
ATLANTIS AMBIENT AND PERVASIVE INTELLIGENCE

VOLUME 5

SERIES EDITOR: ISMAIL KHALIL

Atlantis Ambient and Pervasive Intelligence

Series Editor:

Ismail Khalil, Linz, Austria

(ISSN: 1875-7669)

Aims and scope of the series

The book series 'Atlantis Ambient and Pervasive Intelligence' publishes high quality titles in the fields of Pervasive Computing, Mixed Reality, Wearable Computing, Location-Aware Computing, Ambient Interfaces, Tangible Interfaces, Smart Environments, Intelligent Interfaces, Software Agents and other related fields. We welcome submission of book proposals from researchers worldwide who aim at sharing their results in this important research area.

For more information on this series and our other book series, please visit our website at:

www.atlantis-press.com/publications/books



AMSTERDAM – PARIS

© ATLANTIS PRESS

Computer Vision and Action Recognition

*A Guide for Image Processing and Computer Vision Community for
Action Understanding*

Md. Atiqur Rahman Ahad

Faculty of Engineering, Kyushu Institute of Technology,
1-1, Sensui, Tobata, Kitakyushu, Fukuoka,
T-804-0012, Japan

Applied Physics, Electronics & Communication Engineering,
University of Dhaka, Dhaka – 1000, Bangladesh



AMSTERDAM – PARIS

Atlantis Press

8, square des Bouleaux
75019 Paris, France

For information on all Atlantis Press publications, visit our website at: www.atlantis-press.com

Copyright

This book, or any parts thereof, may not be reproduced for commercial purposes in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system known or to be invented, without prior permission from the Publisher.

Atlantis Ambient and Pervasive Intelligence

Volume 1: Agent-Based Ubiquitous Computing - Eleni Mangina, Javier Carbo, José M. Molina

Volume 2: Web-Based Information Technologies and Distributed Systems - Alban Gabillon, Quan Z. Sheng, Wathiq Mansoor

Volume 3: Multicore Systems On-Chip: Practical Software/Hardware Design - Abderazek Ben Abdallah

Volume 4: Activity Recognition in Pervasive Intelligent Environments - L. Chen, C.D. Nugent, J. Biswas, J. Hoey

ISBNs

Print: 978-94-91216-19-0

E-Book: 978-94-91216-20-6

ISSN: 1875-7669

This book is dedicated to

my loving mother — Hosne Ara Begum,
my late-father — Md. Amir Hossain,
my Sensei — Prof. Seiji Ishikawa
&
my lovely wife — Shahera Hossain (Ruba)

for their continued support, love and encouragement.

Preface

The book is about computer vision and action recognition & understanding. If you have a basic knowledge on image processing and would like to study or to do research on action/activity understanding and recognition in computer vision and related arenas – then this is the book for you!

Throughout my endeavor to write this book, I have undergone many discussions with many experts during various conferences and academic meetings, discussed over emails on many challenging issues. I would like to thank them for their valuable time and inputs. Those exchanges facilitate me to build up and mature different ideas in the line of this book.

The materials presented in this book are covered in seven chapters – from basics to methodologies to challenges ahead. The fundamental goal of this book is to cover the important and related issues in such a manner – so that a student or a researcher can find ample of resources and points to go through. Chapter 1 defines action and activity. There is no clear-cut well-accepted nomenclature for atomic action to activity or behavior. This Chapter presents the key issues and I hope that a new researcher will find it very useful. Chapter 2 covers some important low-level image processing topics, which are more relevant for action recognition. If one has no prior knowledge on image processing and computer vision – this chapter can help him/her to build a background in concise manner. In Chapter 3, state-of-the-art action representation approaches are organized for better understanding on this area. It is one of the core chapters of this book that deals with representations, action recognition approaches, affective computing, action segmentation, gait recognition, and related issues.

Chapter 4 presents the motion history image method, as it is one of the most widely employed methods. However, this chapter not only presents this method and its appli-

cations, but also digs out various developments at the top of this method, from which one can retrieve essences on developmental issues from an existing approach.

