CPHC

BCS

DISTINGUISHED DISSERTATIONS

David Capel

Image Mosaicing and Super-resolution

Springer

David Capel

Image Mosaicing and Super-resolution

Springer

Dr David Capel

Series Editor

Professor C.J. van Rijsbergen
Department of Computing Science, University of Glasgow, G12 8RZ, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Capel, David, 1974-
Image mosaicing and super-resolution / David Capel.
p. cm. – (Distinguished dissertations)
Includes index.
ISBN 1-85233-771-0 (alk. Paper)
1. Image processing—Digital techniques—Mathematics. 2. Resolution (Optics) 3. Matching theory. I. Title. II. CPHC/BCS distinguished dissertations.

TA1637.C475 2003
621.36'7—dc22

2003061882

Distinguished Dissertations ISSN 1439-9768

ISBN 1-85233-771-0 Springer-Verlag London Berlin Heidelberg
Springer-Verlag is a part of Springer Science+Business Media
springeronline.com

Typesetting: Electronic text files prepared by author
Printed and bound in the United States of America
34/3830-543210 Printed on acid-free paper SPIN 10945223

Distinguished Dissertations

Springer
London
Berlin
Heidelberg
New York
Hong Kong
Milan
Paris
Tokyo

Other titles published in this Series:

Search and Planning Under Incomplete Information: A Study Using Bridge Card Play
Ian Frank

Hardware Evolution: Automatic Design of Electronic Circuits in Reconfigurable Hardware by Artificial Evolution
Adrian Thompson

User-Developer Cooperation in Software Development
Eamonn O'Neill

A Combination of Geometry Theorem Proving and Nonstandard Analysis, with Application to Newton's Principia
Jacques Fleuriot

Accurate Visual Metrology from Single and Multiple Uncalibrated Images
Antonio Criminisi

Inheritance Relationships for Disciplined Software Construction
Tracy A. Gardner

Asynchronous System-on-Chip Interconnect
John Bainbridge

Stochastic Algorithms for Visual Tracking
John MacCormick

Automated Theory Formation in Pure Mathematics
Simon Colton

Dynamic Flexible Constraint Satisfaction and its Application to AI Planning
Ian Miguel

Strength or Accuracy: Credit Assignment in Learning Classifier Systems
Tim Kovacs

Image Mosaicing and Super-resolution

Abstract

This book investigates the problem of how information contained in multiple, overlapping images of the same scene may be combined to produce images of superior quality. This area, generically titled *frame fusion*, offers the possibility of reducing noise, extending the field of view, removal of moving objects, removing blur, increasing spatial resolution and improving dynamic range. As such, this research has many applications in fields as diverse as forensic image restoration, computer generated special effects, video image compression, and digital video editing.

An essential enabling step prior to performing frame fusion is *image registration*, by which an accurate estimate of the point-to-point mapping between views is computed. A robust and efficient algorithm is described to automatically register multiple images using only information contained within the images themselves. The accuracy of this method, and the statistical assumptions upon which it relies, are investigated empirically.

Two forms of frame-fusion are investigated. The first is *image mosaicing*, which is the alignment of multiple images into a single composition representing part of a 3D scene. Various methods of presenting the composite image are demonstrated, and in particular, a novel algorithm is developed for automatically choosing an optimal viewing transformation for certain cases. In addition, a new and efficient method is demonstrated for the matching of point features across multiple views.

The second frame-fusion method is *super-resolution*, which aims to restore poor-quality video sequences by removing the degradations inherent in the imaging process. The framework presented here uses a generative model of the imaging process, which is discussed in detail and an efficient implementation presented. An algorithm is developed which seeks a maximum likelihood estimate under this model, and the factors affecting its performance are investigated analytically and empirically.

The use of "generic" prior image models in a Bayesian framework is described and shown to produce dramatically improved super-resolution results. Finally, super-resolution algorithms are developed which make use of image models which are tuned to specific classes of image. These algorithms are shown to produce results of comparable or better quality than those using generic priors, while having a lower computational complexity. The technique is applied to images of text and faces.

Throughout this work, the performance of the algorithms is evaluated using real image sequences. The applications demonstrated include the separation of latent marks from cluttered, non-periodic backgrounds in forensic images; the automatic creation of full 360° panoramic mosaics; and the super-resolution restoration of various scenes, including text and faces in low-quality video.

