Computer Vision

计算机视觉

Edited by

Jiebo Luo

Xiaoou Tang

Dong Xu

University of Science and
Technology of China Press

当代科学技术基础理论与前沿问题研究丛书

Computer Vision

计算机视觉

Edited by

Jiebo Luo

Xiaoou Tang

Dong Xu

University of Science and
Technology of China Press

中国科学技术大学出版社

内 容 简 介

本书是由一些综述性或原始研究论文组成的，涉及了计算机视觉的各个领域.包括图像分割和标注、人脸和生物特征识别、图像配准、基于视频内容的分析和三维重建.每篇论文的作者至少有一位是中国科学技术大学信息学院的毕业生.

本书可供计算机专业高年级本科生、研究生以及相关领域的科研人员使用.

Published by University of Science and Technology of China Press
96 Jinzhai Road, Hefei, 230026, P.R. China

图书在版编目(CIP)数据

计算机视觉 = Computer Vision:英文/ 罗杰波，汤晓鸥，徐东主编. —合肥:中国科学技术大学出版社，2011.1
(当代科学技术基础理论与前沿问题研究丛书:中国科学技术大学校友文库)
"十一五"国家重点图书
ISBN 978-7-312-02750-5

Ⅰ.计… Ⅱ.①罗… ②汤… ③徐… Ⅲ.计算机—文集—英文
Ⅳ.TP302.7-53

中国版本图书馆 CIP 数据核字(2011)第 244079 号

出版发行 中国科学技术大学出版社
地址 安徽省合肥市金寨路 96 号，230026
网址 http://press.ustc.edu.cn
印　　刷 合肥晓星印刷有限责任公司
经　　销 全国新华书店
开　　本 710 mm×1000 mm 1/16
印　　张 28
字　　数 621 千
版　　次 2011 年 1 月第 1 版
印　　次 2011 年 1 月第 1 次印刷
定　　价 88.00 元

总　　序

侯建国
(中国科学技术大学校长、中国科学院院士、第三世界科学院院士)

大学最重要的功能是向社会输送人才. 大学对于一个国家、民族乃至世界的重要性和贡献度，很大程度上是通过毕业生在社会各领域所取得的成就来体现的.

中国科学技术大学建校只有短短的五十年，之所以迅速成为享有较高国际声誉的著名大学之一，主要就是因为她培养出了一大批德才兼备的优秀毕业生.他们志向高远、基础扎实、综合素质高、创新能力强，在国内外科技、经济、教育等领域做出了杰出的贡献，为中国科大赢得了“科技英才的摇篮”的美誉.

2008 年 9 月，胡锦涛总书记为中国科大建校五十周年发来贺信，信中称赞说: 半个世纪以来，中国科学技术大学依托中国科学院，按照全院办校、所系结合的方针，弘扬红专并进、理实交融的校风，努力推进教学和科研工作的改革创新，为党和国家培养了一大批科技人才，取得了一系列具有世界先进水平的原创性科技成果，为推动我国科教事业发展和社会主义现代化建设做出了重要贡献.

据统计，中国科大迄今已毕业的 5 万人中，已有 42 人当选中国科学院和中国工程院院士，是同期（自 1963 年以来）毕业生中当选院士数最多的高校之一. 其中，本科毕业生中平均每 1000 人就产生 1 名院士和七百多名硕士、博士，比例位居全国高校之首. 还有众多的中青年才俊成为我国科技、企业、教育等领域的领军人物和骨干. 在历年评选的“中国青年五四奖章”获得者中，作为科技界、科技创新型企业界青年才俊代表，科大毕业生已连续多年榜上有名，获奖总人数位居全国高校前列. 鲜为人知的是，有数千名优秀毕业生踏上国防战线，为科技强军做出了重要贡献，涌现出二十多名科技将军和一大批国防科技中坚.

为反映中国科大五十年来人才培养成果，展示毕业生在科学研究中的最新进展，学校决定在建校五十周年之际，编辑出版《中国科学技术大学校友文库》，于2008年9月起陆续出书，校庆年内集中出版50种. 该《文库》选题经过多轮严格的评审和论证，入选书稿学术水平高，已列为“十一五”国家重点图书出版规划.