Chapter 5 summarizes shape representations and pattern recognition issues in a concise manner. A reader may need to look into more details on the approaches he/she requires for further study. In Chapter 6, the datasets for action recognition are covered. This is another very important contribution of this book – because no other material is available till-to-date that encompasses the dataset issues in such comprehensive manner, as covered in this book. Finally, Chapter 7 ponders upon the challenges ahead in this important field, through plenty of new dimensions and thought-provoking discussions on present and future challenges.

After each chapter, I put some tasks under ‘Think Ahead!’ for students and researchers – so that they do some brain-storming based on the materials covered in this book. It is written in a simpler manner, with about 650 relevant citations, so that a student or a researcher can look into those references for further study on that specific area.

None is perfect! Therefore, I encourage you to write me for any comment, criticism or feedback on the book for its future development – to atiqahad@univdhaka.edu or atiqahad@yahoo.com on: What are the areas to improve in future? What topics are left out? What topics should be added? What areas need more attention?

I very much confident that this book will contribute the academia as a textbook, to run a regular course in Graduate-level or Undergraduate-level; and also guide a researcher to accomplish the work in a comprehensive manner. Let’s hope that better methodologies in future will lead us towards more human-oriented real-life applications for *better life*. With this high-toned yearning, I would like to welcome you to read this book and inform others!

Best regards,

Md. Atiqur Rahman Ahad

Kitakyushu, Fukuoka, Japan

<http://ijcvsp.com>

<http://iciev.org>

Foreword

Humans receive the great majority of information about their environment through sight. Vision is also the key component for building artificial systems that can perceive and understand their environment. Due to its numerous applications and major research challenges, computer vision is one of the most active research fields in information technology.

In recent years, recognition of human actions has become a topic of great interest in computer vision research. Among the areas of potential application are human-computer interaction, surveillance, video indexing, sport video analysis, video games and smart environments. All these applications have their own demands, but in general, algorithms to be used must be able to detect and recognize various actions in real time. As people look different and move differently, the methods should also handle variations in performing actions and work properly in various kinds of environments. In developing novel methods for different applications, it is important to use representative and publicly available test image datasets in order to be able to compare results to the state-of-the-art.

During the past few years, researchers in computer vision have developed various kinds of approaches for action recognition. In general these methods are able to recognize well relatively simple actions, but dealing with more complex actions in varying environments is still a major research issue.

This book provides an excellent overview and reference to human action recognition. After introduction to the problem area, the most common low-level image processing methods for action representations and different approaches used for action representation are presented in Chapters 2 and 3. The widely used approach based on motion history images (MHI) is described in Chapter 4, and Chapter 5 deals with different shape

representations and methods for feature vector analysis. Chapter 6 provides very valuable information about different datasets used in research on different types of actions and applications. Finally, Chapter 7 discusses what kind of challenges are ahead in this research area.

The book contains few hundreds of good references of many issues, which are important for researchers and students to look for some more details on specific areas. It is well written and easy to read also for non-native English speakers. The book also covers many research dimensions and anyone can get benefit from these points. The author has made an excellent job in writing this book. It will be a valuable resource for researchers, engineers and graduate students working in computer vision and related fields.

Matti Pietikäinen

University of Oulu, Finland

<http://www.cse.oulu.fi/MattiPietik%C3%A4inen>

Acknowledgments

All praise to *Allah*.

I express my sincere gratitude to my research supervisor, Prof. Seiji Ishikawa, Dean of Faculty of Engineering, Kyushu Institute of Technology, for his excellent guidance and motivation throughout my tenure at Kyushu Institute of Technology, Japan. His sincere and kind-heart always have been a motivating factor for my research works.

Unlimited tribute and veneration to my other teachers of my schools, Notre Dame College, University of Dhaka, Bangladesh; University of New South Wales, Australia; and Kyushu Institute of Technology, Japan.

I would like to show my sincere gratitude to Prof. Matti Pietikäinen, University of Oulu, Finland (Fellow of IAPR, Senior Member of IEEE, Associate Editor of IVC, former Associate Editor of IEEE Trans. PAMI, and Pattern Recognition) for his kind time to write a forward for this book. Various researchers assisted me through materials and discussions in different period, which helped me a lot to bring out the book. Of them, I specially mention Greg Mori (Simon Fraser University), Ramin Mehran (University of Central Florida), JK Aggarwal (The University of Texas), Jeff Cohn (University of Pittsburgh), Leonid Sigal (Brown University), and Ronald Poppe (University of Twente). I am very much thankful to Ahmed Boudissa (Kyushu Institute of Technology), Ashik Iftekhar (Kyushu Institute of Technology), Ragib Morshed (University of California Irvine) and Hafiz Imtiaz (Bangladesh University of Engineering & Technology) for their kind time to review and comment on the book.

