G.W. AWCOCK & R. THOMAS

APPLIED IMAGE PROCESSING

Applied Image Processing

G. J. Awcock and R. Thomas

Department of Electrical and Electronic Engineering
University of Brighton

McGraw-Hill, Inc.

New York San Francisco Washington, D.C. Auckland Bogotá Caracas Lisbon London Madrid Mexico City Milan Montreal New Delhi San Juan Singapore Sydney Tokyo Toronto

1 2 3 4 5 6 7 8 9 0 DOC/DOC 9 0 0 9 8 7 6 5

ISBN 0-07 001470-1

First published 1995 by THE MACMILLAN PRESS LTD
Houndmills, Basingstoke, Hampshire RG21 2XS
and London.

Printed and bound by R. R. Donnelley & Sons Company.

Applied Image Processing

Preface

Images are a vital and integral part of everyday life. On an individual, or person-to-person basis, images are used to reason, interpret, illustrate, represent, memorise, educate, communicate, evaluate, navigate, survey, entertain, etc. We do this continuously and almost entirely without conscious effort. As man builds machines to facilitate his ever more complex lifestyle, the *only* reason for NOT providing them with the ability to exploit or transparently convey such images is a weakness of available technology.

Applied Image Processing, in its broadest and most literal interpretation, aims to address the goal of providing practical, reliable and affordable means to allow machines to cope with images while assisting man in his general endeavours.

By contrast, the term 'image processing' itself has become firmly associated with the much more limited objective of modifying images such that they are either:

(a) corrected for errors introduced during acquisition or transmission ('*restoration*'); or
(b) enhanced to overcome the weaknesses of the human visual system ('*enhancement*').

As such, the discipline of 'pure' image processing may be succinctly summarised as being concerned with

> *'a process which takes an image input and generates a modified image output'*

Clearly then, other disciplines must be allied to pure image processing in order to allow the stated goal to be achieved. 'Pattern classification', which may be defined simply as

> *'a process which takes a feature vector input and generates a class number output',*

confers the ability to identify or recognise objects and perform sorting and some inspection tasks. 'Artificial intelligence', which may be defined as

> *'a process which takes primitive data input and generates a description, or understanding or a behaviour as an output',*

confers a wide range of capability from description, in the form of simple measurement of parameters for inspection purposes, to a form of autonomy borne out of an ability to interpret the world through a visual sense.

These disciplines have been evolving steadily and independently ever since computers first became available, but only when they are all effectively harnessed together do machines acquire anything like the ability to exploit images in the way that humans do. In particular, the marriage of one, or both, of the first two disciplines with artificial intelligence has given birth to the new, image specific disciplines, namely 'image analysis', 'scene analysis' and 'image understanding'.

Image analysis is normally satisfied with quantifying data about objects which are known to exist within a scene, or determining their orientation, or recognising them as one of a limited set of possible prototypes. As such it is largely concerned with the development of descriptions of the 2-D relationships between regions within single images. However, for many applications, there is an undoubted need to extend this activity to the description of 3-D relationships between objects within a 2-D view of the real-world scene.

Scene analysis was the original term coined to describe this extension of image analysis into the third dimension. Such work flourished in the 1960s and was concerned with the rigorous visual analysis of three-dimensional polyhedra (the so-called 'blocks-world'), on the mistaken premise that it would be a trivial matter to extend these concepts to the analysis of natural scenes. The work was finally abandoned in the late 1970s when it was realised that the exploitation of application-dependent constraints was no way to research *general-purpose* vision systems.

Consequently, the term scene analysis fell into disuse only to be replaced by that of *image understanding*, which is more fundamentally based upon the physics of image formation and the operation of the human visual system. It aims to allow machines to operate with ease in complex natural environments which feature, for example, partially occluded objects or, ultimately, previously unseen objects.

A broad overview of the literature in the field of machine *perception* of images suggests the existence of two distinct 'camps' whose followers, while sharing common roots, set out to achieve fundamentally different objectives. We have chosen to label these camps as 'computer vision' and 'machine vision', and feel that they are essentially distinguished by their different approaches to the use of artificial intelligence and the degree to which it is employed. ('Robot vision' was also a popular alternative at one time, although it appears to be slowly falling into disuse, perhaps because of rather unfortunate science-fiction connotations.)