入选作者中，有北京初创时期的毕业生，也有意气风发的少年班毕业生；有“两院”院士，也有IEEE Fellow；有海内外科研院所、大专院校的教授，也有金融、IT行业的英才；有默默奉献、矢志报国的科技将军，也有在国际前沿奋力拼搏的科研将才；有“文革”后留美学者中第一位担任美国大学系主任的青年教授，也有首批获得新中国博士学位的中年学者……在母校五十周年华诞之际，他们通过著书立说的独特方式，向母校献礼，其深情厚意，令人感佩!

近年来，学校组织了一系列关于中国科大办学成就、经验、理念和优良传统的总结与讨论. 通过总结与讨论，我们更清醒地认识到，中国科大这所新中国亲手创办的新型理工科大学所肩负的历史使命和责任. 我想，中国科大的创办与发展，首要的目标就是围绕国家战略需求，培养造就世界一流科学家和科技领军人才. 五十年来，我们一直遵循这一目标定位，有效地探索了科教紧密结合、培养创新人才的成功之路，取得了令人瞩目的成就，也受到社会各界的广泛赞誉.

成绩属于过去，辉煌须待开创. 在未来的发展中，我们依然要牢牢把握“育人是大学第一要务”的宗旨，在坚守优良传统的基础上，不断改革创新，提高教育教学质量，早日实现胡锦涛总书记对中国科大的期待：瞄准世界科技前沿，服务国家发展战略，创造性地做好教学和科研工作，努力办成世界一流的研究型大学，培养造就更多更好的创新人才，为夺取全面建设小康社会新胜利、开创中国特色社会主义事业新局面贡献更大力量.

是为序.

2008年9月

Preface

Computer vision is the science and technology of machines that see. As a scientific discipline, computer vision is concerned with the theory and practice for building artificial intelligence systems that extract information from visual data. The visual data can take many forms, such as a single image, a video sequence, views from multiple cameras, or multi-dimensional data from a medical scanner.

Computer vision can also be described as a complement (but not necessarily the opposite) of biological vision. In biological vision, the visual perception of humans and various animals are studied, resulting in models of how these systems operate in terms of physiological processes. Computer vision, on the other hand, studies and describes artificial vision systems that are implemented in software and/or hardware. Interdisciplinary exchange between biological and computer vision has proven increasingly fruitful for both fields.

USTC alumni have been an active and noteworthy part of the recent developments in computer vision. The objectives of this book are two-fold: (1) to provide a cross-section sampling of the diverse topics in modern computer vision, and (2) to present a cross-generation sampling of the USTC alumni working in computer vision.

There is a bigger context for this book on Computer Vision, as a part of a book series in electrical electronic engineering and computer science for promoting the national and international reputations of USTC, reporting the cutting-edge and topical research results from alumnus of USTC, and invoking a sense of belongings and excitement among the members of the greater USTC community in commemoration of its 50-th anniversary.

We received an overwhelming response when we contacted the USTC vision researchers. In the end, a total of sixteen high quality chapters from USTC alumnus were accepted to form this book. As a result, we are extremely pleased that the accepted chapters covered a wide range of topics in computer vision in terms of theories and related applications, including segmentation and registration, face and biometrics, image annotation, video analysis, as well as 3D reconstruction. A high-level overview of the chapters is included to help readers make the most of this book.

Part I Segmentation and Registration

The goal of segmentation is extraction of object boundaries, a fundamental problem in many computer vision applications. To date, this remains an open

problem in general because one has to deal with complications such as discontinuities and ambiguities in object boundaries due to appearance variations and noise. Image registration is another such problem to spatially align two or more images for comparing the difference between them or exploiting complementary information from those images, and non-rigid registration is required when irregular deformation exists in the images.

The chapter by Ning Xu and Nerendra Ahuja describes "Graph Cuts Based Active Contours"(GCBAC) for object segmentation, where the problem of obtaining an optimal object boundary is formulated in terms of energy minimization. GCBAC is a combination of the iterative deformation idea of active contours and the optimization tool of graph cuts. The resulting contour after each iteration is the global optimum within a contour neighborhood (CN) of the previous result. The use of contour neighborhood helps alleviate the bias of the minimum cut in favor of a shorter boundary. GCBAC is shown to work well for 2D objects and is expected to extend to three dimensional objects.

