. It required about 5 hours to label the 17,829 candidates extracted by the vision system, of which 718 were labeled positive and 17,048 were labeled negative. We are investigating additional methods to further reduce the burden of labeling large amounts of training data. Table 1 presents characteristics of the images and the data sets generated for each.

4 Error Costs and ROC Analysis

An important facet of our study is that we evaluated the methods over a range of misclassification costs. Because BUDDS cannot retrieve discarded rooftop candidates, it is better to keep a false positive than to remove a false negative—it removes false positives at later stages of processing where it can draw upon more accumulated evidence. As a result, mistakes on the positive class are more expensive than ones on the negative class. This is complicated by the fact that we have a data set that is highly skewed to-

ward the negative class, which effectively biases learning algorithms toward this class and away from the positive class, the more important of the two. To compensate for these factors, we modified the methods to take into account the cost of classification error.

In a previous study [3], we evaluated several learning methods for the rooftop detection task without taking into account the cost of errors and found that naive Bayes and nearest neighbor showed promise of providing the best tradeoff between the true positive and false positive rates. We continued our experimentation with these two methods but modified them to operate under the influence of a cost heuristic that biases each method toward one of the classes, as described in detail elsewhere [4]. This cost heuristic effectively changes the decision boundary at which a classifier predicts one class versus the other.

Naive Bayes (e.g., [5]) forms probabilistic concept descriptions from training data by estimating the prior probability of each class and the conditional probability of each attribute value given the class. When classifying an instance, this method predicts the class with the highest posterior probability, as computed by Bayes' rule. To incorporate a cost heuristic into naive Bayes, we defined an error cost for each class on the range $[0.0, 1.0]$, where numbers close to one indicate a high cost of making a mistake. We computed the expected cost of a decision as a function of the error cost and the posterior probability, which is minimized for large values of the error cost and the posterior. The cost-sensitive version of naive Bayes predicts the class with the least expected cost.

Nearest neighbor (e.g., [6]) stores each training case in memory. To classify an instance, the method predicts the class of the case in memory that is "nearest" to the instance. For our studies, we used the Euclidean distance function to measure the distance between the query and each example in memory. To incorporate costs into nearest neighbor, we again modified the performance element and computed the expected cost as a function of the error cost and the distance from the query to the closest instances from each class, which is minimized for large values of the error cost and small values of the distance. Cost-sensitive nearest neighbor also predicts the class with the least expected cost.

We made similar modifications to the handcrafted linear classifier, the method currently used in BUDDS. When classifying an instance, a linear classifier predicts the positive class if the weighted sum of the attribute values of the instance surpasses a threshold; otherwise, it predicts the negative class (e.g., [7]). For this method, we used the cost heuristic to move the hyperplane of discrimination *farther* from the hypothetical cluster of examples for the more expensive class, thus enlarging the decision region for that class. The cost-sensitive linear classifier predicts the positive class if the weighted sum surpasses the adjusted threshold; otherwise, it predicts the negative class. We included this method for the purpose of comparison and will refer to it as the BUDDS *classifier*.

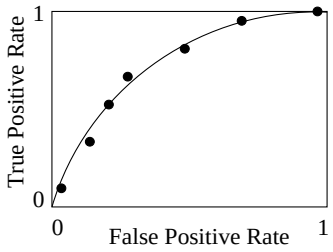

Figure 2: An idealized ROC curve.

Although we knew that detecting rooftops was more important than rejecting non-rooftops, we did not know the exact costs involved. Fortunately, ROC analysis [2] provides a way to evaluate the performance of cost-sensitive methods over a range of costs. An ROC curve plots the false positive and true positive rates for a variety of costs for a given method, as shown in Figure 2. Performance is perfect at the point $(0, 1)$, since the false positive rate is zero and the true positive rate is one. Therefore, we want curves that "push" toward this corner. Traditional ROC analysis uses area under the curve as the measure of performance, which we approximated by summing the areas of the trapezoids produced by each pair of adjacent points on the ROC curve.

5 Experimental Results

When designing our experiments, we wanted to investigate two issues. First, we wanted to know which method performed the best when generalizing over location and aspect, expecting that machine-learned classifiers would outperform the BUDDS classifier.

Second, we wanted to investigate the degree to which each method was able to generalize to unseen images that differed in location and in aspect. We anticipated that the behavior of the learning methods would degrade when generalizing to these unseen images, which we demonstrate by comparing to a baseline performance condition.

