SpringerBriefs in Intelligent Systems

Artificial Intelligence, Multiagent Systems, and Cognitive Robotics

Series Editors

Gerhard Weiss, Maastricht University, Maastricht, The Netherlands

Karl Tuyls, University of Liverpool, Liverpool, UK; Google DeepMind, London, UK

Editorial Board

Felix Brandt, Technische Universität München, Munich, Germany

Wolfram Burgard, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany

Marco Dorigo, Université Libre de Bruxelles, Brussels, Belgium

Peter Flach, University of Bristol, Bristol, UK

Brian Gerkey, Open Source Robotics Foundation, Mountain View, CA, USA

Nicholas R. Jennings, Imperial College London, London, UK

Michael Luck, King's College London, London, UK

Simon Parsons, City University of New York, New York, NY, USA

Henri Prade, IRIT, Toulouse, France

Jeffrey S. Rosenschein, Hebrew University of Jerusalem, Jerusalem, Israel

Francesca Rossi, University of Padova, Padua, Italy

Carles Sierra, IIIA-CSIC Cerdanyola, Barcelona, Spain

Milind Tambe, University of Southern California, Los Angeles, CA, USA

Makoto Yokoo, Kyushu University, Fukuoka, Japan

This series covers the entire research and application spectrum of intelligent systems, including artificial intelligence, multiagent systems, and cognitive robotics. Typical texts for publication in the series include, but are not limited to, state-of-the-art reviews, tutorials, summaries, introductions, surveys, and in-depth case and application studies of established or emerging fields and topics in the realm of computational intelligent systems. Essays exploring philosophical and societal issues raised by intelligent systems are also very welcome.

More information about this series at https://link.springer.com/bookseries/11845

Zhongxu Hu · Chen Lv

Vision-Based Human Activity Recognition

Springer

Zhongxu Hu
School of Mechanical Aerospace Engineering
Nanyang Technological University
Singapore, Singapore

Chen Lv
School of Mechanical Aerospace Engineering
Nanyang Technological University
Singapore, Singapore

ISSN 2196-548X ISSN 2196-5498 (electronic)
SpringerBriefs in Intelligent Systems
ISBN 978-981-19-2289-3 ISBN 978-981-19-2290-9 (eBook)
https://doi.org/10.1007/978-981-19-2290-9

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Preface

In recent years, tremendous progress has been made in various intelligent devices and systems, which can be found in many living scenarios and industrial sectors, including smartphones, automated vehicles, and collaborative robots, are now part of our lives, and the proportion will become larger. Therefore, intelligent human–machine interaction which will play an inseparable role has piqued the public's considerable interest, and as a result, numerous advanced interaction methods and interfaces are being investigated to enhance user experience, acceptance, and trust. Visual sensors are now the most widely used ones due to their low-cost, high-quality, and un-intrusive characters, resulting in vision-based human activity recognition (V-HAR) becoming a critical technique for supporting downstream human–machine interaction applications. On the other hand, the deep learning approaches have achieved significant progress in many fields, which also promote V-HAR-related research. As a result, this field is flourished with a multitude of topics and techniques from various perspectives. This book provides a comprehensive overview of the past and current research studies associated with various vision-based approaches focusing on human activity recognition, especially contemporary techniques. It also sheds light on advanced application areas, and futuristic research topics. This book aims to systematically sort out these related tasks and applications and introduce the advanced deep learning methods to help further our understanding of the current state and future potential of the V-HAR research.

There are a total of 6 chapters included in this book. The chapters cover recent vision-based studies for various human activity recognition, including hand pose estimation, hand gesture recognition, head pose estimation, gaze direction estimation, gaze fixation estimation, body pose estimation, human action recognition, body reconstruction, human attention modelling, and other issues. Chapter 1 gives an overview of human activity recognition research, which introduces the background and the taxonomy of the related topics for the V-HAR. Chapter 2 focuses on the vision-based hand activity recognition, which firstly introduces the depth sensor-based hand pose estimation leveraging the deep learning approach and presents the optimization solutions from the multi-scale feature fusion and the multi-frame complementary perspectives, then an efficient dynamic hand gesture recognition

approach is discussed. Chapter 3 presents the vision-based facial activity recognition approaches, where an end-to-end head pose estimation model is firstly introduced, then a dynamic head tracking system is presented; furthermore, the appearance-based gaze direction and fixation estimation solutions are also introduced. Chapter 4 discusses vision-based body activity recognition, which includes body pose estimation, action recognition, and reconstruction. The corresponding state-of-the-art methods are reviewed and summarized. Chapter 5 introduces human attention modelling, where a context-aware approach is presented to couple with the visual saliency information of the scenarios. The conclusions and the recommendations for the V-HAR are presented in Chap. 6.

