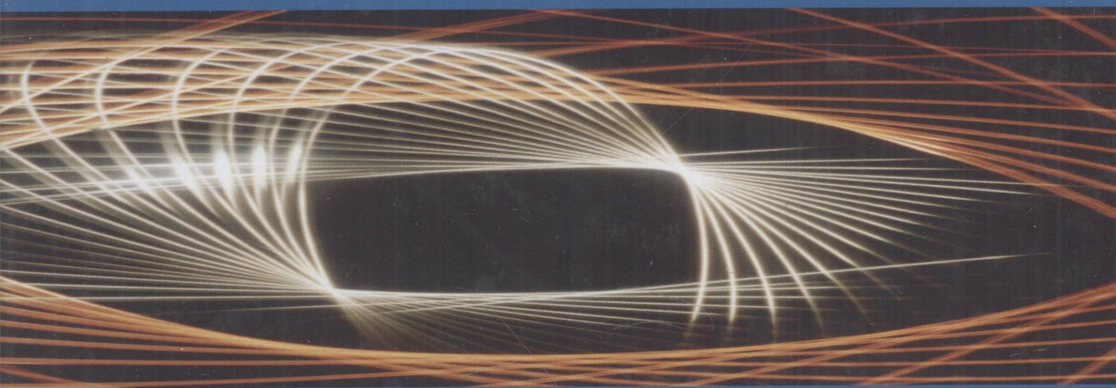
DSP

DIGITAL SIGNAL AND IMAGE PROCESSING SERIES



Visual Perception Through Video Imagery

Edited by Michel Dhome

iSTE

WILEY

Visual Perception through Video Imagery

Visual Perception through Video Imagery

Edited by
Michel Dhome

iSTE
WILEY

First published in France in 2003 by Hermes Science/Lavoisier entitled *Perception visuelle par imagerie video* © LAVOISIER, 2003

First published in Great Britain and the United States in 2009 by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Cataloging-in-Publication Data

[Perception visuelle par imagerie video. English] Visual perception through video imagery / edited by Michel Dhome.
p. cm.
Includes index.
ISBN 978-1-84821-016-5
1. Computer vision. 2. Visual perception. 3. Vision. I. Dhome, Michel.
TA1634.P4513 2007
006.3'7--dc22

2007028789

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN: 978-1-84821-016-5

Printed and bound in Great Britain by CPI Antony Rowe, Chippenham and Eastbourne.

FSC
Mixed Sources
Product group from well-managed forests and other controlled sources
Cert no. SGS-COC-2953
www.fsc.org
© 1996 Forest Stewardship Council

Table of Contents

Introduction

Artificial vision with a main objective of automatic perception and interpretation of the universe observed by a system containing one or several cameras is a relatively new field of investigation. It leads to a surprisingly large range of problems, and most of these are not currently resolved in a reliable way. Although a general theory is not close to being reached, significant progress has been made recently, theoretically as well as methodologically.

In the visible world, images of luminance are the result of two physical processes: the first one is linked to reflectance properties of the surface of observed objects, while the second one corresponds to the projection of these same objects on the light sensitive plate of the sensor used. From a mathematical standpoint, in order to interpret the observed scene, we must solve an inverse problem, i.e. infer the surface geometry (3D) of objects present, from the purely 2D content of the image or from logged images.

This reputedly complex problem in the context of computer vision is solved by man with surprising ease. However, the human vision system operation is clearly not founded on a single concept. Examining the implemented processes during short or long distance vision is sufficient proof. In the first case, the existing disparity between left and right retinal images makes it possible for man to obtain indepth information by triangulation (*stereoscopy*) relating to its close environment which is vital in particular to manually capture objects. In the second case, when looking at long distance, or even more so when contemplating a picture, stereoscopy is obviously no help in interpreting the observed scene. Even under these conditions, however (*total lack of direct 3D information*), man is able to estimate the form and spatial position of objects he observes in the vast majority of cases. This requires mental processes, from 2D information extracted from a luminance image, able to infer 3D information. These are based on the unconscious use of prior knowledge relating to the principle of retinal image composition and the form of 3D objects surrounding us. The surprising capabilities of the human vision system are because this knowledge is continuously enhanced from early childhood.

In the last few decades, researchers in the artificial vision community have attempted to develop perception systems that would work from data emanating from video cameras. This book presents a few tools emerging from recent advances in the field.

In Part 1, the reader will find three chapters dedicated to *calibration* or *self-calibration* of video sensors. Chapter 1 presents a finite estimation approach of the intrinsic parameters of a video camera, greatly inspired by the world of photogrammetry. It is based on the interpretation of images from a calibration test chart which is not generally known with great precision. Chapter 2 addresses the complex self-calibration problem from a series of matching points between different images from a single scene. The recommended method is based on an elegant and simple decomposition of Kruppa equations. Chapter 3 explores the self-calibration problem of cameras with *specific movements* making the implementation of simplified development of the main matrix possible.

Part 2 mainly involves the estimation of the *relative object/sensor position* by introducing prior knowledge (CAD object model). The reader will discover how to treat the localization problem of a rigid object observed by a monocular system. The formalism presented is then extended to understand cases as different as multi-ocular localization and hand-eye calibration, and research on the posture of articulated objects such as robotic arms.

Part 3 addresses volume information inference in two chapters. Chapter 5 discusses the *reconstruction* problem of a fixed scene observed by a multi-ocular system. The notions homo-log points, epipolar geometry, fundamental matrix, essential matrix and trifocal tensor are first introduced as well as different approaches for obtaining these entities. The problem of dense matching between image pairs is addressed before a few reconstruction examples are shown. Chapter 6 discusses the notion of *active dynamic vision*. It is the control of the path of a camera embedded in a robotic arm in order to reconstruct the surrounding scene. An underlying problem involves the definition of optimal movements necessary for the reconstruction of different primitives (points, straight lines, cylinders). Finally, perception strategies are proposed in order to ensure a complete reconstruction taking inter-object blanking into consideration.

The *recognition of forms in images* is the heart of Part 4. Chapter 7 is dedicated to proposing tools for the identification in a database of the images containing visual elements identical to those contained in a request image; the differences of acquisition between images which could involve the point of view, conditions of illumination and global composition of the scene observed. The methods presented are based on a common principle: the use of quasi-invariants associated with local descriptors. These are tested against large image databases.