5.1 Aspect Experiment

In the first experiment, we controlled for differences in location to test how well the methods generalized to unseen images of different aspects. To perform the

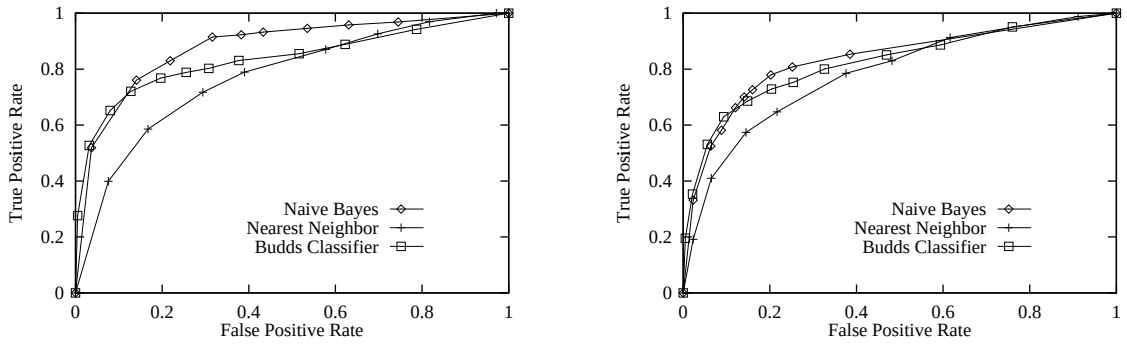

Figure 3: ROC curves for the aspect experiment in which we trained on images from one aspect and tested on images from the other aspect. Left: trained on oblique, tested on nadir. Right: trained on nadir, tested on oblique.

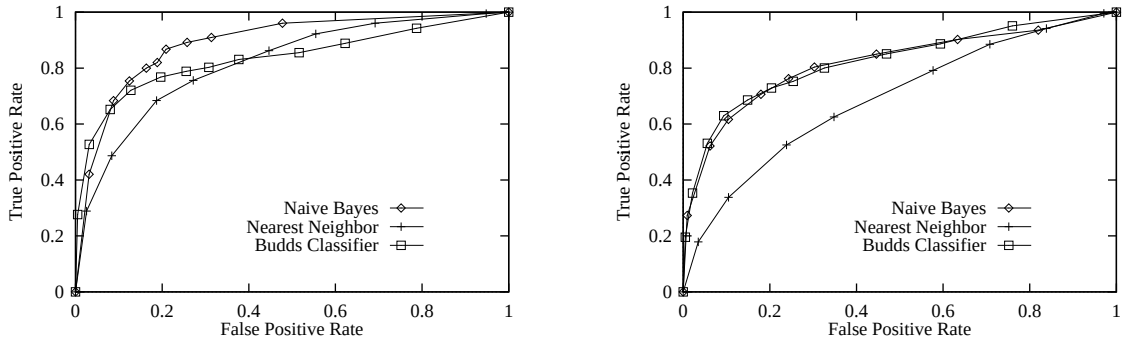

Figure 4: ROC curves for the location experiment in which we trained and tested on images from the same aspect but different locations. Left: trained and tested on nadir. Right: trained and tested on oblique.

Table 2: Approximate areas under the ROC curves with 95% confidence intervals from the location and aspect experiments. The 'Nadir' and 'Oblique' labels indicate the testing images for each condition.

	Aspect Experiment		Location Experiment	
Classifier	Nadir	Oblique	Nadir	Oblique
Naive Bayes	0.878±0.042	0.842±0.063	0.901±0.079	0.831±0.067
Budds Classifier	0.837±0.085	0.831±0.068	0.837±0.085	0.831±0.068
Nearest Neighbor	0.795±0.035	0.785±0.053	0.819±0.058	0.697±0.027

Table 3: Approximate areas under the ROC curves and 95% confidence intervals for the baseline performance condition (i.e., controlling for both aspect and location).

Classifier	Nadir	Oblique
Naive Bayes	0.900±0.012	0.851±0.022
Nearest Neighbor	0.851±0.019	0.791±0.020

experiment, we selected an image from a given aspect and location, constructed classifiers,[1] and tested the resulting concept descriptions over a range of misclassification costs on the image of the same location but from the other aspect. For example, referring to Table 1, we would select image 1 for training and image 2 for testing. We repeated this procedure for each location and both aspects, plotting the average true positive and false positive rates for the methods as ROC curves, which are shown in Figure 3. The approximate areas under these curves appear in Table 2.

