Sensor Modeling, Probabilistic Hypothesis Generation, and Robust Localization for Object Recognition

Mark D. Wheeler

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213-3891
email: mdwheel@cs.cmu.edu

Katsushi Ikeuchi

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213-3891
email: ki@cs.cmu.edu

Abstract

In an effort to make object recognition efficient and accurate enough for applications, we have developed three techniques—sensor modeling, probabilistic hypothesis generation, and robust localization—which form the basis of a probabilistic object recognition algorithm. To minimize recognition time, our techniques exploit prior knowledge to reduce the number of verifications (the most expensive and critical part of the algorithm) required during recognition. Our approach utilizes statistical constraints generated by modeling the entire sensing process—resulting in more accurate constraints on matches. Hypotheses are pruned by a probabilistic algorithm which selects matches based on image evidence and prior statistical constraints. The reliability of the verification decision is increased by robust localization. We have implemented these techniques in a system for recognizing polyhedral objects in range images. Our results demonstrate accurate recognition while greatly limiting the number of verifications performed.

1 Introduction

We present novel techniques for model-based object recognition that makes use of prior knowledge to make recognition more efficient and accurate. To make object recognition more efficient, we present a probabilistic approach that significantly prunes the number of hypotheses that must be tested by the recognition algorithm. To improve the accuracy of the recognition result, we use accurate statistical constraints and a robust localization method.

Our approach to object recognition strives to exploit our prior knowledge of the sensing process to improve recognition performance and accuracy. Hypothesis generation must compensate for inaccurate prior models by loosening the constraints, thereby increasing the number of incorrect hypotheses that are generated. Most systems model only the geometry of the object, ignoring other factors in the sensing process; the result is inaccurate constraints. *Sensor modeling* is used to generate constraints that are more accurate than previous approaches. The performance of most recognition algorithms degrades when dealing with images that contain unknown objects. By pruning unlikely hypotheses, our *probabilistic hypothesis-generation* algorithm reduces the effect of unknown objects on the recognition efficiency. Correct verification of a hypothesis relies on accurate localization of the hypothesized object. Typical approaches to pose refinement are sensitive to missing object features in the image (i.e., partial occlusion of the object). *Robust localization* provides us with reliable pose estimates that are relatively insensitive to missing image features. These three techniques form the basis of a system that demonstrates the potential to make object recognition efficient and accurate enough for a wide range of applications. These methods were implemented and tested in a model-based vision system that recognizes polyhedral models in range images.

The next three sections detail the main contributions of this paper: sensor modeling, probabilistic hy-

*This research was sponsored by the Avionics Laboratory, Wright Research and Development Center, Aeronautical Systems Division (AFSC), U. S. Air Force, Wright-Patterson AFB, OH 45433-6543 under Contract F33615-90-C-1465, Arpa Order No. 7597 and in part F33615-93-1-1282. The first author was supported by a National Science Foundation Graduate Fellowship. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency, the U.S. Government, or the National Science Foundation.

0-8186-5310-8/94 $03.00 © 1994 IEEE

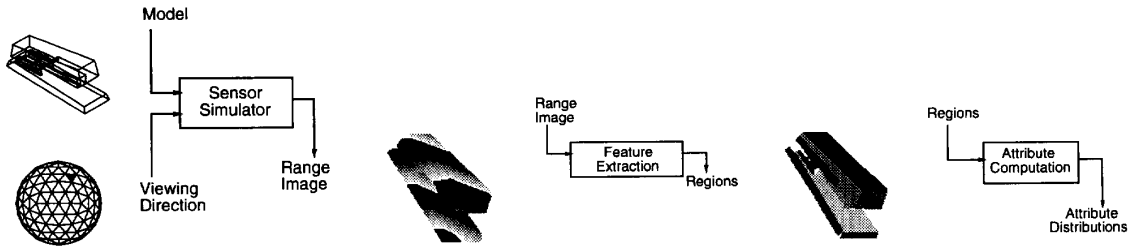

Figure 1: **Generation of feature attribute distributions. An object model and viewing direction are selected. The simulator produces a range image of the object which is segmented into regions for computing the distributions.**

pothesis generation, and robust localization. Experimental results are reported in Section 5. In Section 6, our approach is compared with other approaches.