As a professional reference and research monograph, this book covers multiple popular research topics and includes cross-domain knowledge, which will benefit readers from various levels of expertise in broad fields of science and engineering, including professional researchers, graduate students, university faculties, etc. This book will help the readers to systematically study the related topics and give an overview of this field.

The completion of this book owes to not only the work of the authors but also many other individuals and groups. Special thanks would first go to all our group members of Automated Driving and Human–Machine System (AutoMan) Lab, especially Peng Hang, Yang Xing, Chao Huang, Xiangkun He, Shanhe Lou, Hao Chen, Jingda Wu, Yiran Zhang, Yanxin Zhou, Xiaoyu Mo, Tianchu Su, Wenhui Huang, and Haohan Yang, for their generous assistance. We are grateful to all members of the Rehabilitation Research Institute of Singapore (RRIS) and the Continental-NTU Corporate Lab, Nanyang Technological University, for the constant feedback and support to this work. Then, we thank Jingying Chen and Sivananth Siva Chandran in the publication team at Springer for their assistance. Finally, sincere thanks to our beloved families for their consideration as well as encouragement.

Research efforts summarized in this book were supported in part by the A*STAR Grant (No. W1925d0046) of Singapore, National Key Research, in part by the Alibaba Innovative Research Program and the Alibaba–Nanyang Technological University Joint Research Institute (No. AN-GC-2020-012), and in part by the RIE2020 Industry Alignment Fund–Industry Collaboration Projects (IAF-ICP) Funding Initiative, as well as cash and in-kind contributions from the industry partner(s).

Singapore
March 2022

Zhongxu Hu

Contents

Acronyms

2D	Two-Dimensional
3D	Three-Dimensional
AI	Artificial Intelligence
AKF	Adaptive Kalman Filter
AR	Augmented Reality
BDD-A	Berkeley Deep Drive Attention
BP	Back Propagation
CC	Correlation Coefficient
CNN	Convolutional Neural Network
CRF	Conditional Random Field
CRNN	Convolutional Recurrent Neural Network
DC	Deep Convolutional Neural Network
DL	Deep Learning
DT	Decision Tree
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyography
ERF	Effective Receptive Field
GAN	Generative Adversarial Network
GNN	Graph Neural Network
HAR	Human Activity Recognition
HCI	Human Computer Interaction
HCPS	Human–Cyber–Physical System
HOG	Histograms-Oriented Gradients
HRNet	High-Resolution Representation Network
IG	Information Gain
IMU	Inertial Measurement Unit
IR	Infrared
KD	K-Dimensional
KF	Kalman Filter
KL	Kullback–Leibler

KNN	K-Nearest Neighbors
LSH	Locality Sensitive Hashing
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MHI	Motion History Image
MLP	Multi-Layer Perceptron
MRF	Markov Random Fields
MSE	Mean Square Error
NAS	Neural Architecture Search
NeRF	Neural Radiance Field
NLP	Natural Language Processing
NN	Neural Network
NNb	Nearest Neighbor
NSFC	National Natural Science Foundation of China
NSS	Normalized Scanpath Saliency
NYU	New York University
PIFu	Pixel-aligned Implicit Function
ReLU	Rectified Linear Unit
RF	Random Forest
RGB	Red-Green-Blue
RGB-D	RGB-Depth
RNN	Recurrent Neural Network
RPN	Region Proposal Network
SAE	Society of Automotive Engineers
SIFT	Scale-Invariant Feature Transform
SMPL	Skinned Multi-Person Linear
SSC	Similarity Sensitive Coding
ST-ViT	Spatial-Temporal Vision Transformer
SURF	Speed Up Robust Features
SVM	Support Vector Machine
SVR	Support Vector Regression
T-ViT	Temporal Vision Transformer
V-HAR	Vision-based Human Activity Recognition
VR	Virtual Reality
WoS	Web of Science