Huaping Liu · Fuchun Sun

Robotic Tactile Perception and Understanding

A Sparse Coding Method

Springer

Robotic Tactile Perception and Understanding

Huaping Liu · Fuchun Sun

Robotic Tactile Perception and Understanding

A Sparse Coding Method

Springer

Huaping Liu
Department of Computer Science
and Technology
Tsinghua University
Beijing
China

Fuchun Sun
Department of Computer Science
and Technology
Tsinghua University
Beijing
China

ISBN 978-981-10-6170-7 ISBN 978-981-10-6171-4 (eBook)
https://doi.org/10.1007/978-981-10-6171-4

Library of Congress Control Number: 2018932545

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

Robotic manipulation and grasping is one of the most challenging problems in the field of robotics. It requires the robot to have the ability to perceive and understand its environment via multimodal sensing and strategies.

Compared with visual sensing modality, human's understanding of the tactile sensing modality remains limited. It is mainly because of the complexity of the tactile signals, the restriction of the tactile perception techniques, and the lack of the available tactile data. Moreover, since tactile sensing is highly coupled with other sensory modalities, investigating its mechanism can largely improve the development of cognitive science.

Recently, with the rapid development of artificial intelligence, and especially machine learning techniques, the area of robotics has revealed great advances and potential. I am pleased to see this book by Huaping and Fuchun. To the best of my knowledge, this is the first book for a comprehensive approach to tactile perception using machine learning. For the problem of tactile sensing in robotic manipulation, they have established a novel technical framework, sparse coding, and dictionary learning. With this framework, the complex tactile signals can be reconstructed as new coding vectors. The sparsity is utilized to characterize many features such as the correlations between multiple fingers and different tactile attributes. Moreover, under the proposed framework, the authors also successfully solve the heterogeneous visual–tactile sensing fusion problem.

Therefore, I believe there are mainly three contributions in this book. Firstly, it provides a comprehensive survey of tactile object recognition and of visual–tactile fusion recognition technology, together with an analysis of the different representations for tactile and visual modalities. Secondly, it systematically unravels the object attribute recognition problem in the field of robotic tactile perception and understanding. Finally, it establishes a complete machine learning approach for the multimodal sensing fusion task. This work provides a good way of solving robotic manipulation and grasping in unstructured and complex environments.

This book provides readers with an intuitive understanding and exciting applications in robotic tactile sensing. The tactile sensing promises to play a critical role in robotic manipulation. I believe this book will reveal enormous practical impact as well as scientific insights into tactile sensing research and education.

Prof. Angelo Cangelosi
University of Manchester

Preface

Intelligent service robots have great potential in various application scenarios such as home services, public health, and warehouse logistics. Robotic manipulator and dexterous finger system are two key components of service robots to perform tasks which require manipulation and grasp capability, for example, caring for the elderly, surgical operations, and space or underwater exploration.

The technical challenges of manipulation and grasp involve a number of aspects including mechanical structure, hand material, object property, environment perception, and grasp planning. Among them, environment perception of service robots brings obvious challenges to existing technologies for industrial robots used for structured environments, due to the fact that service robots usually work in more complex, dynamic, and uncertain environments. This requires the robots to perceive and understand its environment in an accurate and timely manner. Referring to humans' approaches for sensing the environment through looking, listening, tasting, smelling, touching, and then unintentionally integrate the information from all channels, it is tempting to equip service robots with various sensors.

For both humans and robots, tactile sensing is the core approach used for exploration and manipulation of objects. Unlike visual sensors, tactile sensors are capable of perceiving some physical properties (e.g., softness/hardness, texture, temperature) of an object. Incorporating tactile perception to the robots can not only simulate human perception and cognitive mechanisms but also enable robots to perform more satisfyingly at practical applications.