2 Sensor Modeling

Without an accurate model of the sensing process, the hypothesis-generation procedure must compensate for the inaccuracies by loosening constraints and, thus, increasing the number of incorrect hypotheses that are generated. Our solution is to use *sensor modeling* to build accurate prior models of constraints due to sensor and feature-extraction characteristics in addition to model geometry.

Our current system's sensing modality is range data and our image features are planar regions extracted from the range image. A hypothesis, (R_i, M_j), represents a match between a planar region R_i of the image and model face M_j. Each region R_i is described by a vector of attribute values $\vec{f}_{R_i} = (f^1_{R_i}, f^2_{R_i}, ..., f^n_{R_i})$. The attributes are specified over 3-D surfaces corresponding to planar regions extracted from range images. For this application, our first-order attributes include region area, maximum second moment, minimum second moment, and maximum axis length. Second-order attributes include simultaneous visibility, relative orientation, and maximum distance between surfaces.

The constraints used by our hypothesis generation algorithm are in the form of statistical distributions of the appearance of model faces represented by conditional distributions $P(f^k_{R_i}|M_j)$ and prior probabilities $P(M_j)$. The prior distributions are approximated by generating many (here, 320) sample images of our object models.

Our sensor model is composed of an appearance simulator and a feature-extraction algorithm. We use the Vantage solid modeler [1] to model our objects using constructive solid geometry. Sample range images are generated using an appearance simulator developed by Fujiwara et al. [2]. Regions are extracted from the simulated images, and the attributes of each region are calculated. Prior distributions $P(f^k_{R_i}|M_j)$ are computed for each attribute and each model face. Figure 1 shows an example iteration of this process.

Since this is a simulation, we know the correspondence between the model surfaces and the image regions. Thus, we can build a list of the sampled attribute values for each model surface. The attribute values for each model face are tabulated and used to form the prior distribution. Figure 2 shows a sample distribution of a model face's area value as computed using our sensor model. By comparing the various distributions (see Figure 2), one gets a feel for how inaccurate the constraints are when normal distributions or simple thresholds are used. With simplified models, a large fraction of incorrect hypotheses will not be filtered by the constraints. An additional benefit of our approach is that model features that are not detectable by the feature-extraction program will not affect the hypothesis generation.

3 Probabilistic Hypothesis Generation

Given a set of primitive features (i.e., planar regions) extracted from the input image by a feature-extraction algorithm (i.e., segmentation), the hypothesis-generation procedure produces a set of *possible* model feature to image feature matches (hence referred to simply as hypotheses). Optimally, the gen-

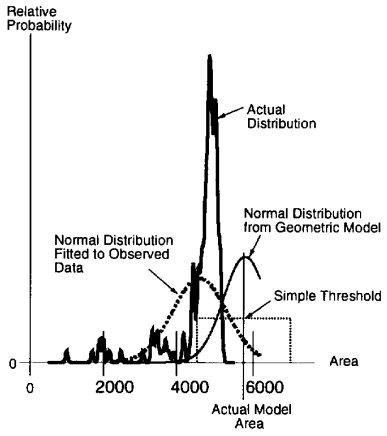

Figure 2: **An example distribution of a given attribute value (area) over a model face generated by sampling resulting area values from synthetic images of an object. Other candidate constraint distributions are shown for comparison.**

erated hypotheses include all of the correct correspondences and exclude as many incorrect ones as possible. To exclude incorrect matches, we must apply constraints derived from our prior knowledge.