I am thankful to many of my friends in Japan for their time and lovely interactions to make my life warm and lively. I would like to mention specially a few of them for provid-

ing me the required *mental food* through their love and care — Kenji Kurokawa, Mohin Mahtab and Keiko Matsuoka. I sorely recall my father who spent his life to guide his children for education and better lives. He is no longer here on this earth but my love, my passion, my sincere *duah* for him always. My mother strived a lot for me to bring me up to this stage and her love and pray always motivated me. Thanks to my brothers and sisters.

Finally, I must mention my wife and the best-friend Shahera Hossain (Ruba) for her constant support and her demand to do more & better in my work. She is always inspirational for my development though I failed to follow her expectations!

Md. Atiqur Rahman Ahad

Colorado Springs, USA, 2011

Contents

Preface	vii
Foreword	ix
Acknowledgments	xi
List of Figures	xix
List of Tables	xxi
1. Introduction	1
1.1 Introduction	1
1.2 What is Action?	1
1.3 Action Recognition in Computer Vision	2
1.4 Application Realms of Action Recognition	3
1.5 Categorization	4
1.6 Think Ahead!	8
2. Low-level Image Processing for Action Representations	9
2.1 Low-level Image Processing for Action Representations	9
2.2 Pre-processing Steps	9
2.3 Segmentation and Extraction	10
2.3.1 Feature Detection from an Image	11
2.3.2 Edge Detection	11
2.3.3 Corner Points	11
2.3.4 Blob Detectors	12
2.3.5 Feature Descriptors	13
2.3.6 Segmentation	14
2.4 Local Binary Pattern	20
2.4.1 LBP — Variants	21
2.5 Structure from Motion (SFM)	23
2.5.1 Constraints of FM	25
2.5.2 Improvements of FM	27
2.6 Other Issues	34
2.6.1 Intensity Normalization	34
2.6.2 Image Matching and Correspondence Problem	35
2.6.3 Camera Calibration	36
2.7 Think Ahead!	36
3. Action Representation Approaches	39
3.1 Action Representation Approaches	39

3.2	Classification of Various Dimensions of Representations	39
3.2.1	Bag-of-Features (BoF) or Bag-of-Visual-Words (BoVW)	40
3.2.2	Product Manifold Approaches	44
3.3	Action Recognition Approaches	49
3.3.1	Interest-point-based Approaches	49
3.3.2	Hidden Markov Model-based Approaches	53
3.3.3	Eigenspace-based Approach	55
3.3.4	Approaches to Manage Occlusion	56
3.3.5	Other Approaches	56
3.4	View-invariant Methods	61
3.5	Gesture Recognition and Analysis	62
3.6	Action Segmentation and Other Areas	63
3.7	Affective Computing and Expression Analysis	63
3.7.1	Games with Emotional Involvement	64
3.7.2	Interactive Arts	66
3.7.3	Anatomically-based Talking Head	66
3.8	Action Segmentation	67
3.8.1	Gestures	67
3.8.2	Basic Motion	67
3.8.3	Fundamental Gesture	68
3.8.4	Motion Alphabet	68
3.8.5	Atomic Movement	68
3.8.6	Direction-based Basic Motion	68
3.8.7	Distinct Behaviors	69
3.8.8	Motion Patterns based on Symbols	69
3.8.9	Basic Movement Transition Graph	69
3.9	Gait Analysis	69
3.10	Action Recognition in Low-resolution	71
3.10.1	Application Areas	71
3.10.2	Related Works on Low-Resolution Video Processing	72
3.11	Discussion	73
3.11.1	Salient Region and Its Associated Salient Region Construction	74
3.11.2	Biologically-inspired Visual Representations	75
3.12	Conclusion	75
3.13	Think Ahead!	75
4.	MHI – A Global-based Generic Approach	77
4.1	Motion History Image (MHI)	77
4.2	Why MHI?	78
4.3	Various Aspects of the MHI — A Tutorial	79
4.3.1	Formation of an MHI Image	79
4.3.2	Motion Energy Image (MEI)	80
4.3.3	Parameter — τ	82
4.3.4	Parameter — δ	82
4.3.5	Temporal Duration vs. Decay Parameter	82
4.3.6	Update Function $\psi(x, y, t)$	83
4.3.7	Feature Vector for the MHI	85
4.4	Constraints of the MHI Method	86
4.4.1	Self-occlusion Problem	86
4.4.2	Failure in Dynamic Background	87

