# SIGMA

## A Knowledge-Based Aerial Image Understanding System

/3.87
M634

# SIGMA
## A Knowledge-Based Aerial Image Understanding System


**TAKASHI MATSUYAMA**

*Okayama University*
*Tsushima, Okayama, Japan*

**and**

**VINCENT SHANG-SHOUQ HWANG**

*Mitre Corporation*
*McLean, Virginia*




PLENUM PRESS • NEW YORK AND LONDON

Library of Congress Cataloging-in-Publication Data

Matsuyama, Takashi, 1951-
SIGMA : a knowledge-based aerial image understanding system /
Takashi Matsuyama and Vincent Shang-Shouq Hwang.
p. cm. -- (Advances in computer vision and machine intelligence)
Includes bibliographical references.
ISBN 0-306-43301-X
1. Photographic interpretation--Data processing. 2. Image processing--Digital techniques. 3. Expert systems (Computer science) I. Hwang, Vincent Shang-Shouq. II. Title. III. Series.
TR810.M38 1990
006.3--dc20 89-29221
CIP


A Division of Plenum Publishing Corporation
233 Spring Street, New York, N.Y. 10013



Printed in the United States of America

# ADVANCES IN COMPUTER VISION AND MACHINE INTELLIGENCE

Series Editor: Martin D. Levine
*McGill University*
*Montreal, Quebec, Canada*

---

**COMPUTER VISION FOR ELECTRONICS MANUFACTURING**
L. F. Pau

**SIGMA: A Knowledge-Based Aerial Image Understanding System**
Takashi Matsuyama and Vincent Shang-Shouq Hwang

---

A Continuation Order Plan is available for this series. A continuation order will bring delivery of each new volume immediately upon publication. Volumes are billed only upon actual shipment. For further information please contact the publisher.

# Preface

It has long been a dream to realize machines with flexible visual perception capability. Research on digital image processing by computers was initiated about 30 years ago, and since then a wide variety of image processing algorithms have been devised. Using such image processing algorithms and advanced hardware technologies, many practical machines with visual recognition capability have been implemented and are used in various fields: optical character readers and design chart readers in offices, position-sensing and inspection systems in factories, computer tomography and medical X-ray and microscope examination systems in hospitals, and so on. Although these machines are useful for specific tasks, their capabilities are limited. That is, they can analyze only simple images which are recorded under very carefully adjusted photographic conditions: objects to be recognized are isolated against a uniform background and under well-controlled artificial lighting.

In the late 1970s, many image understanding systems were developed to study the automatic interpretation of complex natural scenes. They introduced artificial intelligence techniques to represent the knowledge about scenes and to realize flexible control structures. The first author developed an automatic aerial photograph interpretation system based on the blackboard model (Naga1980). Although these systems could analyze fairly complex scenes, their capabilities were still limited; the types of recognizable objects were limited and various recognition errors occurred due to noise and the imperfection of segmentation algorithms.

In the 1980s, many vision researchers studied the computational aspects of visual information processing, such as methods for recovering three-dimensional information from two-dimensional images (e.g., stereo vision), and a few image understanding systems emphasizing so-called high-level vision were developed.

In artificial intelligence, the 1980s were a very fruitful period. Many practical expert systems were developed in various fields, including medical diagnosis, troubleshooting of electronic circuits, design of computer systems, and control of energy plants. Recently, many software tools to develop knowledge-based systems and new ideas have been proposed to cope with various problems encountered in the development of such practical systems.

We believe that this is an appropriate time to study extensively the problem of high-level vision:

1. Although many computational vision algorithms have been developed, all have advantages and disadvantages, so that a combination of them suggests a way to realize flexible and robust analysis. Such combination requires control by a high-level vision module.
2. Reasoning about time and geometric space and reasoning with uncertainty (or erroneous information) are crucial problems in achieving practical knowledge-based systems as well as image understanding systems. Thus the study of high-level vision contributes not only to image understanding but also to a wide spectrum of other research fields.
3. Recently neural networks have attracted many researchers in computer science. They can be considered as an alternative approach to the problem of intelligence: neurocomputing is based on pattern-directed (distributed) representation, while ordinary artificial intelligence uses symbolic representation. Since we can realize large-scale neural networks with modern LSI technologies, more capable machines may be created. Moreover, parallel computation and reasoning is an important problem to be investigated in all areas of computer science.

