Synthesis Lectures on Computer Science

SYNTHESIS
COLLECTION OF TECHNOLOGY

Santosh Kumar Tripathy · Roshan Singh · Rajeev Srivastava · Akash Kumar Bhoi · Santosh Kumar Satapathy

Advances in Human Activity Detection and Recognition (HADR) Systems

Springer

Synthesis Lectures on Computer Science

The series publishes short books on general computer science topics that will appeal to advanced students, researchers, and practitioners in a variety of areas within computer science.

Santosh Kumar Tripathy · Roshan Singh · Rajeev Srivastava · Akash Kumar Bhoi · Santosh Kumar Satapathy

Advances in Human Activity Detection and Recognition (HADR) Systems

Springer

Santosh Kumar Tripathy
School of Computing Science and Engineering
VIT Bhopal University
Kothri Kalan, Sehore, Madhya Pradesh, India

Rajeev Srivastava
Department of Computer Science
and Engineering
IIT (BHU) Varanasi
Varanasi, Uttar Pradesh, India

Santosh Kumar Satapathy
Information and Communication Technology,
School of Technology
Pandit Deendayal Energy University
Gandhinagar, Gujarat, India

Roshan Singh
Centre for Computing and Information
Services
IIT(BHU) Varanasi
Varanasi, Uttar Pradesh, India

Akash Kumar Bhoi
Directorate of Research
Sikkim Manipal University
Gangtok, Sikkim, India

ISSN 1932-1228 ISSN 1932-1686 (electronic)
Synthesis Lectures on Computer Science
ISBN 978-3-031-51659-7 ISBN 978-3-031-51660-3 (eBook)
https://doi.org/10.1007/978-3-031-51660-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

About This Book

Detection or recognition of human activities is a prominent research area in the field of Computer Vision and Artificial Intelligence. An intelligent-based human activity detection or recognition (HADR) system is highly required for several applications like activity monitoring in public transport areas, monitoring activities of daily livings (ADLs), health monitoring, anomaly detection in traffic, crowd and parking lots, smart home, smart offices, elderly care, border security control, assistant for physically handicap people and so on. There is a huge number of applications that demand automatic human activity detection or recognition. We can observe the journey of research in this area starting from a particular activity detection to recognition and their behavior analysis. Activities can be of three types like normal, abnormal or suspicious. Both abnormal and suspicious activities are unusual activities, but suspicious activities possess a high degree of theft. We can see different surveys and reviews made in this area, but in our proposed survey paper, we finely divided different activities according to their criticality, clearly discussed different motivations, issues and challenges, we presented a framework for activity detection or recognition using different types of vision sensors and multimodal sensors and minutely described every individual step in it. A brief review of the available human activity recognition datasets is also presented in this book. In addition, we have identified ten different applications of HADR and briefly discussed recent developments in each of these applications. According to the presented framework, application wise, we made a tabular comparative study on recent researches. Finally, we proposed some prominent future studies in these applications and then concluded.

Contents

1 Introduction 1
- 1.1 Background 1
- 1.2 Motivation 4
  - 1.2.1 Video Surveillance or Security Surveillance 4
  - 1.2.2 Health Monitoring 5
  - 1.2.3 Human Computer Interface or Interaction (HCI) 5
  - 1.2.4 Ambient Assisted Living (AAL) and Elderly Care 5
  - 1.2.5 Child Activity Recognition 6
- 1.3 Issue and Challenges 6

2 Basic Framework of HADR 9
- 2.1 Data Collection 9
- 2.2 Pre-processing 12
  - 2.2.1 Pre-processing with Data Fusion 12
  - 2.2.2 Pre-processing Without Data Fusion 14
- 2.3 Feature Identification for Segmentation 15
- 2.4 Segmentation 15
  - 2.4.1 Video Segmentation and Human Tracking 15
- 2.5 Post-processing After Segmentation 18
- 2.6 Feature Extraction 19
  - 2.6.1 Feature Identification 19
  - 2.6.2 Feature Fusion 23
  - 2.6.3 Feature Selection Methods 25
- 2.7 Different Approaches Used for HADR 26
  - 2.7.1 Machine Learning Approaches Used for HADR 26
  - 2.7.2 Conventional Approaches Used for HADR 28
- 2.8 Performance Evaluation 29