'*Computer vision*' is ultimately concerned with the goal of enabling machines to understand the world that they see, in real-time and without any form of human assistance. Thus, application-specific constraints are rejected wherever possible as the world is 'interpreted on-line'. The complexity of this task is easily under-estimated by those who take human vision for granted, but it is fraught with many immensely difficult problems, and seriously hampered by inadequate processing power.

'*Machine vision*', on the other hand, is concerned with utilising existing technology in the most effective way to endow a degree of autonomy in specific applications. The universal nature of the computer vision approach is sacrificed by deliberately exploiting application-specific constraints. Thus knowledge about the world is 'pre-compiled', or engineered, into machine vision applications in order to provide cost-effective solutions to real-world problems.

Thus one group of workers, primarily from engineering backgrounds, is application specialists, while the other group is more strongly motivated by the quest for knowledge and the desire to establish a solid research base for a 'universal' visual capability for machines. Both communities are vital for the successful development of the field, and the scope of their interests will continually converge as the performance of cost-effective computer capability improves. Clearly, the goal of producing a universal vision system which compares favourably with the human visual system is a very long way off, but progress towards that goal will continue to be absorbed in raising the level of autonomy exhibited by industrial automata etc.

However, while the labels 'computer vision' and 'machine vision' address a major sector of the field which aims to offer 'image manipulation for human advancement', they do not adequately embrace the full range of disciplines involved in meeting that aim. For example, consider the 'information technology' roles of image manipulation, such as 'document image processing' in business or 'image reconstruction' in medicine, which are critical to the effective utilisation of images in their respective application domains. Underpinning each of these and many other applications is image data compression, which is a vital part of *practical* storage and transmission of image-based information.

So, what should a book be called which aims to introduce its readership to the exciting new field of 'image manipulation for human advancement'? How should such a text be structured?

We have targeted the book at people with an engineering or general scientific background who are now in a position to *exploit* this new technology, rather than at computer scientists and AI workers who might have *developed* it. Therefore we have chosen to return to first principles and name the book *Applied Image Processing*. This deliberately uses the term 'image processing' in its broadest, colloquial sense and prefixes it with the word 'applied' to reflect the practical bias that pervades the whole book.

The book divides naturally into two parts – theory and applications – although the theory is always treated with a strongly pragmatic bias. This is reflected in the choice of industrial machine vision as a vehicle through which to investigate many of the disciplines defined above. Such an approach also allows the book to achieve a second important objective: that of providing readers with an insight into the design methodology of effective machine vision systems. This is intended to address one of the main weaknesses of the machine vision approach with respect to that of computer vision, i.e. the amount of 'bespoke' engineering that is required to realise effective solutions, coupled with the scarcity of people qualified to undertake it.

Thus the theoretical treatment is underpinned by a 'systemic' philosophy which is introduced in Chapter 1 and which aims to ensure that pragmatic and cost-effective solutions can be achieved through the well-placed application of a little forethought. This places rather uncommon emphasis upon the acquisition of good quality images (where 'good' implies *fit for purpose*), particularly through the exploitation and application of 'scene constraints' and appropriately matched image acquisition techniques (Chapters 2 and 3). Chapter 4 discusses the commonly encountered image processing techniques, but the systemic philosophy ensures that these are not used simply to compensate for poor quality image acquisition.

All image processing operations up to this point in the text have sought simply to modify the array of stored image data in order that it might better serve its intended purpose. As such, these processes are generally classified as 'low level' or pre-processing operations. 'High level' operations are concerned with the analysis, description and understanding of images, where the information representation format of an image changes from an array of numbers to symbolically meaningful strings of text. This treatment begins with discussion of segmentation of an image into meaningful regions and subsequent feature description (Chapters 5 and 6). Chapter 7 introduces the three major pattern classification strategies, including a treatment of the rapidly developing field of 'neural networks' in this context.

Until this point in the text, it is assumed that the machine is intent on deriving a two-dimensional description of the scene under investigation. While this dramatically simplifies the processing problem, and also suits an introductory text, it clearly imposes excessive constraints on many desirable applications. Thus Chapter 8 addresses the problem of helping a machine to understand and interpret a two-dimensional view of a three-dimensional world. It inevitably leans towards computer vision and includes a discussion of Marr's pivotal contribution to that field.

