Mark D. Wheeler
Research Overview

Research Interests and Philosophy

My research interests include graphics and modeling, image processing, shape estimation, stereo vision, machine learning and neural networks. My primary research focus is on the object recognition problem, specifically, *identifying and locating* objects from intensity images or range (depth) images. Object modeling is an implicit but extremely important aspect of recognition. In fact, the object recognition problem touches on many fundamental aspects of computer vision, graphics and object modeling. Through my research, I have acquired a broad understanding of the fundamental problems in these fields.

Computer vision is plagued by unfavorable combinatorics, uncertainty in measurements, and, in general, ill-posed problems. This leads to solutions with many implicit assumptions, heuristics, and implicit restrictions. My research philosophy is that a formal understanding of the problems and solutions is necessary to build systems that are robust and at the same time efficient. Once the problems, solutions, and algorithms are formalized, it is then possible to develop an improved or new solution using well-understood mathematical techniques. Furthermore, we can then characterize the applicability and performance of the algorithm. I am also a pragmatic researcher; I believe that analysis of the successes and failures of experimental systems is required to elucidate the mathematical nature of problems *in practice*.

Thesis Summary

In my thesis work, I present a paradigm for probabilistic object recognition. The paradigm is a variation of the alignment method due to Huttenlocher and Ullman, which in its most basic form is "hypothesize a minimal number of matches between image and object features, compute the position of the object, project the object into the image, and verify." In my probabilistic paradigm, each stage uses statistical representations and probabilistic criteria. In addition, I consider the object modeling problem and provide a method for automatically acquiring these representations from real and synthetic images of an object. I believe that statistical distributions and prior probabilities provide a unifying framework for representing recognition constraints on object appearance. I feel that the statistical representation coupled with probabilistic optimization methods will lead to efficient and robust recognition in practice.

Previous Research

In the initial part of my thesis research, I developed complete systems to recognize 3D polyhedral models in both range images and intensity images. I introduced the use of non-parametric statistical distributions to represent the constraints on feature appearance for selecting matches. These distributions can represent a wide range of phenomena and provide strong cues about the utility of each feature for recognition. Random sampling of synthetic views of the object was used to generate the distributions. By processing the synthetic images as we would a real input image, the constraints were an improvement over the results of previous analytic techniques. The statistical constraints led to a natural formulation of the match selection problem as *maximum a posteriori* estimation resulting in a dramatic pruning of the search space.

After much work, it was apparent that reliable verification could only be possible with accurate position estimation. Because of the limited accuracy of position estimates provided by alignment, further refinement of the position via a local search (known as *localization*) was required. I then developed a localization algorithm using techniques from robust statistics. In experiments, this algorithm demonstrated significantly better convergence than previous techniques. The localization and verification algorithms operate on the range data points and edge points and thus do not depend on any high level features. The restriction to polyhedral models is thus isolated in the matching procedure; however, a variety of more general features (such as distinguished points) are available to augment line segments and planar surfaces for purposes of matching, and, thus, we can remove the polyhedral restriction. In fact, the current work handles occluding contours and arbitrary curved surfaces.

A common criticism of feature based recognition is that data-driven feature extraction is not reliable enough for robust recognition; however, my results show that even imperfect feature extraction can be effectively used for recognition. The real problem is the use of unrealistic assumptions about feature extraction performance. In my system, features are only used as cues to generate initial hypotheses which are then fed to the data-driven localization scheme. My only assumption is that the feature's appearance follows a distribution similar to its appearance distribution over the images used to build the model.

Current Research

Analysis of the initial systems' failures revealed a serious limitation: inaccuracies in the sensor and object models were responsible for most of the recognition errors and reduced efficiency. My current research is aimed at addressing this problem by replacing the ray tracer with a real sensor—building the recognition model from real images of the object. Much of the previous work on learning/acquiring object models uses the 2D view-based approach where the model is represented by a large number of representative views. Though these techniques have shown promise within restricted domains, I believe that a 3D representation is necessary for localization and, thus, reliable verification. Accurate appearance prediction is the key to reliable localization and verification; thus, the recognition model must enable us to predict the appearance of object primitives (e.g., range data points or edge points) given the object position.

For the range image domain, a model of the object's surface is sufficient for appearance prediction. I have built upon previous techniques to merge multiple range image views of an object into a single geometric surface model. This geometric surface model is then used to build the model of the edge-point appearance for the intensity image domain. The surface model provides a mapping from the 2D edge points to 3D coordinates. I have developed techniques to merge multiple views of intensity edge images into a unified geometric edge-point representation. Current work is focused on inferring the geometric and photometric appearance constraints from the sample images. The appearance model will also be used for selecting statistically significant groups of features for feature matching.

Future Plans

In the future, I plan to extend the probabilistic recognition paradigm to use contextual indicators (e.g., the environmental conditions) and goals to drive the prior probabilities (i.e., using expectation to improve efficiency). I believe this direction of research is a key to achieving efficient recognition over a wide range of scenarios with large model bases. My thesis research only considers rigid objects (i.e., 6 degrees of freedom). Extending the techniques to apply to articulated objects such as human forms is another line of research I plan to pursue. I am familiar with machine learning and neural networks and am interested in exploring the use of techniques from these fields to aid or enhance solutions to computer vision problems. Recently, the graphics and vision communities have demonstrated simple techniques that process visual data to produce astonishing graphical effects (e.g., image mosaicing and morphing). The door is open for applications of computer vision techniques to automate these effects, and I am very anxious to collaborate with graphics researchers on these problems.