We are considering many hypotheses simultaneously and wish to choose the most likely subset of these. We can think of the hypotheses as a set of variables each of which can be assigned a discrete value of **on** or **off**. A hypothesis labeled **on** indicates that the hypothesis is assumed to be correct. The hypotheses display Markovian characteristics. For example, if two hypotheses provide mutual support for each other, and one of them is correct, it is more likely that the other is correct. A similar dependency exists between contradicting hypotheses. These dependencies can be thought of in terms of conditionally dependent probability distributions represented by a Markov random field (MRF). The minimum energy state of the MRF represents the state that best satisfies our constraints—balancing the weight of supporting evidence and contradictory evidence. For a review of MRFs, we refer the reader the description found in [3].

3.1 Formulation of Hypothesis Generation using Markov Random Fields

MRFs are used to represent the probability distribution of values, ω_i, of a set of random variables, X_i. Each variable may be conditionally dependent on a set of *neighbor* variables. Given a set of independent observations, O_i, the most likely state of the MRF variables can be found by minimizing its posterior energy function

$$U(\omega|O) = \sum_{c \in C} V_c(\omega) - \sum_i \log P(O_i|\omega_i) \qquad (1)$$

where C is the set of cliques of related (neighbor) variables in the MRF, and $V_c(\omega)$ measures the potential (energy) of clique c under assignment ω. The posterior distribution is in terms of things we may be able to calculate or specify: clique potentials $V_c(\omega)$ (represent higher-order, prior constraints among related variables) and prior distributions for our observations $P(O_i|\omega_i)$.

By defining our constraints in terms of clique potentials and likelihoods in the MRF framework, we can formulate the search for the most likely hypotheses as a *maximum a posteriori* (MAP) estimate of the MRF. With this formulation, we can apply a MAP estimation procedure to our MRF. The result of the estimation is the set of hypotheses with the highest probability of occurring based on our prior constraints.

For our application, each variable in the MRF represents the hypothesis, (R_i, M_j), that region R_i arose from model face M_j. The variables can be labeled either **on** or **off** indicating our belief or disbelief in the hypothesis. Our observations O_i are the regions, R_i, extracted from the image.

The ith region is described by a vector of attribute values $\vec{f}_{R_i} = (f^1_{R_i}, f^2_{R_i}, ..., f^n_{R_i})$. For computational reasons, we assume that these attributes are independent for a given model face, giving us:

$$P(\vec{f}_{R_i}|M_j) = \prod_k P(f^k_{R_i}|M_j). \qquad (2)$$

We need to determine the likelihood that an image region, R_i, arose from the presence of a model face, M_j, in the scene. This is the probability of observing R_i assuming that the match hypothesis (R_i, M_j) is correct. In terms of our label set, we equate

$$P(R_i|(R_i, M_j) = ON) = P(R_i|M_j) = P(\vec{f}_{R_i}|M_j). \qquad (3)$$

To calculate the probability that R_i was observed given that hypothesis (R_i, M_j) is incorrect (R_i resulting from some other face or background), we use:

$$P(R_i \quad |(R_i, M_j) = OFF) = P(R_i|\neg M_j) = \frac{\sum_k P(R_i|M_k)P(M_k) + P(R_i|B)P(B) - P(R_i|M_j)P(M_j)}{1 - P(M_j)} \qquad (4)$$

where $P(M_j)$ is the prior probability of detecting M_j ($\sum_j P(M_j) + P(B) = 1$), and B represents the possibility of background or no label. $P(B)$ and $P(R_i|B)$ are set to constant values (here, 0.3 and 0.310^{-4}).

Equations 3 and 4 provide us with the prior probabilities of the observations ($P(O_i|\omega_i)$) required for the posterior energy function of Equation 1. Next, we need to specify the clique potentials which first requires a definition of the neighborhoods of the hypotheses (variables).