When generalizing over aspect for both conditions (i.e., testing on nadir images and testing on oblique images), naive Bayes performed the best, yielding ROC curves with areas of 0.878 and 0.842, respectively. The BUDDS classifier produced curves with areas of 0.837 for the nadir condition and 0.831 for the oblique. Finally, nearest neighbor yielded curves of area 0.795 and 0.785 when testing on the nadir and oblique images, respectively.

5.2 Location Experiment

For the second experiment, we controlled for differences in aspect and investigated how well the methods generalized to images of different locations. We selected a pair of images from a given aspect, trained each method over a range of costs, and then tested the resulting concept descriptions on the third image of the same aspect but a different location. As an example, for the nadir aspect, we trained on images 1 and 3 and tested on image 5. We did this for all pairs of images for each aspect, plotting the average results as ROC curves, as shown in Figure 4. Approximate areas for this experiment also appear in Table 2.

When generalizing over location, naive Bayes outperformed the other methods when testing on nadir images, but tied with the BUDDS classifier when testing on the oblique images. For the nadir aspect, naive Bayes yielded an ROC curve with an area of 0.901, while the BUDDS classifier and nearest neighbor produced curves of area 0.837 and 0.819, respectively. For the oblique aspect, naive Bayes and the BUDDS classifier yielded curves of area 0.831, with nearest neighbor producing a curve of 0.697.

5.3 Baseline Performance Condition

To determine the degree to which generalization occurred, we must establish a *baseline performance*. This will help us understand how well each method performs on the rooftop detection task when differences in aspect or location are not factors.

To this end, we split the data from each of the six images into training (60%) and testing sets (40%), and ran each method over a range of misclassification costs. After ten runs for each image, we computed the average area under the ROC curves for the runs involving the nadir images and for the oblique images, which we present in Table 3.

5.4 Analysis

We first sought to determine the best performing method when generalizing over aspect and over location, anticipating that the learned classifiers would outperform the BUDDS classifier. Although nearest neighbor consistently performed worse than the handcrafted linear classifier, naive Bayes outperformed the BUDDS classifier in three of the experimental conditions and tied it in the fourth. Hence, we view these results as positive and generally supportive of our first research hypothesis.

We were also interested in the degree to which each method was able to generalize to unseen images that differed in location and in aspect. Recall that we predicted that the performance of the learning methods would degrade when generalizing to unseen images differing in aspect and location.

To perform this analysis, we compared each method's performance from the generalization experiments to those from the baseline condition. If we compare the baseline performances of the methods (see Table 3) with the results from the aspect experiment (see Table 2), we see that the performance of naive Bayes and nearest neighbor decreases for both the nadir and oblique conditions.

Conducting the same analysis for the location experiment, we see a similar situation: compared to the baseline condition, the performance of the methods decreased when generalizing over location. The exception is naive Bayes, which performed equally well on the location experiment and the nadir images of the baseline condition. Although we want the performance of the learning methods to degrade as little as possible, we predict that further experimentation with additional images will produce such a degradation. Even with this exception, we view these results as supportive of our second research hypothesis.

It is important to note that generalizing to oblique images appears to pose a more difficult problem than generalizing to nadir images, since the areas under the curves for the oblique conditions are less than those for the nadir conditions. However, notice that the baseline performances for the oblique condition were also less than those for the nadir condition. We suspect that oblique imagery simply poses a more difficult problem than nadir imagery. Additionally, since BUDDS was originally developed using nadir images

[1] We simply applied the BUDDS classifier to the test set.

and later extended to oblique images, the features may not represent oblique rooftops as well as nadir rooftops, which could contribute to this effect.

6 Related Work

Much of the work in visual learning relates to ours but does so along different dimensions. For example, Beymer and Poggio [9] take an *image-based* approach that entails presenting images, usually after a filtering step, directly to a neural network that learns a mapping from images to classes (e.g., faces or gestures). This contrasts our framework, since BUDDS forms explicit 3-D representations of objects. Connell and Brady [10] used learning to generalize 3-D object models of commercial aircraft extracted from overhead images, but they did not apply learning to intermediate steps of processing or present a rigorous experimental evaluation.