Acknowledgements

There are a great many friends and colleagues to whom I am indebted for the parts that they have played in the production of this book. I am extremely grateful to my supervisor, Professor Andrew Zisserman. I have benefitted tremendously from his considerable experience, enthusiasm, encouragement and advice, offered freely and generously throughout the years that we have worked together. I also extend my sincere thanks to Dr. Andrew Fitzgibbon for his guidance and assistance in all areas of my research. I would like to thank my examiners, Dr. Stephen Roberts and Dr. Michal Irani, for their many helpful comments and suggestions. I am also grateful to Simon Bramble and David Compton of the Forensic Science Service for supplying the latent mark images. I would especially like to thank Beth whose love, understanding and company were a great comfort throughout the dark days of thesis-writing. Finally, and perhaps most importantly, I am forever grateful to my parents for their encouragement and support in all my endeavours.

Contents

Chapter 1

Introduction

This book investigates sets of images consisting of many overlapping views of a scene, and how the information contained within them may be combined to produce single images of superior quality. The generic name for such techniques is *frame fusion*. Using frame fusion, it is possible to extend the field of view beyond that of any single image, to reduce noise, to restore high-frequency content, and even to increase spatial resolution and dynamic range. The aim in this book is to develop efficient, robust and automated frame fusion algorithms which may be applied to real image sequences.

An essential step required to enable frame fusion is *image registration*: computing the point-to-point mapping between images in their overlapping region. This sub-problem is considered in detail, and a robust and efficient solution is proposed and its accuracy evaluated. Two forms of frame fusion are then considered: *image mosaicing* and *super-resolution*. Image mosaicing is the alignment of multiple images into a large composition which represents part of a 3D scene. Super-resolution is a more sophisticated technique which aims to restore poor-quality video sequences by modelling and removing the degradations inherent in the imaging process, such as noise, blur and spatial-sampling.

A key element in this book is the assumption of a *completely uncalibrated camera*. No prior knowledge of the camera parameters, its motion, optics or photometric characteristics is assumed. The power of the methods is illustrated with many real image sequence examples.

1.1 Background

The camera is a device which measures scene intensities and, just like any other measuring instrument, has a transfer function which introduces information loss into the measurement process. Bandwidth reduction, quantization, frequency aliasing and noise are common degradations found in imaging systems. Consequently, the images are often unable to completely capture the fine detail in a scene.

Image restoration techniques use a model of the imaging process which is "deconvolved" from the measured images in an attempt to recover the undegraded scene

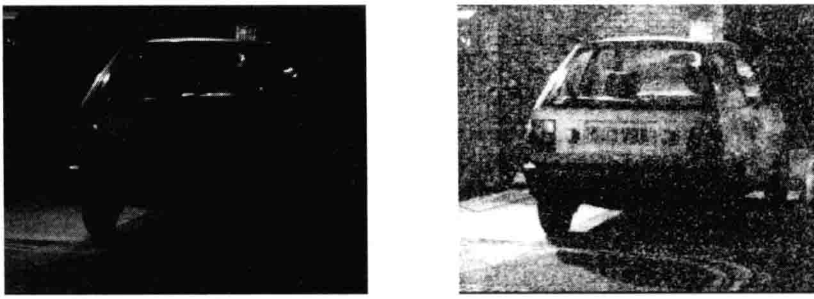

Figure 1.1: (Left) A single frame from a video sequence of a static scene taken under very poor lighting conditions. (Right) After histogram equalization. The license plate is still unreadable due to the noisy nature of the image.

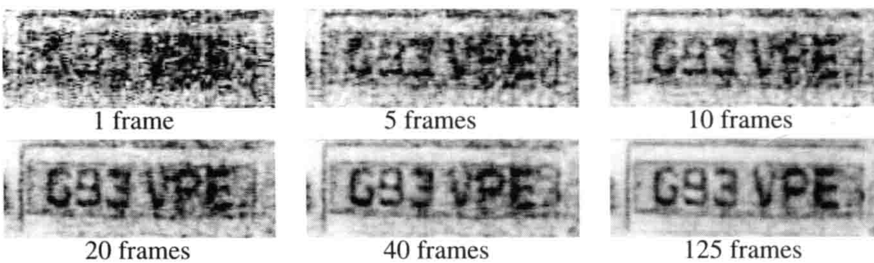

Figure 1.2: Close-up of the license plate of Figure 1.1 (histogram equalized). Each image is the result of averaging together an increasingly large number of frames. Eventually the text becomes readable.

intensities. Traditional methods have only applied to single images and are therefore fundamentally limited by the inherent loss of information in the imaging process. The restoration problem is ill-posed (under-constrained) and the techniques must rely heavily on prior assumptions about the scene intensities.

By combining information from multiple images of the same scene, the number of constraints on the reconstructed image can be greatly increased, thus improving the condition of the problem and reducing the extent to which a strong prior model of the image is required.