This book describes results of our joint studies on image understanding over several years. We designed the overall architecture of an image understanding system, SIGMA, and implemented a prototype system in 1983–1984 while the first author was at the University of Maryland as a visiting researcher. Later, the second author improved the prototype system to complete his Ph.D. thesis (Hwan1984a), and the first author studied knowledge-based image segmentation, which is one of the major research goals underlying SIGMA, at Kyoto University and Tohoku University, Japan.

Several papers on SIGMA have been published in academic journals and conference proceedings (Hwan1984b, Mats1985, Hwan1986, Mats1986, Mats1987). In writing this book, however, we did not want to simply bundle together the contents of these papers; instead we wanted to discuss various fundamental problems in image understanding and how we can cope with them by using modern artificial intelligence techniques. We have included surveys of aerial image understanding systems, spatial reasoning methods, and expert systems for image processing so as to clarify characteristics of SIGMA. In addition, we have discussed the theoretical foundations of our spatial reasoning in terms of the first-order predicate calculus. Descriptions in terms of formal logic enabled us to formulate the essential reasoning mechanisms used in SIGMA.

Major characteristics of SIGMA can be summarized as follows:

1. *Evidence accumulation for spatial reasoning.* A new flexible reasoning method based on spatial relations among objects is proposed. It integrates both bottom-up and top-down analyses into a unified reasoning process. It has much to do with nonmonotonic reasoning in artificial intelligence, which permits reasoning with incomplete knowledge.
2. *Distributed problem solving based on object-oriented knowledge representation.* In SIGMA, all recognized objects are regarded as active reasoning agents which perform reasoning about their surrounding environments based on their own knowledge. The system coordinates such local reasoning by independent reasoning agents to construct a globally consistent description of the scene. To model active reasoning agents, object-oriented knowledge representation is used.
3. *Knowledge-based image segmentation.* Many different image analysis processes can be used to extract features from an image. The knowledge-based image segmentation module in SIGMA reasons about the most effective image analysis process based on both knowledge about image processing techniques and the quality of the image under analysis. (We call such systems *expert systems for image processing.*) Guided by the knowledge, it performs trial-and-error image analysis automatically to extract features specified in a given goal.

It would be interesting to compare SIGMA with our former image understanding system based on the blackboard model (Naga1980); SIGMA emphasizes spatial reasoning and top-down goal-directed image segmentation, while the former system incorporated sophisticated image analysis procedures to recognize objects based on their various spectral (i.e., color) and spatial properties.

The system as implemented analyzes black-and-white aerial photographs of suburban housing developments and locates cultural structures such as houses, roads, and driveways. Thus, the book will be interesting and valuable to those who are engaged in remote sensing as well as in image processing, image understanding, and artificial intelligence. We hope that this book contributes to the further development of these research areas.

Takashi Matsuyama
Vincent Shang-Shouq Hwang

*Okayama and McLean*

# Acknowledgments

We are grateful to Professors Azriel Rosenfeld and Larry Davis of the University of Maryland for their support of this research and for constructive comments on technical issues. The first author also thanks Professor Makoto Nagao of Kyoto University and Professor Takayasu Ito of Tohoku University for giving him this research opportunity and for their continual encouragement and advice.

Since this research was jointly performed at the Computer Vision Laboratory of the University of Maryland, the Department of Electrical Engineering of Kyoto University, and the Department of Information Engineering of Tohoku University, many students and members of the technical staffs of these universities helped us to implement the system and conduct experiments. We would like to thank all of them for their assistance.