Furthermore, visual and tactile modalities are quite different from each other. First of all, the format, frequency, and range of perceived object information are different. Tactile sensing obtains information through constant physical contact with target object, while the visual modality can simultaneously obtain multiple different features of an object at a distance. Furthermore, some features can only be obtained by one single perceptual mode. For example, the color of an object can only be obtained visually, while the texture, hardness, and temperature of a surface are obtained through tactile sensing.

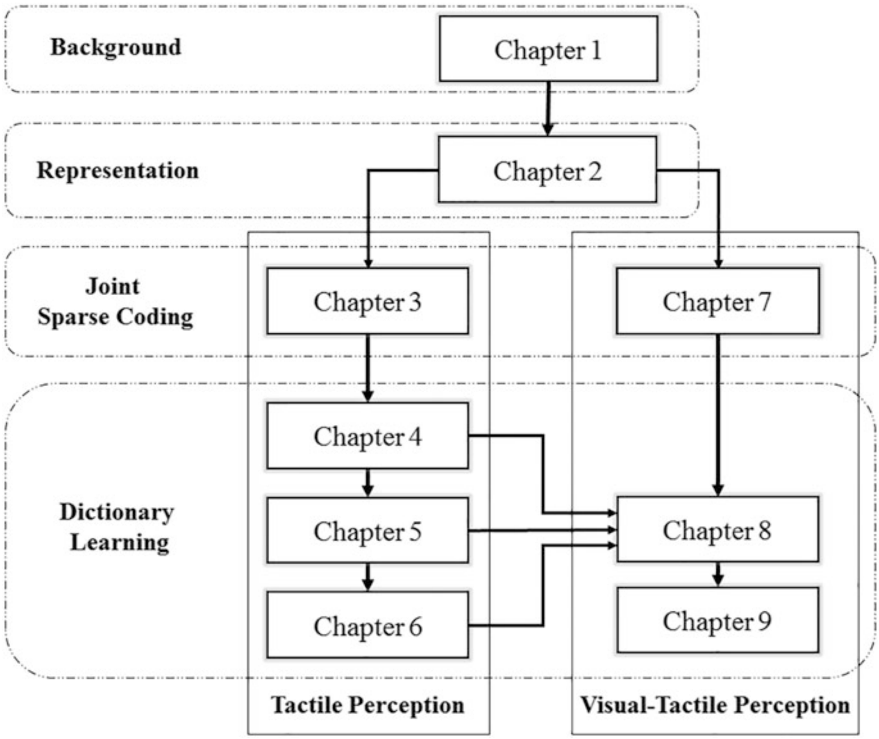

Fig. 1 Organization of the book: logical dependency among parts and chapters

To tackle those intrinsically difficult problems in tactile perception and visual–tactile fusion problems, we establish a unified sparse coding and dictionary learning framework, which forms the main contents of this book. Furthermore, a set of structured sparse coding models is developed to address the issues of dynamic tactile sensing. The book then proves that the proposed framework is effective in solving some challenging problems in the field of robotics and automation, e.g., multifinger tactile object recognition, multilabel tactile adjective recognition, and multicategory material analysis. The proposed sparse coding model can be used to tackle the challenging visual–tactile fusion recognition problem, and the book develops a series of efficient optimization algorithms to implement the model.

This book is divided into four parts. Part I presents the research background and motivation and introduces the representation and kernel of the concerned tactile and visual modalities. Part II focuses on the tactile perception problem. In Chap. 3, a joint sparse coding method for multifingered tactile fusion task is presented. In Chaps. 4–6, more complicated dictionary learning methods are developed to tackle the difficult tasks of object recognition, tactile adjective property analysis, and material identification. Part III presents more advanced applications of sparse coding and dictionary learning methodology on the heterogeneous visual–tactile fusion problems. Similarly, the joint sparse coding is firstly used to establish the basic framework to tackle the intrinsic problems in visual–tactile fusion in Chap. 7. Chapters 8 and 9 present complicated dictionary learning methods to address the material identification and cross-modal retrieval tasks. Part IV contains Chap. 10, which summarizes this book and presents some prospects. For clear illustration, an outline of the logical dependency among chapters is demonstrated in Fig. 1. Note that we try our best to make each chapter self-contained. Nevertheless, the sparse coding and dictionary learning methods developed in Chaps. 3–9 are always dependent on the kernel representation presented in Chap. 2.