We define two neighborhood relations between pairs of hypotheses: N^+ for supporting hypotheses and N^- for contradictory hypotheses. Two hypotheses are N^- neighbors if they correspond to the same region. Hypotheses are N^+ neighbors if they are spatially and geometrically consistent with the hypothesized model. The consistency is measured by comparing the relational attributes between the regions to the relational attribute statistics of the model faces.

For efficiency concerns, we limit our energy function to 1-cliques and 2-cliques. The first order clique energies correspond to the prior probabilities, $P(M_j)$, of the hypothesized model face. When the hypothesis (R_i, M_j) is **on**, the first order clique energy is $V_{(R_i,M_j)}((R_i, M_j) = ON) = \log P(M_j)$ and when the hypothesis (R_i, M_j) is **off**, the first order clique energy is $V_{(R_i,M_j)}((R_i, M_j) = OFF) = \log(1 - P(M_j))$.

The second order clique potentials used in our experiments are: $V_{N^+}(ON, ON) = -60.0$, $V_{N^+}(OFF, OFF) = 40.0$, $V_{N^-}(ON, ON) = 60.0$, $V_{N^-}(ON, OFF) = -10.0$ (other combinations are given zero energy). For example, if a hypothesis is **on** and a consistent (N^+) neighbor hypothesis is **on**, -60.0 is the potential of that 2-clique.

We can now construct an MRF to perform the search for the most likely hypotheses. To help the reader visualize a typical resulting MRF, a simple example is shown in figure 3. In this example, the model-base contains two similar geometric models. An MRF constructed for the given model-base and an image of the first model. In this case, a hypothesis is generated for all pairs of regions R_i and model faces M_j that have nonzero conditional probabilities $P(R_i|M_j)$. The neighborhood relation of these hypotheses is simply that hypotheses for the same region are inconsistent, and hypotheses for the same model are consistent.

3.2 Hypothesis Generation: Run Time

At run time, the recognition program extracts regions from the image and computes the first-order attributes over all regions and relational attributes over all pairs of regions. Using equations 3 and 4,

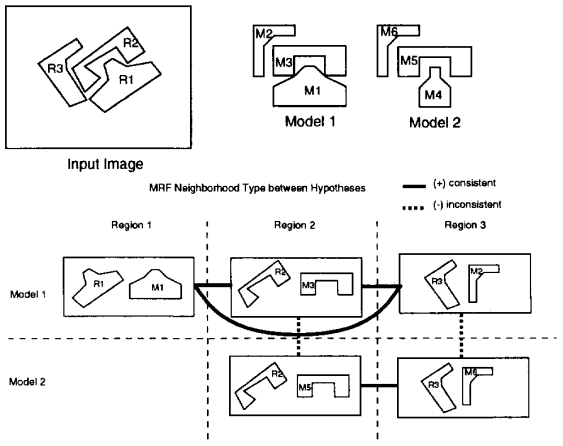

Figure 3: **An example MRF produced from a scene containing three regions with a model base containing two similar geometric objects.**

we first compute the log-likelihoods of the observation of R_i, given that hypothesis (R_i, M_m) is correct, $\log P(R_i|(R_i, M_m) = ON)$, and incorrect, $\log P(R_i|(R_i, M_m) = OFF)$ (throwing out hypotheses that have zero probability). With the remaining hypotheses, we then determine the neighborhood relations N^- and N^+. This information is then used to build the MRF.

Once the MRF is created, we wish to find the most likely set of hypotheses based on the constraints of the image. We use an estimation procedure called Highest Confidence First (HCF) [3]—chosen because of its efficiency and performance.

After HCF estimation is completed, the hypotheses labeled **on** are considered for verification. From the active hypotheses and the neighbor relation N^+, we can create a list of consistent (N^+) cliques (for this domain 3-cliques and 2-cliques are used) of matches. We order the hypothesis cliques by the the energy contribution of the constituent hypotheses to the posterior energy function (see Equation 1)—approximating the relative likelihood of the hypothesis. Thus, we are checking the most likely hypotheses first, given the image regions and our constraints.