Pingkun Yan and Fei Wang present a chapter on "A Novel Region Constrained Non-rigid Image Registration Framework". Regions are first segmented using a hierarchical mean shift segmentation algorithm and assigned distinct labels. Next, both the images and the region labels are registered as two pairs under different transformations, subject to a unified transformation with specified properties. The proposed method exploits the benefits of both feature-based and intensity-based methods to register images, leading to more robust registration for both synthetic and real medical image data without requiring any human interactions or modification to any given underlying registration algorithm.

Part II Face and Biometrics

Biometrics is the study of the recognition of humans based on their physical or behavioral characteristics, such as face, fingerprint, voice, signature, palm, iris, gait, vein, and NDA. It has many applications in security and surveillance, and has been one of the hottest research areas in computer vision in the past decade. Face and facial expression recognition has attracted more researchers than other biometrics topics. This is because it is convenient to use without the need of those being watched to cooperate and cameras are already ubiquitous in public places. Many face recognition systems have been successfully used in controllable environments such as customs and building entrances. However, robust face recognition is still a great challenge when face images are captured from uncontrollable environments (e.g., streets) where low resolution of the

images, shadows, poor lighting, large face pose variations, and occlusions may cause a sharp drop in the performance of a system deployed in the field. New approaches that can handle these problems are still in great demand.

The chapter "Parallel Image Matrix Compression for Face Recognition" by Dong Xu, Shuicheng Yan, Lei Zhang, Zhengkai Liu, and Hong-Jiang Zhang presents a parallel image matrix compression (PIMC) algorithm for face recognition. It is extended from a 2D matrix representation of face images called 2DPCA. The authors show that 2DPCA is a localized PCA and propose a two-stage strategy to reduce the dimension of 2DPCA and improve the recognition accuracy. Comprehensive experimental results on the ORL and CMU PIE databases demonstrate that PIMC significantly outperforms 2DPCA and Eigenface, and PIMC+LDA achieves much better recognition accuracy than 2DPCA+LDA and Fisherface, especially in cases with a small number of training samples and strong pose and illumination variations.

In the chapter "Facial Expression Recognition Based on Statistical Local Features", Caifeng Shan, Shaogang Gong, and Peter W. McOwan carry out a comprehensive empirical study of facial representation for facial expression recognition based on local binary patterns (LBPs). Different machine learning methods such as SVM and AdaBoost are systematically examined using several public databases. LBP features for low-resolution facial expression recognition are also investigated, together with the evaluation of the generalization ability of the LBP features.

In the chapter "A Hierarchical Compositional Model for Face Representation and Sketching", Zijian Xu, Hong Chen, Song-Chun Zhu, and Jiebo Luo propose a hierarchical-compositional representation for modeling human faces in the form of an And-Or graph model, which simultaneously accounts for face regularity and dramatic structural variability caused by scale transitions and state transitions. Experiments show that the model helps reconstruct face images with great structural variations and rich details, and facilitates the generation of vivid cartoon sketches. The model can be used in a number of face-related applications, including recognition, non-photorealistic rendering, super-resolution, and low-bit rate coding.

The chapter "A Brief Introduction to Skeleton-based Fingerprint Minutiae Extraction" by Feng Zhao discusses how to extract fingerprint minutiae based on the skeletons of fingerprints. Zhao develops several efficient preprocessing techniques to enhance skeleton images for minutiae extraction, as well as several post-processing techniques to effectively remove

spurious minutiae. The promising performance of the proposed algorithm is demonstrated by extensive experiments with comparisons to other state-of-the-art algorithms.

Part III Image Annotation

Image annotation is an effective way for managing and retrieving the massive amount of images on the internet or personal archives. Content-based image retrieval (CBIR) was proposed to index images using visual features and to perform image retrieval based on visual similarities. However, due to the well-known semantic gap, the performance of CBIR systems is far from satisfaction. To overcome such limitations, researchers have turned attention to automatic image annotation, which can simplify the problem of image retrieval to one of text retrieval problems. Many well developed textual retrieval algorithms can then be applied to search for images by ranking the relevance between image annotations and textual queries.