This book is suitable as a reference book for graduate students with a basic knowledge of machine learning as well as professional researchers interested in robotic tactile perception and understanding, and machine learning.

Beijing, China
July 2017

Huaping Liu
Fuchun Sun

Acknowledgements

This book refers to our research work at Department of Computer Science and Technology, Tsinghua University, and State Key Laboratory of Intelligent Technology and Systems, TNLIST, China.

Five years ago, we started looking into the challenging field of robotic tactile perception. Dr. Wei Xiao conducted the first experiment for tactile data acquisition with us under very difficult conditions. With him, we launched the research work and published some preliminary results. Meanwhile, one of our visiting students, Rui Ma, who constructed a more complete tactile dataset, also published our first journal paper on this topic. About three years ago, visiting students Wen Wang and Liuyang Wang carried out the research work on dynamic time sequence classifications. This joint work established a good foundation for the development of object classification based on the tactile sequence. We would like to thank everyone who have participated for their support, dedication, and cooperation.

We would like to sincerely thank our visiting student Jingwei Yang. With her, we were able to explore the idea of using sparse coding for tactile object recognition. In 2015, we completed our first joint paper on solving the problem of tactile classification using joint sparse coding. Since then, we have gradually exploited the advantages of sparse coding in multimodal information processing and have carried out a series of research work on visual and tactile fusion. Visiting students Peng Bai and Fengxue Li also conducted a series of experiments on tactile recognition and provided important support for data acquisition and experimental verification. Our graduate and undergraduate students, Yupei Wu, Yifei Ma, Jiang Lu, and Junyi Che, helped build elegant experimental platforms for tactile perception research. Thanks for their strong support.

In terms of theoretical algorithms, we are particularly grateful to Dr. Minnan Luo and Dr. Wenbing Huang. We are extremely impressed by their strong mathematical knowledge and skills. Their work on sparse coding is a great inspiration to the authors. In addition, our Master students, Mingyi Yuan, Yulong Liu and Yunhui Liu, have successfully applied sparse coding methods to different visual fields, completed excellent Master’s theses, and provided critical support for our research methodologies. We also appreciate Hui Zhang, Tao Kong, and Yuan Yuan for their valuable help with deep learning and cyber-physic systems.

We would like to give special thanks to Dr. Bin Fang and Dr. Di Guo, for providing us with cooperation and fruitful discussions in robotic grasp and manipulation. Their deep understanding of tactile perception has provided great insights for the authors. In addition, majority of work they have done has played an important part in accomplishing this book.

In the course of our research, we have obtained a great deal of support from industrial officials. We would like to thank Mr. Liming Huang from Siemens Research, Mrs. Hong Zhang from Intel Research, Mrs. Yan-yan He from P&G Corporation, and Dr. Shan Lu from Shanghai Aerospace Control Technology Institute for their extensive support in our research of tactile perception and visual–tactile fusion.

We would like to express our sincere gratitude to Jie Qin, Yupei Wu, Haibin Fu, Xiaonan Liu, Jing Fang, Liang He, and Ge Gao who have provided immense help with preparation of figures and with the proofreading of the book.

The completion of this book cannot be separated from the strong support of Springer Publishing. We would like to thank our commissioning editors, Lanlan Chang and Jane Li, for their great support.

A great deal of this research was supported by the National Natural Science Foundation of China under the grant U1613212 for a key project "Multi-Modal Fusion Perception and Human-Robot Interaction for Robotic Manipulation," the grant 61673238 for a project "Structured Sparse Coding for Robotic Visual-Tactile Fusion Recognition." This work was also supported partially by the National Natural Science Foundation of China under Grant 91420302 and Grant 61327809.