# Contents

# Chapter 1

# Introduction

The term *image understanding* has been widely used since work on image understanding in the United States started in 1975. It refers to knowledge-based interpretation of visual scenes by computers and consequently denotes an interdisciplinary research area including signal processing, statistical and syntactic pattern recognition, artificial intelligence, psychology, and even neurophysiology. In the early literature, it was called *scene analysis*. *Computer vision* is also widely used to refer to a similar research area; but while computer vision emphasizes computational aspects of visual information processing, such as measurement of three-dimensional shape information by visual sensors, image understanding stresses knowledge representation and reasoning methods for scene interpretation. Although understanding time-varying scenes is a very important topic in image understanding and computer vision, in this book we confine ourselves to image understanding of static scenes.

In this chapter, we first describe, as an introduction, a general framework for image understanding in which the importance of discriminating levels of information is emphasized. Then, we provide an overview of the image understanding systems for aerial photograph interpretation that have been developed so far and discuss their limitations. Three general problems in image understanding are noted: (1) unreliable segmentation, (2) representation of and reasoning based on geometric information, and (3) reasoning with incomplete information.

This discussion is followed by surveys of various approaches to spatial reasoning and of control structures in image understanding. In the last section, we outline our fundamental ideas to solve the above problems, based on which we designed our image understanding system, SIGMA.

## 1.1. FRAMEWORK FOR IMAGE UNDERSTANDING

A primary objective of image understanding systems (IUSs) is to construct a symbolic description of the scene depicted in an image. In contrast, image processing transforms one image into another and pattern recognition classifies and labels objects represented by feature vectors. IUSs analyze an image or images to *interpret* the scene in terms of the object models given to the IUSs as knowledge about the world. Here *interpretation* refers to the correspondence (i.e., mapping) between the description of the scene and the structure of the image (Fig. 1.1). It associates objects in the scene (e.g., houses, roads) with *image features* in the image (e.g., points, lines, regions). Once the description of the scene has been constructed, computer systems can answer various queries about the scene (e.g., "How many houses exist in the scene?") and can perform physical operations by controlling robot manipulators (e.g., pick up and move physical objects). It is in this sense that we can say IUSs *understand* the scene.

IUSs consider a scene as a heterogeneous structure composed of mutually related objects of different kinds, while object detection (recognition) systems dichotomize the world into the *target object* and the *background*. Since the objective of the latter systems is to extract and recognize specific target objects, most of the knowledge used is confined to that about their intrinsic properties (e.g., color, shape) and little knowledge about the entire world is used. However, this dichotomy is valid only in *artificial* environments, for example, an isolated industrial part on a conveyer of uniform color. In reality, when we look around, we always see the scene jumbled with various objects: tables, chairs, lighting devices, coffee cups, and so on.

Most *natural* scenes are composed of objects of various kinds which are related to each other through their *functions*: coffee cups are on tables, chairs are near tables, table tops are parallel to the floor, and so on. Thus, to understand the scene, we need knowledge about (spatial) relations between objects as well as knowledge about their intrinsic properties. Since some objects may be occluded by others and only partial information about mutual relations plays an especially important role in understanding natural scenes. Using knowledge about relations,