Jiebo Luo, Matthew Boutell, Robert T. Gray, and Christopher Brown introduce "Image Transform Bootstrapping and Its Applications to Scene Classification". Various boosting schemes have been proposed in machine learning, focusing on the feature space. However, the performance of an exemplar-based scene classification system depends largely on the size and quality of its set of training exemplars, which are often limited in practice. The interesting concept of image-transform bootstrapping using transforms in the image space is proposed to address such issues. In particular, the authors designed three major schemes in augmenting training, testing, or both using this concept. For a number of image classification problems, appropriate transforms and meta-classification methods are combined to improve performance.

In the chapter by Xiaoguang Rui and Mingjing Li, "Bipartite Graph Reinforcement Model for Web Image Annotation", a bipartite graph reinforcement model (BGRM) is proposed for web image annotations. Given a web image, a set of candidate annotations is extracted from its surrounding text and other textual information in the hosting web page. As this set is often incomplete, it is extended to include more potentially relevant annotations by searching and mining a large-scale image database. All candidates are modeled as a bipartite graph, where a reinforcement algorithm is performed to re-rank the candidates such that only those with the highest ranking values are reserved as the final and effective annotations for real web images.

Part IV Video Analysis

Video Analysis covers a broad scope of research areas including video tracking, motion estimation, video clustering, retrieval and annotation, video event recognition, and so on, from various video sources (e.g., sports, surveillance, user-generated video, etc.). Video tracking is the process of locating moving objects from videos. The goal of motion estimation is to determine motion vectors that describe the transformations among adjacent frames of videos. Video clustering and retrieval aim to group or search semantically similar videos, and video classification deals with detecting the presence of the concepts in videos. Event recognition aims to recognize predefined events such as actions and activities or abnormal events from videos.

In "Motion Estimation Based on Trilinear and Optical Flow Constraints", Zhaohui Sun presents a novel three-frame based motion estimation scheme, in which the geometric constraint and the appearance constraint are simultaneously imposed to tackle the inherent aperture problem and over-smoothness of motion field across boundary. This method is applicable to both 2D and 3D rigid scenes with unconstrained camera motion.

Shaohua Kevin Zhou, Rama Chellappa, Zhanfeng Yue, and Baback Moghaddam present in "Appearance Modeling for Visual Tracking" a novel particle filter based tracking system by using an adaptive appearance model, an adaptive velocity motion model and an adaptive number of particles. Their method is further extended to deal with occlusion in a monocular sequence and sequences of two views. The effectiveness and robustness of their system are demonstrated for tracking different types of visual objects (e.g., car, tank and human faces) in realistic outdoor and indoor scenarios.

In the chapter "Robust Monocular 3D Tracking of Articulated Arm Movement", Gang Qian and Feng Guo propose a new 3D arm movement tracking system to cope with the challenging cases from shoulder flexion or extension and singular movement. In order to achieve persistent tracking, they incorporate a new sampling and mapping method based on inverse kinematics into the particle filter based tracking framework. The efficacy of their system is demonstrated by using real videos.

The chapter "Video Classification via Local 3D Eigen Analysis" by Jie Wei and Ze-Nian Li presents a new video classification method by analyzing the spatial/temporal properties of the eigenvalues and eigenvectors of the autocorrelation matrices for small 3D macro-blocks. Their

method effectively takes advantage of the efficacy of eigen analysis as well as the spatial and temporal correlations of videos. Promising classification results are achieved on several publicly available videos.

The chapter by Meng Wang, Tao Mei, and Xian-Sheng Hua, entitled: "Video Annotation: Supervised, Semi-Supervised and Active Learning Approaches", first groups video semantic annotation (also known as video concept detection) methods as three categories, namely, supervised, semi-supervised and active learning. After comparing Support Vector Machine (SVM) and Kernel Density Estimation (KDE), they further propose a new semi-supervised learning method as well as an active learning algorithm to alleviate the human labeling effort for collecting training data set.