Possibly the simplest example of frame fusion is temporal averaging, which is an extremely effective means of reducing image noise in video sequences of static scenes. Figure 1.1 shows a frame from a video sequence of a stationary car captured under very poor lighting conditions with a fixed camera. Histogram equalization of a single frame fails to reveal the license plate due to the extremely noisy nature of the images. However, as an increasing number of frames are averaged together, the noise is reduced and the plate eventually becomes clearly readable, as shown in Figure 1.2.

The range of problems to which frame fusion is applicable is greatly increased when the single view-point constraint is removed. If multiple cameras/views are to be used, there is then the requirement that every image be accurately mapped into some global reference frame, or equivalently, that the same scene point may be accurately

Figure 1.3: (Top) 8 frames from a sequence of 25 captured by a hand-held video camera. (Bottom) All 25 frames are combined by image mosaicing. The field of view is greatly increased.

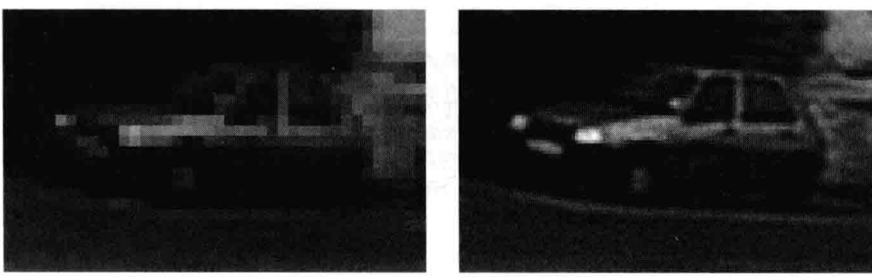

Figure 1.4: (Left) A region-of-interest in an image captured by a hand-held Mini-DV camera. (Right) A super-resolution estimate of the underlying scene combining information from 50 such images. The reconstruction is at $3\times$ the original resolution.

located in every image in which it appears. The process of computing these mappings is called *image registration*. Having registered a set of images with a common frame of reference, there are two principal forms of frame fusion which may then be applied.

Image mosaicing The images themselves may be geometrically warped and combined in a manner which both reduces noise and greatly increases the effective field of view. This is known as *image mosaicing*, and it may be used to compose tens or hundreds of images into wide-angle, panoramic views, such as that shown in Figure 1.3, which is composed from 25 images.

Super-resolution If the image registration is of high enough accuracy, the many overlapping images may be used to increase the spatial sampling density of the scene, thus allowing recovery of image frequencies above the Nyquist limit of any single image. This *super-resolution* idea is one of the main topics of this book . Figure 1.4 shows a super-resolution reconstruction which combines information from 50 images into a single, still image at $3\times$ the original resolution.

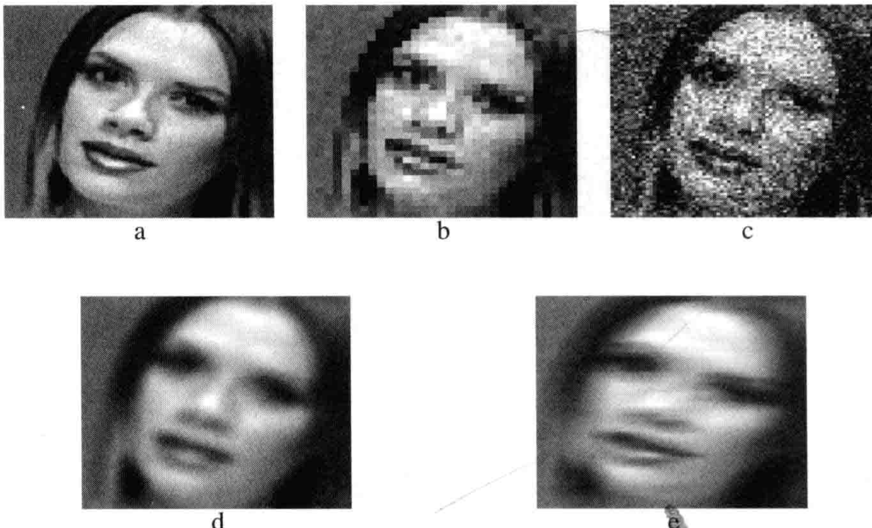

Figure 1.5: The effects of various imaging degradations. (a) An undegraded image. (b) The result of subsampling the image. (c) Noise is introduced to the intensity values. (d) The original image is subject to optical blur. (e) Motion in the scene causes anisotropic blurring.