4 Robust Localization and Verification

Given a hypothesized set of matches, we must determine where the object is (**localization**) and whether it is really present in our image or not (**ver-**

ification). Unfortunately, even slight inaccuracies of location can cause rejection of a hypothesis. In many cases, a small refinement of the location estimate may be the difference between throwing away a valid hypothesis and recognizing the object.

Several factors exacerbate the localization problem. We may not have enough constraints from our matches to determine the location of the object accurately. Inaccuracies in our region data due to noise and partial occlusion will lead to errors in location estimates. Our objects may vary slightly from the CAD models causing errors in alignment along edges and surfaces. Our initial location estimate based on our matches is assumed to be inaccurate but can serve as a good starting point for a local search for the best pose estimate.

4.1 3-D Template Matching (3DTM)

We define a parameterized template to model our object. An energy function is specified over the model parameters which relates how closely the model matches the image data. Then, we find the best model parameters by minimizing the energy function. Since we are dealing with range images, we define the 3-D template of a model to be a set of points sampled from the surface of the model. Our constraint on the templates is that visible points on the model surface match range data points in the image. The template of a rigid model is parameterized by rigid-body transformation parameters (rotation and translation).

Assuming an error distribution of $P(z) \propto e^{-\rho(z)}$ (z is the error) over the points in our model with respect to the image, we can find the MAP estimate by minimizing the energy function

$$E = \sum_{i=1}^{n} \rho(z_i). \tag{5}$$

where z_i is defined to be the error of the ith point in the model—the distance between the model point and the data point nearest the model point:

$$z_i(q) = \min_{\vec{a} \in D} \| \vec{x}_i(q) - \vec{a} \| \tag{6}$$

where D is the set of three-dimensional image points, and $\vec{x}_i(q)$ is the world coordinate of the ith model point transformed using the model parameters q.

By taking the derivative of E with respect to our model (template) parameters q, we get

$$\frac{\partial E}{\partial q} = \sum_{i=1}^{n} \psi(z_i) \frac{\partial z_i}{\partial q} \tag{7}$$

where $\psi(z) = \frac{d\rho(z)}{dz}$. To minimize E, we can use the gradient-descent update rule $\Delta q \propto -\frac{\partial E}{\partial q}$.

For our purposes, we assume that parts of the object surfaces are often occluded. Occlusion can be due to self-occlusion, nearby objects in the scene, or sensor shadows. Occluded points are considered to be outliers as are noisy points due to illumination irregularities and sensor error.

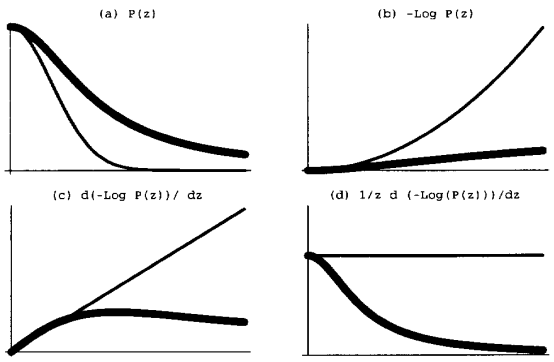

Figure 4: **Comparison of Gaussian and Lorentzian distributions and their effect on outliers. The Lorentzian is in bold. (a) the probability distributions, (b) $\rho(z)$, (c) $\psi(z)$, (d) the relative weight of the data as a function of the error magnitude.**

In this work, we use a *Lorentzian* error distribution

$$P(\frac{z}{\sigma}) \propto \frac{1}{1 + \frac{1}{2}(\frac{z}{\sigma})^2}$$

to perform the MAP estimate of our model parameters. Using

$$\rho(\frac{z}{\sigma}) = -\log P(\frac{z}{\sigma}) = \log(1 + \frac{1}{2}(\frac{z}{\sigma})^2), \tag{8}$$

we can find the MAP estimate by minimizing the energy function of Equation 5. The Lorentzian is similar in shape to the Gaussian distribution, but the tail of the distribution is much larger, indicating that outliers are assumed to occur with a relatively higher probability than in the Gaussian error model. Figure 4 compares the (unnormalized) Gaussian distribution with the Lorentzian distribution. The important graph is Figure 4(d) which shows the weighting (relative to magnitude) of the error vectors under the Gaussian and Lorentzian distributions. The effect of the Lorentzian is to significantly decrease the influence of the true outliers. This has the desirable result of more robust and accurate parameter estimation.