For undergraduate courses or self-study at any level, these first eight chapters provide a complete and thorough introduction to machine vision and through that, all the supporting disciplines. Chapter 9 uses a wide range of case studies to introduce and illustrate the breadth of application of 'image manipulation for human advancement'. It aims to fire the imagination of the readership, and inspire them to seek applications within their own sphere of influence and personal experience. It also serves as a practical introduction to many of the techniques, such as image data compression, which are not adequately addressed by the machine vision orientation of the earlier chapters.

Affordable computer performance has at last begun to become equal to the demanding nature of image processing, and the situation can only get better as the years go by. Therefore the stage is set for massive exploitation of 'Applied Image Processing'.

Acknowledgements

The authors would like to acknowledge the support of their families and the patience of their publisher, Malcolm Stewart, throughout the gestation period of this book. A particular debt of thanks is due to David Thomas for his efforts in originating many of the line drawings in the book. Thanks are also due to the companies, and individuals within them, who supplied photographs to help illustrate the book. Finally, a big thank-you to all our friends and colleagues, in particular David Lawrence, who contributed to the book through their ideas, comments and suggestions.

Contents

1 Design of Industrial Machine Vision Systems

1.1 The importance of a visual sense for machines

There seems to be little disagreement that vision is the most valuable sense that an automaton can possess. The information that it conveys is extremely rich. It can provide absolute and relative position, range, scale, orientation etc., and all this is achieved without the need for physical contact.

The term '*computer vision*' serves to associate the machine world of computing and electronics with the human attribute of vision. The 'computer' aspects of such a system are related to the hardware elements of optical sensors, parallel processing architectures, computer graphics and displays, and the software elements of data manipulation and calculation. The 'vision' aspects mirror the human visual system and encompass the functional aspects of the eye, optic nerve, and brain.

Ideally machines should be endowed with the same visual sensing capabilities as humans. In attempting to define a viable computer vision system, which emulates the essential functionality of the human system, a decision must be made on which characteristics of the human system must be, or should be, included. This may seem a deceptively trivial objective to the uninformed, but this is because humans are so good at vision that we take it for granted. In fact the human visual system is extremely complex, with many stages of processing both in the eye and in the visual cortex of the brain [1].

The distinguishing feature of higher order animate vision is the perception of images rather than their simple *sensation*. Perception, coupled with the ability to *actively* investigate a scene confers the incredible flexibility which most humans take for granted.

1.2 The role of understanding in vision

The need to emulate perception makes image understanding a vital ingredient of any computer vision system. For many years now workers in the field of computer vision have striven to develop universal vision systems which see and understand the world in much the same way as we do. Since the pivotal contribution made by Marr [1] and others in the late 1970s, worthwhile progress has been made in understanding the processes of visual perception. However, the problem with a *universal* vision system is that it must cope with operation in unconstrained environments containing objects that have never been specifically encountered before. This highly prized human attribute of 'generalisation' requires an incredible depth and breadth of knowledge and understanding about the world to be achieved. In the absence of computers featuring the sheer power and architectural elegance of the brain, the implementation of vision systems offering the versatility of biological vision must be considered a very remote possibility.

Fortunately, despite this gloomy prediction, computer vision does not have to remain totally in the realms of science fiction. Provided that we accept sensible limitations and do not become seduced by the ideal of fully emulating the human visual capabilities, useful visual perception for machines can be brought within the bounds of realistic processing power. Throughout the rest of this book, this pragmatic approach to computer vision will be distinguished from the more generalised aims of computer vision research *per se*, by use of the term 'machine vision'.

1.3 Machine vision in context

Thus the term 'machine vision' is used to describe any work which aims to provide a practical and affordable visual sense for any type of machine, which works in 'real-time'. In order to satisfy this definition it is necessary to operate in relatively constrained environments. Fortunately, the modern manufacturing environment is characterised by a high degree of order and industrial automata are generally required to perform repetitive tasks on a limited range of well-defined objects. Therefore such an environment allows exploitation and imposition of application-specific constraints and it is here where machine vision has made the most progress.

Exploitation of *a priori* knowledge about the working environment of the machine considerably eases the problem of understanding that environment. For example, an assumption can often be safely made from a single visual cue without having continually to support its validity with other cues. Reliable recognition of complex objects can often be achieved by evaluating a simple set of features which have been determined to be uniquely characteristic during an 'off-line' training process.

Using structured light to actively probe the machine's environment is much like the way a human uses a torch as a tool to investigate an unknown scene. The essential difference is that the machine must be told which tool to use and when. The pattern of light is also specially formulated and used so