Finally, on a personal note (HL), I would like to thank my parents and my wife for standing beside me and supporting me throughout my research and writing this book. Special thanks to my little Xinyuan (Harry). During the final stage of writing the book, I didn't have much time to spend with him. I am delighted that he learned how to swim during this period. On a personal note (FS), I would like to thank my parents, wife, and son for supporting my research and writing this book.

Beijing, China
July 2017

Huaping Liu
Fuchun Sun

Contents

Acronyms

ADMM	Alternating Direction Method of Multipliers
BioTacs	Biomimetic Tactile sensors
BPC	Blue Portable Cup
CDC	Coffee Disposable Cup
C-DL	Common Dictionary Learning
CKSC	Concatenation Kernel Sparse Coding
CMOS	Complementary Metal Oxide Semiconductor
DL	Dictionary Learning
DTW	Dynamic Time Warping
EBC	Empty Beer Can
EEG	Electroencephalogram
ELM	Extreme Learning Machine
EMB	Empty Mizone Bottle
EPB	Empty Pocari Bottle
ETC	Empty Tea Can
FBC	Full Beer Can
FMB	Full Mizone Bottle
FPB	Full Pocari Bottle
FSRs	Force Sensing Resistors
FTC	Full Tea Can
GA	Global Alignment
GCDL	Generalized Coupled Dictionary Learning
GPC	Green Portable Cup
HDC	Hard Disposable Cup
HRI	Human–Robot Interaction
ISS	International Space Station
JGKSC	Joint Group Kernel Sparse Coding
JKSC	Joint Kernel Sparse Coding
k-NN	k-Nearest Neighborhood
k-NN-T	k-Nearest Neighborhood with Tactile modality

k-NN-V	k-Nearest Neighborhood with Visual modality
K-RRSS	Kernelized Robust Representation and Structured Sparsity
KSC-T	Kernel Sparse Coding with Tactile modality
KSC-V	Kernel Sparse Coding with Visual modality
LBP	Linear Binary Pattern
LIBSVM	A Library for Support Vector Machine
LS-SVM	Least Squares Support Vector Machine
M2PDL	Multimodal Projective Dictionary pair Learning
MIS	Minimal Invasive Surgery
ML-kNN	Multi-Label kNN
NN	Nearest Neighborhood
PALM	Proximal Alternating Linearization Minimization
PDL	Projective Dictionary pair Learning
PDL-A	Projective Dictionary pair Learning-pixel Averages
PDL-D	Projective Dictionary pair Learning-Depth information
PDL-F	Projective Dictionary pair Learning-Fourier feature
PDL-Gray	Projective Dictionary pair Learning-Gray pixels
PDL-H	Projective Dictionary pair Learning-Haptic information
PDL-LBP	Projective Dictionary pair Learning-Linear Binary Pattern
PDL-RGB	Projective Dictionary pair Learning-RGB information
PDL-V	Projective Dictionary pair Learning-Visual information
PHAC	Penn Haptic Adjective Corpus
PHAC-2	Penn Haptic Adjective Corpus 2
PLC	PLastic Cup
PR2	Personal Robot2
RCovDs	Region Covariance Descriptors
RKHS	Reproducing Kernel Hilbert Space
SCDL	Semi-coupled dictionary learning
SDC	Soft Disposable Cup
S-kNN	Separate k-Nearest Neighborhood
SKSC	Separate Kernel Sparse Coding
SliM2	Supervised coupled dictionary learning with group structures for Multimodal retrieval
SO-DL	Structured Output-associated Dictionary Learning
SPAMS	SPArse Modeling Software
SR-DL	Semantics-Regularized Dictionary Learning
SVM	Support Vector Machine
TD	Toy DRagon
TDO	Toy DOll
TPA	Toy PAnda
TPE	Toy PEnguin
WMCA	Weakly paired Maximum Covariance Analysis