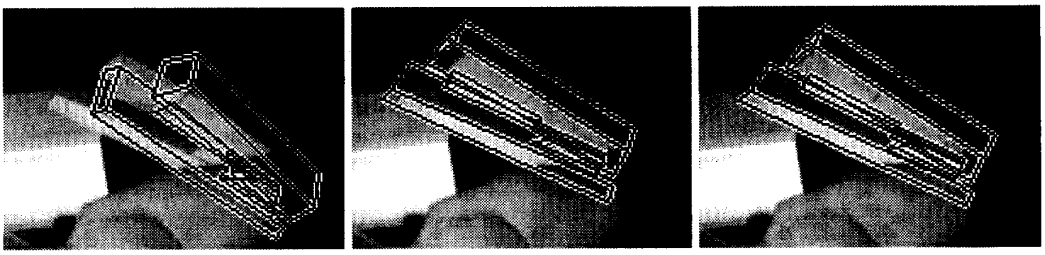

Figure 5: **Comparison of the localization results of the Gaussian and Lorentzian formulation: initial model location (left), Gaussian solution (middle), and Lorentzian solution (right).**

4.2 Verification

The hypothesis-generation phase produces a sorted list of cliques of hypothesized matches. The verification procedure must determine which of these hypotheses describe objects present in the scene. Each hypothesis is first localized (using 3DTM) then verified.

Currently, our verification metric is based on a metric proposed by Breuel [4]. Each model point is considered matched if it is within a distance σ (see Equation 8) of an image point. We define a first-order statistic,α, to be the percentage of visible model points that match an image point. A second-order statistic, β, is defined as the percentage of neighbor model points that match an image point. β measures the local consistency of the match over the model and penalizes random scattered matches compared to locally coherent matches. For the results described here, the hypothesis was accepted if $\alpha \geq 0.5$ and $\beta \geq 0.425$.

5 Experiments and Results

We first present example localization results using 3DTM to determine how accurate and robust our localization method is. This is followed by an example recognition result which demonstrates the performance of the complete recognition system.

5.1 Localization Results

To illustrate the utility of the Lorentzian distribution for accurate localization, we present an example (see Figure 5) comparing the location estimates of our method using both Gaussian and Lorentzian distributions. The example is of a partially occluded stapler. The result of the Gaussian estimate is close but has a noticeable translation error resulting from the occlusion of the back part of the stapler. This error is enough to cause a false negative verification decision for the stapler hypothesis. The Lorentzian estimate is very accurate and is not noticeably biased by the occlusion.

In this and many other tests, the least-squares solutions are noticeably occlusion sensitive while the results using the Lorentzian distribution are relatively insensitive to occlusion. The Lorentzian-based approach is able to converge to the correct location from a wider range of initial locations and is much less sensitive to occlusion than the Gaussian approach. Even when the initial pose is correct, Gaussian estimates will diverge if there is a significant number of outliers.

5.2 Recognition Results

To evaluate the performance of our object-recognition system, we are interested in how well the constraints prune the hypotheses and how well the ordering of hypotheses limits the number of verifications required. Experiments were conducted using the model-base of 8 polyhedral objects of moderate complexity.