Draper *et al.* [11] tested a cost-sensitive decision-tree algorithm on an image labeling task using eye-level images of roads, but they did not evaluate their method over a range of costs. Other researchers (e.g., [12, 13]) have also evaluated a variety of cost-sensitive learning algorithms but did not use visual tasks or ROC analysis. Draper's [14] recent work complements ours: we assume that a human specifies the visual processing steps required to recognize an object, while his approach learns the sequence of operators necessary to perform recognition. Finally, several vision researchers have used ROC analysis to evaluate different neural network configurations for face detection [16], as well as ensembles of classifiers for detecting abnormal tissue in mammograms [17] and Venusian volcanos in synthetic aperture radar imagery [15].

7 Conclusion

In this paper, we have examined how two learning methods generalize to unseen images that differ in location and in aspect on the task of rooftop detection. Experimental results demonstrated that, over a range of costs, naive Bayes outperformed nearest neighbor and a handcrafted linear classifier, using area under an ROC curve as the performance metric. We anticipate that our approach will prove beneficial to other levels of scene analysis within BUDDS. In the future, we plan to investigate this notion by applying our approach to higher (e.g., at the building description level) and lower levels of processing, and we anticipate that we will see similar gains in performance over handcrafted solutions. Our ultimate goal is to incorporate learning into all levels of processing so, when deployed, BUDDS can improve its performance and adapt to novel and unforeseen circumstances with user supervision.

Acknowledgments

We thank Melinda Gervasio, Wayne Iba, Stephanie Sage, and Chen Tsung-Liang for helpful comments, and Andres Huertas and Andy Lin for providing the images and rooftop data. This research was conducted at ISLE and in the Center for the Study of Language and Information at Stanford University. It was supported by DARPA, under grant N00014-94-1-0746, administered by ONR, and by Sun Microsystems, through a generous equipment grant.

References

[1] C. Lin and R. Nevatia, "Building detection and description from monocular aerial images," in *Proceedings of the IU Workshop*, pp. 461–468. Morgan Kaufmann, 1996.

[2] J.A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol. 240, pp. 1285–1293, 1988.

[3] M.A. Maloof, P. Langley, S. Sage, and T.O. Binford, "Learning to detect rooftops in aerial images," in *Proceedings of the IU Workshop*, pp. 835–845. 1997.

[4] M.A. Maloof, P. Langley, T.O. Binford, and S. Sage, "Improving rooftop detection in aerial images through machine learning," Technical report, ISLE, Palo Alto, CA, 1998.

[5] P. Langley, W. Iba, and K. Thompson, "An analysis of Bayesian classifiers," in *AAAI-92*, pp. 223–228.

[6] D.W. Aha, D. Kibler, and M.K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, pp. 37–66, 1991.

[7] J.M. Zurada, *Introduction to artificial neural systems*, West Publishing, St. Paul, MN, 1992.

[8] O. Firschein and T.M. Strat, Eds., *RADIUS: image understanding for imagery intelligence*, Morgan Kaufmann, 1997.

[9] D. Beymer and T. Poggio, "Image representations for visual learning," *Science*, vol. 272, pp. 1905–1909, 1996.

[10] J.H. Connell and M. Brady, "Generating and generalizing models of visual objects," *Artificial Intelligence*, vol. 31, pp. 159–183, 1987.

[11] B. Draper, C. Brodley, and P. Utgoff, "Goal-directed classification using linear machine decision trees," *IEEE PAMI*, vol. 16, no. 9, pp. 888–893, 1994.

[12] M. Pazzani, *et al.*, "Reducing misclassification costs," in *ICML '94*, pp. 217–225. Morgan Kaufmann.

[13] P.D. Turney, "Cost-sensitive classification," *JAIR*, vol. 2, pp. 369–409, 1995.

[14] B. Draper, "Learning control strategies for object recognition," in *Symbolic visual learning*, K. Ikeuchi and M. Veloso, Eds., pp. 49–76. Oxford Press, 1997.

[15] L. Asker and R. Maclin, "Feature engineering and classifier selection," in *ICML '97*, pp. 3–11. Morgan Kaufmann.

[16] H.A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," in *CVPR '96*, pp. 203–208. IEEE Press.

[17] K. Woods, K. Bowyer, and W.P. Kegelmeyer, "Combination of multiple classifiers using local accuracy estimates," in *CVPR '96*, pp. 391–396. IEEE Press.