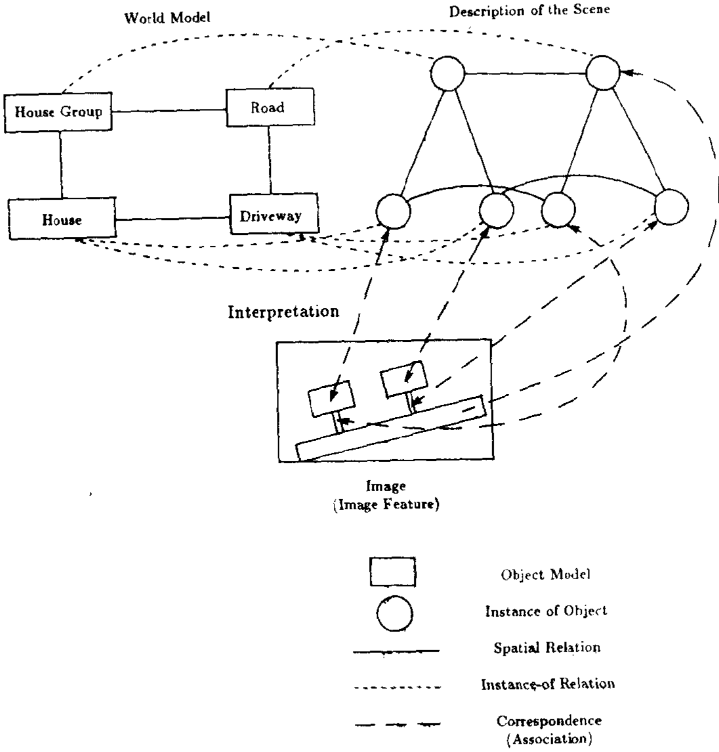


**Figure 1.1.** Interpretation in image understanding.

IUSs conduct reasoning about the structure of the scene. This clearly discriminates image understanding from statistical pattern recognition, in which all processing is done based on *attributes* of objects. The knowledge provided to the system, including information on the intrinsic properties of and mutual relations between objects in the world, is called the *world model* (Fig. 1.1).

In image understanding, discrimination between information about the scene and that about the image is crucial. The scene is described in terms of objects in the world, while the structure of the image is described in terms of image features. There exists a large gap between these two levels of information: usually the scene is three-dimensional (3D) while the image is two-dimensional (2D). The task of IUSs is to fill in the gap by computation and reasoning and to construct an interpretation (i.e., an association between the information at these two levels). A major reason for the limited performance of early IUSs (Barr1971, Prep1972, Yaki1973, Tene1977) is that they tried to jump over this gap in a single step.

In modern IUSs, the information is organized at several different levels. Many analysis processes are incorporated to analyze the information at each level and to transform the information from one level to another. Figure 1.2 illustrates the levels of information used in IUSs (see also Fig. 1.1) and the analysis processes which transform the information

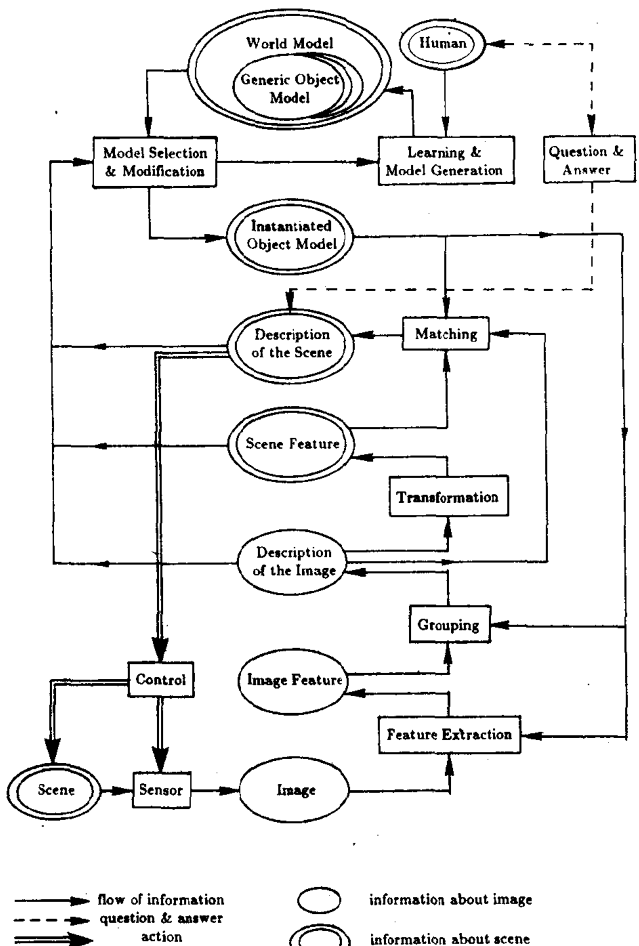


**Figure 1.2.** General architecture of image understanding systems.