An example of our results from a real image is shown in Figure 6. For this example, 34 planar regions were extracted from the image. The total number of possible hypotheses (correspondences to model faces) is over 5600. The constraints of our models reduce the number of hypotheses to 460. Using these hypotheses, the MRF was built and HCF was applied to select the hypotheses for verification. 242 cliques (sets of 3 or 2 consistent hypotheses) were selected—215 of which were actually tested. One reason for the high number of required verifications was that only two correspondences of the rolodex regions satisfied the first-order constraints (most likely due to modeling inaccuracies). This placed the correct hypothesis for the rolodex near the bottom of the list of hypotheses (since 3-cliques are considered more likely) and resulted in a signifi-

Figure 6: **Example recognition result (stapler and rolodex): region image (left), wire-frame overlay of recognized models (right).**

cant number of unnecessary verifications. Typically, if 3 or more correspondences are available, they are verified early in the process and reduce the number of required verifications.

After analyzing the unnecessary verifications made by our recognition system, we made a couple of observations. The first is that complex objects are easiest to recognize; the constraints of complex objects are more redundant than those of simple objects. Thus, the correct hypotheses will have more mutual support for each other and the likelihood of these hypotheses will be higher. Secondly, object symmetry increases the effective size of the model-base. If symmetry exists, hypotheses are generated for all of the symmetric combinations of features—greatly increasing the number of hypothesis cliques produced by the hypothesis-generation system.

Robust localization had a significant impact on the reliability of the verification decision. Without robust localization, it was clear that the thresholds on α and β (see Section 4.2) for accepting hypotheses would need to be so low that false positives would become very common. In our prototype recognition program, the bottleneck was the localization algorithm: it takes from 1 to 4 minutes on a SPARC IPC workstation. We have not yet spent effort on making the implementation more efficient.

6 Relation to Other Work

In this section, we compare our approach with other approaches for object recognition.

The majority of recognition techniques are fragile because they only account for model geometry when modeling the mapping between object and image features. Some techniques that avoid this involve automatic model acquisition [5, 6]. Wells [6] builds models of detectable object edges by choosing edges that appear over a variety of lighting conditions. Fan [5] uses a multi-view approach which matches a graph structure of the segmented image to stored graph structures of the representative views. Camps et al. [7] utilize an analytical model of the feature formation process. They fit a normal distribution to the data, while we prefer to use a smoothed version of the observed distribution for increased accuracy. We have found that the normal distribution is not adequate for representing the observed distributions.

Two important techniques in object recognition are invariance [8] and perceptual grouping [9]. Our statistical representation of the mapping between object and image features provides a general framework for discovering and (implicitly) representing invariants as well as bounds on the value of the invariant. The distribution of the invariant feature is represented explicitly and thus avoids the need to specify some approximating distribution (like a normal distribution) or thresholds over allowable values of the invariant. When used for matching, the statistics implicitly favor features that are truly invariant and easily detected over a wide range of views. Statistics over relations between features (here pairs of features) similarly provide a way of discovering perceptually salient or invariant groups of features. Our approach makes these "viewpoint-dependent" (view-variant) features useful for selecting matches by the hypothesis-generation procedure. The inclusion of imaging and processing effects is the essential difference between our prior models and the distributions used by [10, 11, 12].

Several researchers have investigated the idea of using statistical priors to optimize the recognition search, including [9, 7, 6, 12, 11, 10]. We have found that probabilistic selection and ordering greatly reduces the number of hypotheses that are verified during recognition. Since localization and verification is usually the most expensive step in a recognition algorithm, this has a significant impact on efficiency. By pruning the unlikely hypotheses and ordering the hypotheses in most likely first order, the number of wasted verifications is reduced significantly—resulting in more efficient performance.

The behavior of HCF search is very similar to the idea of the "focus feature" method of Bolles et al. [13]. When there are obvious matches available, HCF search dives in by turning on the most obvious match first. This creates a ripple effect for matches consistent with obvious matches. Our approach should not

be confused with the idea of relaxation labeling used by Bhanu [14] or the MAP model matching technique of Wells [15]. We are not searching for *the* most likely matching; we are searching for sets of matches that are likely enough to consider verifying.