Mathematical Notation

M	The number of the visual training samples
N	The number of the tactile training samples
$\mathbb{T}_i$	The ith tactile training sample
$\mathbb{T}$	Testing tactile sample
$\mathbb{V}_i$	The ith visual training sample
$\mathbb{V}$	Testing visual sample
$\mathfrak{T}$	The set of tactile training samples
$\mathfrak{I}$	The set of visual training samples
$\mathscr{T}$	The manifold in which the tactile sequences lie
$\mathscr{V}$	The manifold in which the visual descriptors lie
$\mathfrak{D}$	Tactile dictionary in the space of $\mathscr{T}$
$\mathfrak{P}$	Visual dictionary in the space of $\mathscr{V}$
$\mathscr{H}_T$	Higher-dimensional (possibly infinite dimensional) inner product space for the tactile modality
$\mathscr{H}_V$	Higher-dimensional (possibly infinite dimensional) inner product space for the visual modality
$\kappa(\mathbb{T}_i, \mathbb{T}_j)$	Kernel function between tactile samples $\mathbb{T}_i$ and $\mathbb{T}_j$
$\kappa(\mathbb{V}_i, \mathbb{V}_j)$	Kernel function between visual samples $\mathbb{V}_i$ and $\mathbb{V}_j$
$\Phi(\cdot)$	Kernel-induced implicit feature mapping for tactile modality
$\Psi(\cdot)$	Kernel-induced implicit feature mapping for visual modality
$\|\boldsymbol{x}\|_0$	The number of the nonzero elements in the vector $\boldsymbol{x}$
$\|\boldsymbol{X}\|_{row-0}$	The number of the nonzero rows in the matrix $\boldsymbol{X}$
$\|\boldsymbol{x}\|_1$	The sum of the absolute values of all elements in the vector $\boldsymbol{x}$
$\|\boldsymbol{X}\|_{2,1}$	The sum of the Euclidean norms of all row vectors in the matrix $\boldsymbol{X}$
$\|\boldsymbol{x}\|_2$	The Euclidean norm of the vector $\boldsymbol{x}$
$\|\boldsymbol{X}\|_F$	The Frobenius norm of the matrix $\boldsymbol{X}$
$\|\boldsymbol{x}\|_\infty$	The maximum values of the absolute values of all elements in the vector $\boldsymbol{x}$
$\sigma(x)$	Sigmoid activation function for the scalar x
$\mathscr{E}_C$	The set of the elementary C-dimensional vectors

$\mathbf{V}$	All-one matrix with compatible dimensions
$\boldsymbol{I}$	Identity matrix with compatible dimensions
$\delta^{(c)}$	The characteristic function that selects the coefficients associated with the cth class

Part I
Background

This part of the book comprises two chapters. In Chap. 1, a survey about the tactile object recognition and visual–tactile fusion recognition technology is presented. The technical challenges for tactile perception and visual–tactile fusion understanding are also analyzed in this chapter. Chapter 2 serves as a basis of the whole book, by providing different representations for the tactile and visual modalities.

Chapter 1
Introduction

Abstract For robots, tactile perception is a key function utilized to obtain information from environment. Unlike vision sensors, tactile sensors can directly measure various physical properties of objects and the environment. Similarly, humans also use touch sensory receptors as an important approach to perceive and interact with the environment. In this chapter, a detailed discussion associated with tactile object recognition is presented. Current studies on tactile object recognition are divided into three sub-categories, and detailed analyses are provided. In addition, some advanced topics such as visual–tactile fusion, exploratory procedure, and datasets are discussed.