4.4.3	Improper Implementation of the Update Function	87
4.4.4	Label-based Recognition	87
4.4.5	Failure with Motion Irregularities	87
4.4.6	Failure to Differentiate Similar Motions	87
4.4.7	Not View-invariant Method	87
4.4.8	Problem with Varied Motion Duration	88
4.4.9	Non-trajectory Nature	88
4.5	Developments on the MHI	88
4.5.1	Direct Implementation of the MHI	88
4.5.2	Modified MHI	91
4.6	Solutions to Some Constraints of the Basic MHI	95
4.6.1	Solutions to Motion Self-occlusion Problem	95
4.6.2	Solving Variable-length Movements	100
4.6.3	<i>timed</i> -Motion History Image	101
4.6.4	Hierarchical-MHI	101
4.6.5	Pixel Signal Energy	102
4.6.6	Pixel Change History	102
4.6.7	Motion Flow History	102
4.6.8	Contour-based STV	102
4.6.9	Solving View-Invariant Issue	103
4.7	Motion Analysis	105
4.8	Implementations of the MHI	105
4.8.1	The MHI and its Variants in Recognition	106
4.8.2	The MHI and its Variants in Analysis	106
4.8.3	The MHI and its Variants in Interactions	107
4.9	MHI and its Future	111
4.10	Conclusion	112
4.11	Think Ahead!	113
5.	Shape Representation and Feature Vector Analysis	115
5.1	Feature Points Tracking	115
5.1.1	Two-frame-based Approaches	115
5.1.2	Long-sequence-based Approaches	116
5.1.3	Discussions	116
5.2	Shape Representation Schemes	119
5.2.1	Contour-based Methods	121
5.2.2	Region-based Methods	122
5.3	Moment Invariants	122
5.3.1	Hu Moments for Feature Sets	122
5.3.2	Zernike Moments for Feature Sets	126
5.4	Component Analysis Methods	128
5.4.1	Appropriate Feature Selection	128
5.4.2	Dimension Reduction	129
5.5	Pattern Classifications	133
5.5.1	Supervised Learning	134
5.5.2	Unsupervised Learning	137
5.5.3	Nearest Neighbor	138
5.5.4	Cross-validation — Partitioning Scheme	142
5.6	Evaluation Matrices	143
5.7	Think Ahead!	145

6. Action Datasets	147
6.1 Action Datasets	147
6.2 Necessity for Standard Datasets	147
6.2.1 Motion Capture System	148
6.3 Datasets on Single-person in the View	149
6.3.1 KTH Dataset	149
6.3.2 Weizmann Dataset	150
6.3.3 IXMAS Dataset	151
6.3.4 CASIA Action Database	152
6.3.5 UMD Dataset	152
6.3.6 ICS Action Database	152
6.3.7 Korea University Gesture Database	152
6.3.8 Wearable Action Recognition Database (WARD)	153
6.3.9 Biological Motion Library (BML)	153
6.3.10 HDM05 (Hochschule der Medien) Motion Capture Database	153
6.4 Gesture Datasets	153
6.4.1 Cambridge Gesture Dataset	153
6.4.2 NATOPS Dataset	153
6.4.3 Keck Gesture Dataset	154
6.5 Datasets on Social Interactions	154
6.5.1 Youtube Dataset	154
6.5.2 Youtube Video Dataset	155
6.5.3 Hollywood2 Human Action (HOHA) Datasets	155
6.5.4 UCF Sports Dataset	155
6.5.5 Soccer Dataset	156
6.5.6 Figure-skating Dataset — Caltech Dataset	156
6.5.7 ADL — Assisted Daily Living Dataset	156
6.5.8 Kisses/Slaps Dataset	156
6.5.9 UIUC Action Dataset	157
6.6 Datasets on Other Arenas	157
6.6.1 Actions in Still Images	158
6.6.2 Nursing-home Dataset	158
6.6.3 Collective Activity Dataset	158
6.6.4 Coffee and Cigarettes Dataset	161
6.6.5 People Playing Musical Instrument (PPMI)	161
6.6.6 DARPA's Mind's Eye Program	162
6.6.7 VIRAT Video Dataset	162
6.6.8 UMN Dataset: Unusual Crowd Activity	162
6.6.9 Web Dataset	162
6.6.10 HumanEva Dataset	162
6.6.11 University of Texas Dataset	164
6.6.12 Other Datasets	164
6.7 Challenges Ahead on Datasets	165
6.8 Think Ahead!	168
7. Challenges Ahead	173
7.1 Challenges Ahead	173
7.2 Key Challenges Ahead in Action Recognition	173
7.3 Few Points for New Researchers	177