Because of errors in location of the image features, minimal sets of correspondences often give only rough estimates of the object's pose. A favored method [9, 6] is to use all of the available image features and model features to overconstrain the pose estimate of the object. Typically, systems use a local search for the pose that minimizes the least-squared error between visible object features and image features. The techniques fail when a significant portion of object features are occluded or undetected. One common approach is outlier removal. The problem with outlier removal is that when the pose is incorrect, there may be a few incorrect matches that have small error—nullifying the effect of a majority of correct correspondences with higher error. We find it more reliable to consider all model features simultaneously while reducing the effect of the matches with high error.

7 Conclusions

We have presented new techniques for object recognition that exploit prior knowledge to make the search process efficient and the results accurate. We have described the sensor-modeling approach for generating statistical constraints for object recognition. We developed a probabilistic hypothesis-generation algorithm that uses statistical constraints to accurately select hypotheses and order them for efficient search. To ensure that our pose estimates are as accurate as possible, we designed a pose refinement algorithm that is robust to partial occlusion and other outliers. When combined into a complete system, these techniques make progress towards improving efficiency and accuracy by minimizing unnecessary verifications and improving the accuracy of the verification decision with robust location estimates.

Acknowledgments

The authors thank Martial Hebert for the use of his segmentation code, and Sing Bing Kang for providing helpful comments.

References

[1] P. Balakumar, J.-C. Robert, R. Hoffman, K. Ikeuchi, and T. Kanade, "VANTAGE: A frame-based geometric modeling system," tech. rep., Carnegie Mellon University, 1988.

[2] Y. Fujiwara, S. Nayar, and K. Ikeuchi, "Appearance simulator for computer vision research," Tech. Rep. CMU-RI-TR-91-16, Carnegie Mellon University, 1991.

[3] P. B. Chou and C. M. Brown, "The theory and practice of Bayesian image labeling," *International Journal of Computer Vision*, vol. 4, pp. 185–210, 1990.

[4] T. M. Breuel, *Geometric aspects of visual object recognition.* PhD thesis, Massachusetts Institute of Technology, 1992.

[5] T.-J. Fan, *Describing and Recognizing 3-D Objects Using Surface Properties.* New York: Springer-Verlag, 1990.

[6] W. M. Wells, *Statistical Object Recognition.* PhD thesis, Massachusetts Institute of Technology, 1992.

[7] O. I. Camps, L. G. Shapiro, and R. M. Haralick, "Premio: An overview," in *IEEE Workshop on Directions in Automated CAD-Based Vision*, IEEE, 1991.

[8] C. A. Rothwell, A. Zisserman, J. Mundy, and D. Forsyth, "Efficient model library access by projectively invariant indexing functions," in *Proceedings of Computer Vision and Pattern Recognition*, IEEE, 1992.

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

[10] J. Ben-Arie, "The probabilistic peaking effect of viewed angles and distances with application to 3-d object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 760–774, August 1990.

[11] J. B. Burns and E. M. Riseman, "Matching complex images to multiple 3d objects using view description networks," in *Proceedings of Computer Vision and Pattern Recognition*, pp. 328–334, IEEE, 1992.

[12] C. Olson, "Fast alignment by eliminating unlikely matches," Tech. Rep. UCB/CSD 92/704, University of California at Berkeley, 1992.

[13] R. C. Bolles, P. Horaud, and M. J. Hannah, "3DPO: A three-dimensional part orientation system," in *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms* (M. A. Fischler and O. Firschein, eds.), pp. 355–359, Morgan Kaufmann, 1987.

[14] B. Bhanu, "Representation and shape matching of 3-d objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 3, 1984.

[15] W. M. Wells, "Statistical object recognition with the expectation-maximization algorithm in range-derived features," in *DARPA IUS Workshop*, pp. 839–850, unknown, 1993.