Preface

Object recognition —or, in a broader sense, scene understanding— is the ultimate scientific challenge of computer vision: After 40 years of research, robustly identifying the familiar objects (chair, person, pet), scene categories (beach, forest, office), and activity patterns (conversation, dance, picnic) depicted in family pictures, news segments, or feature films is still far beyond the capabilities of today's vision systems. On the other hand, truly successful object recognition and scene understanding technology will have a broad impact in application domains as varied as defense, entertainment, health care, human-computer interaction, image retrieval and data mining, industrial and personal robotics, manufacturing, scientific image analysis, surveillance and security, and transportation.

Although research in computer vision for recognizing 3D objects in photographs dates back to the 1960s, progress has been relatively slow and we only now see the emergence of effective techniques for recognizing object categories with different appearances under large variations in the observation conditions. While much of the early work relied almost exclusively on geometric methods, modern recognition techniques are appearance-based, in which methods from standard statistical pattern recognition are applied to image descriptors. Tremendous progress has been achieved in the past five years, thanks in large part to the integration of new data representations, such as invariant semi-local features, developed in the computer vision community with the effective models of data distribution and classification procedures developed in the statistical machine learning community.

This book exemplifies this progress. It is the outcome of two workshops that were held in Taormina in 2003 and 2004, and brought together about 40 prominent vision and machine learning researchers interested in the fundamental and applicative aspects of object recognition, as well as representatives of industry. The main goals of these two workshops were (1) to promote the creation of an international object recognition community, with common datasets and evaluation procedures, (2) to map the state of the art and identify the main open problems and opportunities for synergistic research, and (3) to articulate the industrial and societal needs and opportunities for object recognition research worldwide.

These concerns are reflected in this book. Collecting all the workshops' contributions into a single book would have been impossible. We have chosen instead to select a relatively small number of papers that illustrate the breadth of today's object recognition research and the arsenal of techniques at its disposal, and discuss current achievements and outstanding challenges.

The book is divided into five parts. Each part includes a series of chapters written by contributors to the workshops. Most of the chapters are descriptions of technical approaches, intended to capture the current state of the art. Some

of the chapters are of a tutorial nature. They cover fundamental building blocks for object recognition techniques.

Part I of the book introduces general background material on the state of object recognition research. We begin with a review of the history of the field, which sets the stage for the more recent developments reported later in the book. We then discuss the need for consistent evaluation procedures and common, challenging, datasets. This is a crucial aspect since, as the field matures, systematic evaluation of the different approaches becomes increasingly important. We conclude Part I with a discussion of the industrial needs and opportunities. As we shall see, the technology has matured to a point at which exciting applications are becoming possible.

Part II focuses on recognizing *specific* objects, an area where significant progress has occurred over the past five years. This is in part due to the advent of effective techniques for detecting and describing image patches with a controlled degree of invariance, together with efficient matching and indexing algorithms that exploit both local appearance models and powerful global geometric constraints arising from perspective imaging. As demonstrated by the five chapters making up this part of the book, reliable methods for localizing specific objects in photographs and video clips despite occlusion, clutter, and changes in viewpoint are now available.

Part III of the book attacks the difficult problem of category-level object recognition. In the methods described in these chapters, object categories are represented by collections of image patches (fixed image windows or invariant patches such as those used in Part II), potentially augmented with weak spatial layout constraints. The emphasis is on the generative or discriminative techniques used to learn the distribution of these features and their relationships, and subsequently used to classify the image instances.

Part IV investigates part-based object models that incorporate stronger structural components in the form of explicit geometric constraints, or tree-structured part assemblies for example. The emphasis there is on the definition and identification of parts as well as on efficient algorithms for detecting object instances as part assemblies in images.

Finally, Part V of the book is concerned with classifying the image pixels into object foreground vs background (as opposed to simply detecting an object instance). As shown in the chapters making up this part, this process leads to a new, well-posed view of image segmentation incorporating both bottom-up and top-down interpretation processes.

This book is a testimony to the amazing progress achieved in object recognition research in the past five years. But much remains to be done: We can now recognize a limited number of categories in constrained settings (e.g. from particular viewpoints). However, *understanding* an image or video still remains an open problem. We must also improve current datasets and evaluation criteria to avoid toy problems and to allow meaningful comparisons (see the chapter on "Datasets" in Part I, for more on this issue). Further, category-level object recognition is today essentially viewed as a statistical pattern matching prob-

lem. The emphasis is in general on the features defining the patterns and the machine learning techniques used to learn and recognize them, rather than on the representation of object, scene, and activity categories, or the integrated interpretation of the various scene elements. Future progress will require explicitly addressing the representational issues involved in object recognition and, more generally, scene understanding. Contextual issues, and hierarchical, incremental learning of a large number of categories must also be addressed. Exciting times lie ahead.

Acknowledgments. The two workshops were supported in part by the National Science Foundation under grant IIS-0335780, DARPA, the Institut National de la Recherche en Informatique et Automatique, the PASCAL European Network of Excellence IST-2002-506778, France Telecom, General Electric, Intel, Lockheed Martin, Microsoft Research, Toyota, and Xerox.

Jean Ponce
Martial Hebert
Cordelia Schmid
Andrew Zisserman