7.3.1	Intelligent Transport System	180
7.4	Conclusion	182
7.5	Think Ahead!	182
Bibliography		183

List of Figures

1.1	Various applications of action understanding.	4
1.2	Concept of <i>moveme</i> and action hierarchy.	5
1.3	Hierarchy according to [48, 91].	6
2.1	Example of naive background subtraction method.	15
2.2	Example of frame subtraction method.	17
2.3	Optical flow vectors in an action frame.	18
2.4	Streaklines for crowd segmentation.	19
2.5	Streaklines in a video.	20
2.6	Streaklines vs. optical flow.	21
3.1	A typical concept of Bag-of-Words.	40
3.2	Some examples of pose-based body part segmentation.	41
3.3	A concept of learning from the Bag-of-Words as per [110].	41
3.4	A flow-diagram for spatio-temporal interest points to classify.	42
3.5	An example for mixed approach with large-scale and small features.	46
3.6	A typical ML approaches for video-based human motion analysis.	47
3.7	Concept of localization and recognition based on prior model and training data.	59
3.8	Detecting group of people to understand their activities.	60
3.9	Understanding the actions of a group of people — talk	60
3.10	Examples for the concept of joint action recognition and localization.	64
3.11	An approach for action localization and understanding.	65
3.12	An approach for crowd segmentation by exploiting optical flow and streaklines.	65
3.13	An approach for abnormal behavior detection by exploiting streaklines.	66
4.1	Template-based recognition.	77

4.2	Applications of the MHI and its variants. Related to activity analysis.	78
4.3	Applications of the MHI and its variants. Related to interactive applications. .	81
4.4	Different <i>tau</i> values and the corresponding MHI images for an action.	83
4.5	Delta dependency.	84
4.6	Optic flow into four channels.	97
4.7	A flow diagram for the DMHI method for action recognition.	98
4.8	HMHH images for for patterns for six actions.	100
5.1	Shape representation — various approaches.	120
5.2	Classifier black-box.	134
5.3	Example of an unsupervised learning	138
5.4	Choice of <i>k</i> : an example.	141
6.1	A typical motion capture system layout.	149
6.2	Few frames for KTH action dataset.	150
6.3	Sample frames for Weizmann action dataset.	151
6.4	Few action samples for HOHA2 actions dataset.	155
6.5	Nursing-home dataset: Samples with some results.	159
6.6	Nursing-home dataset: Samples with some results of <i>fall</i> detection.	160
6.7	Sample images from the collective activity dataset.	161
6.8	UMN dataset: Unusual Crowd Activity — example with detection	169
6.9	Abnormal crowd dataset — few sample images.	170
6.10	Sample frames from HumanEva dataset.	171
6.11	Human interaction dataset.	171
6.12	Aerial-view activity dataset — person extraction from each class.	172
6.13	Wide-area activity dataset.	172
7.1	Choose the best- <i>pose</i> based on some <i>score</i>	179

List of Tables

2.1 Background subtraction methods and their properties. 17

4.1 Recognition and related areas. 107

4.2 On motion analysis. 109

4.3 On motion analysis. 110