1.1 Robotic Manipulation and Grasp

The robotic manipulator and the dexterous finger system are most important components for service robots to perform tasks such as heavy domestic work, caring for the elderly, surgical operations, and space or underwater exploration. All of those operations require a manipulation and grasp capability, which remains a challenging problem for intelligent robots. Though many scholars have investigated related problems for several decades [6, 9, 12, 78, 86], the available robotic hand applications are still far from satisfying practical usage. This restricts development of many applications such as electronic commerce, which has benefitted from successful mobile robots. Figuratively, the problem of *Last Mile* can be solved with the outdoor mobile robots; the problem of *Last Foot* can be solved with the indoor mobile robots; and the problem of *Last Inch* must be solved with robotic manipulation and grasp technology. Recently, some major Internet companies have started promoting research on robotic manipulation and grasp. For example, Amazon held the Amazon Picking Challenge (APC)[1] in 2015 (see the left panel of Fig. 1.1[2]) at the 2015 International Conference on Robotics and Automation (ICRA) in Seattle, Washington. After that event, APC

[1] https://www.amazonrobotics.com/site/binaries/content/assets/amazonrobotics/pdfs/2015-apc-summary.pdf.

[2] This image is adopted from the website http://robohub.org/team-rbo-from-berlin-wins-amazon-picking-challenge-convincingly/.

H. Liu and F. Sun, *Robotic Tactile Perception and Understanding*,
https://doi.org/10.1007/978-981-10-6171-4_1

Fig. 1.1 LEFT: Amazon Picking Challenge which was held at ICRA2015. RIGHT: Robotic Grasping and Manipulation Competition which was held at IROS2016

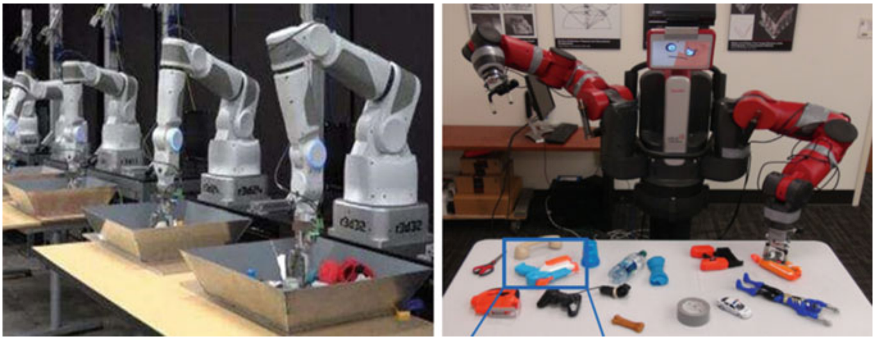

Fig. 1.2 LEFT: Google's collaborative manipulations. Copyright (2016) Sage. Reprinted, with permission, from Ref. [60]. RIGHT: Deep learning robot developed by CMU. Copyright (2016) IEEE. Reprinted, with permission, from Ref. [75]

was held at RoboCup. The goal of the APC is to strengthen ties between the industrial and academic robotic communities in order to promote shared and open solutions to some of the major problems in unstructured automation. In 2016, the Robotic Grasping and Manipulation Competition[3] was held at International Conference on Intelligent Robots and Systems (IROS) in Korea (see the right panel of Fig. 1.1). Google also reported some appealing results on collaborative learning of grasp skills (see the left panel of Fig. 1.2). MIT Technology Review reported on research in a story headlined *Deep-Learning Robot Takes 10 Days to Teach Itself to Grasp* (see the right panel of Fig. 1.2). All of these events show that robotic manipulation and grasp attract considerable attention from both industry and academia.

The technical challenges of manipulation and grasp involve a number of aspects including mechanical structure, hand material, object property, environment perception, and grasp planning. Among them, environment perception is especially difficult compared with industrial robots used for structured environments. In fact, service robots encounter complex, dynamic, and uncertain environments. This requires the robots to perceive and understand its environment in an accurate and timely manner.

[3] http://www.rhgm.org/activities/competition_iros2016/competition_iros_summary.pdf.