1.2 Modelling assumptions

The information loss in camera images is due to a combination of different image degradations, each of which is characterized by a particular transfer function. The highest spatial frequency which can be captured by the camera is limited by the resolution of its imaging transducer - usually a *charge-coupled device* (CCD) array. This leads to spatial quantization of the image, which is illustrated in Figure 1.5(b). The CCD array is subject to various sources of noise, including thermal noise, shot noise, and electronic noise in the amplifier circuitry. This is particularly evident in low-lighting conditions when the camera gain is very high (see Figure 1.5(c)). When the image is digitized it also suffers intensity quantization, usually to 8 bits of precision.

It is common to consider the camera to be the familiar *pin-hole camera*, in which case everything is always in focus. In reality, of course, this is not the case and the image is therefore subject to optical blur which occurs prior to the spatial quantization. The effect of optical blur is illustrated in Figure 1.5(d). Finally, if the scene is moving, the energy entering the camera from a point in the scene will be integrated over several CCD cells during the shutter time of the camera. This causes motion blur, an example of which is shown in Figure 1.5(e).

For the purposes of image registration, this book is concerned only with images that are related by an 8 degree-of-freedom (dof) plane projective transformation (or *homography*), for which we can automatically compute the dense image-to-image correspondence. This motion model is suitable for images of a planar scene, and also for images taken by a stationary camera rotating about its optic centre.

1.3 Applications

Image mosaicing has already made an impact on the "prosumer" digital photography market, with the emergence of several products which allow a handful of photos or even a video stream from a hand-held camera to be stitched together into a wide-field mosaic, such as Peleg and Herman's VideoBrush system [112]. Mosaicing also forms the basic technology underlying node-based virtual reality systems such as QuickTime VR [1, 35] and SmoothMove [5]. The technique is also finding applications in video compression, digital steady-cam, and in the enhancement of digital video editing and matte placement software.

Super-resolution from uncalibrated video is still a relatively young field, and algorithms that are robust enough to be applied to real image data are only now beginning to emerge. The most obvious area of application is in forensic image processing. In the recent trial of the beating of Reginald Denny in the 1992 Los Angeles riots, one of the assailants was uniquely identified by a rose-shaped tattoo, enhanced in a single image using Cognitech's Video Investigator software [2]. The success of such cases had led to the rapidly increasing acceptance of digitally restored images in courtroom evidence. One fairly successful non-forensic application is the Salient Stills software [4] which attempts to generate photo-quality still images from low-quality, interlaced video streams. More recently emerging applications include resolution enhancement of mosaic images, and the generation of high-quality texture and environment maps for use in computer-generated special effects.

1.4 Principal contributions

The remaining chapters and their principal contributions are as follows.

Chapter 2: Literature survey

- A detailed survey is made of the literature pertaining to accurate image registration, image mosaicing and spatial-domain methods for uncalibrated super-resolution.

Chapter 3: Registration: geometric and photometric

- Robust methods are described for the accurate geometric and photometric registration of images, and their performance is demonstrated using real images. The accuracy of the geometric registration algorithm is investigated empirically.

- The registration methods are applied to forensic images in an application which allows latent marks to be separated from confusing, non-periodic backgrounds. The success of the method is demonstrated by several examples.

Chapter 4: Image mosaicing

- A novel algorithm is described for the efficient matching of image features across multiple views which are related by projective transformations.
- An automatic method is demonstrated for choosing an optimal viewing transformation, which aims to minimize projective distortion in the rendered mosaics.

Chapter 5: Super-resolution: Maximum likelihood and related approaches

- A detailed discussion is given of the implementation and efficiency issues regarding the generative imaging model used in super-resolution reconstruction. An effective and efficient implementation is described.
- The behaviour of the maximum-likelihood super-resolution estimator, its sensitivity to observation noise and modelling error, is explored both empirically and analytically. The method is compared in detail with Irani and Peleg's classic super-resolution algorithm [84].

Chapter 6: Super-resolution using Bayesian priors

- The use of generic image priors in a Bayesian approach to super-resolution is discussed. The performance of several such models is investigated empirically. A novel algorithm is described which uses cross-validation to determine the optimal weighting to be given to the prior in these estimators.

Chapter 7: Super-resolution using sub-space models

- A novel approach to super-resolution is described which uses sub-space image models, both in ML and MAP estimation. In some cases, the range-space of these models may be learnt from training data, and thus tuned to a particular class of image. The compactness of the model has a regularizing effect on the reconstruction problem, without imposing undesirable smoothness on the solution. The approach is applied to the enhancement of text and human faces, and is shown to be comparable or superior to methods using generic priors.

Chapter 8: Conclusions and extensions

- Several possible avenues of future research are discussed in detail.

An unfortunate feature of much of the research into uncalibrated super-resolution is the failure of some authors to present any results of their algorithms applied to real image sequences, opting instead to use purely synthetic data. In contrast, all of the algorithms presented in this book are demonstrated by application to real image sequences.