Part V 3D Reconstruction

The purpose of 3D reconstruction in computer vision is to obtain the 3D shapes of objects from single or multiple 2D images. 3D reconstruction has a variety of applications in virtual reality, computer aided design, scene understanding, object recognition, advertisements, movies, and games. 3D reconstruction is often called Shape from X, where X can represent multiple images, single images, shading, texture, focus, defocus, structured light, line drawings, motion, rotation, etc.

In the chapter "Interactive 3D Modeling from a Single Image", Jin Zhou and Baoxin Li present an approach to 3D modeling from a single image with easy user interaction. Methods are given for modeling a set of 3D objects (point, plane, cone, and cylinder). Combined with simple push/pull interactions, these objects are used as primitives in 3D modeling and rendering from a single image. Both synthetic and real data experiments demonstrate the good performance of the proposed approach.

The chapter "Quasiconvex Optimization for Robust Geometric Reconstruction" by Qifa Ke and Takeo Kanade investigates geometric reconstruction problems where a cost function characterizing reprojection errors in 2D images is minimized. The authors find that the reprojection error functions share a common quasiconvex formulation. Based on the quasiconvexity, they present a quasiconvex optimization framework to formulate the problems as a small number of small-scale convex programs that are easy to solve. The effectiveness of the proposed algorithm is demonstrated by experiments on both synthetic and real data.

In the chapter "Deformable Structure from Motion: A Factorization Scheme", Jing Xiao proposes a factorization scheme to deal with the problem of 3D deformable structure from motion. The author first presents a linear closed-form solution to the reconstruction problem under the weak-perspective camera model, and then develops a two-step factorization algorithm under the perspective camera model, which not only solves the reconstruction problem, but also recovers varying camera parameters. The accuracy and robustness of the methods are shown quantitatively on synthetic data and qualitatively on real image sequences.

Concluding Remarks

As we look back to the first 50 years of USTC, we realize that it loosely coincides with the first 50 years of computer vision. As we look forward to the future, we also realize that just as the field of computer vision accelerates its growth tremendously in the most recent years, we expect, with full confidence, that USTC will follow a similar trajectory in this and other disciplines.

In closing, we would like to thank the authors for their momentous effort and steadfast patience as this book gradually came together. We would also like to thank Professor Changwen Chen, for his vision and encouragement. Finally, we also acknowledge the support from the colleagues at the USTC Press.

Jiebo Luo, Xiaoou Tang, Dong Xu

Contents

Part V 3D Reconstruction

Part I

Segmentation and Registration

Chapter 1

Graph Cuts Based Active Contours (GCBAC)

Ning Xu[1], Nerendra Ahuja[2]

[1] DMS Lab, Samsung Information Systems America,
3345 Michelson Dr., Suite 250, Irvine, CA 92612, USA
[2] ECE Department, University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA

Abstract

In this chapter, a graph cuts based active contours (GCBAC) approach to object segmentation is presented. Accurate extraction of object boundaries is an important problem in a wide range of computer vision applications. There are many difficulties in extracting accurate and optimal boundaries of the objects, such as discontinuities in object boundaries and noise in the data. The problem of obtaining an optimal object boundary can be naturally formulated in terms of energy minimization. However, it is difficult to define a cost function that leads to the globally optimal result, and even when a cost function is defined, the minimization problem is difficult to solve. The Graph Cuts Based Active Contours (GCBAC) approach presented in this chapter is a combination of the iterative deformation idea of active contours and the optimization tool of graph cuts. It differs from traditional active contours in that it uses graph cuts to iteratively deform the contour and its cost function is defined as the summation of edge weights on the cut. The resulting contour at each iteration is the global optimum within a contour neighborhood (CN) of the previous result. Since this iterative algorithm is shown to converge, the final contour is the global optimum within its own CN. The use of contour neighborhood alleviates the well-known bias of the minimum cut in favor of a shorter boundary. GCBAC approach easily extends to the segmentation of three and higher dimensional objects, and is suitable for interactive correction.

Keywords: object segmentation, active contours, snakes, graph cuts.