3 Datasets 33
- 3.1 Evolution of Modern Datasets 34
- 3.2 Characteristics of Datasets 35

3.2.1 Focus . . . 35
3.2.2 Modality . . . 36
3.2.3 Data Source . . . 37
3.2.4 Annotation Method . . . 38
3.2.5 Annotation Type . . . 38
3.2.6 Evaluation . . . 39
3.3 Video Datasets . . . 40
3.3.1 Hollywood2Tubes (2016), Hollywood2 (2009), Hollywood (2008) . . . 42
3.3.2 High Five (2010) . . . 42
3.3.3 UCF101 (2012), UCF50 (2010) and UCF11 (2009) . . . 45
3.3.4 THUMOS Challenges (2013, 2014, 2015) . . . 46
3.3.5 MultiTHUMOS (2017) . . . 46
3.3.6 HMDB51 (2011) and JHMDB (2013) . . . 47
3.3.7 Action Similarity Labeling Dataset (2012) . . . 47
3.3.8 Sports-1M (2014) . . . 47
3.3.9 SVW (2015) . . . 48
3.3.10 ActivityNet (2015) . . . 48
3.3.11 Charades Ego (2018) and Charades (2016) . . . 48
3.3.12 Something Something (2017) . . . 49
3.3.13 DALY (2017) . . . 50
3.3.14 AVA (2017) . . . 50
3.3.15 A2D (2017) . . . 50
3.3.16 Kinetics600 (2017) and Kinetics400 . . . 51
3.3.17 VLOG (2017) . . . 52
3.3.18 SoccerNet (2018) . . . 53
3.3.19 EPIC-Kitchens (2018) . . . 53
3.3.20 MLB-YouTube (2018) . . . 53
3.3.21 STAIR Actions (2018) . . . 54
3.3.22 Moments in Time (2018) . . . 55
3.3.23 HACS (2019) . . . 55
3.3.24 Coventry-2018 . . . 56
3.3.25 KU-HAR (2021) . . . 56
3.3.26 POLIMI-ITW-S (2022) . . . 57
3.4 Conclusion . . . 57

4 Applications for HADR . . . 59
4.1 Detection or Recognition of Normal Human Activities/Activity Daily Livings (ADLs) . . . 60
4.2 Detection of Abnormal/Suspicious Human and Human Activities . . . 62
4.3 Abandoned and Stolen Object Detection and Tracking . . . 65
4.4 Detection of Violence and Vandal Activities . . . 67

4.5 Human Fall Detection and Recognition by Using Video Surveillance and Other Sensors (Inertial Sensors) 70
4.6 Detection, Recognition, and Monitoring of Fire, Flames, and Smoke ... 73
4.7 Anomaly Detection and Behavior Analysis of Crowd 75
4.8 Detection and Recognition of Human Activities in Reference with Traffic Control System 77
4.9 Detection and Recognition of Activities in Reference with Parking Space 79
4.10 Detection and Recognition of Child Activities 80

5 Comparative Analysis 83

6 Conclusion and Future Directions 117

References 119

List of Figures

Fig. 1.1 Different types of human activities 2
Fig. 1.2 Basic work-flow diagram of HADR 3
Fig. 2.1 Basic framework of HADR system 10
Fig. 2.2 Sensors used in HADR 11
Fig. 2.3 Data preprocessing 13
Fig. 2.4 Video segmentation and tracking methods 16
Fig. 2.5 Different process and methods used for feature extraction 20
Fig. 2.6 Deep-CNN convolutional layers, each has 32 feature maps, and each convolution layer is followed by a Max Pooling layer with a size of 2×2 [75] 25
Fig. 2.7 Different methods used for HADR 27
Fig. 3.1 Datasets Evolution from left (Scripted) to right (Unscripted) as seen through common activities: Top Row (Running), Bottom Row (Walking). The variation in contents of unscripted datasets can be seen 34
Fig. 3.2 Actions along with their sub-categories 36
Fig. 3.3 Camera and sensor based modality 37
Fig. 3.4 Data Source for dataset generation 37
Fig. 3.5 Annotations types 39
Fig. 3.6 Charades ego dataset with third person (Top) and first person (Bottom) videos for the same actions 49
Fig. 3.7 AVA dataset, Example classes selected at random 51
Fig. 3.8 Label counts for A2D dataset with invalid classes left blank 51
Fig. 3.9 Kinetics dataset actions: Abseiling, scrambling eggs, country line dancing, flipping pancakes, mowing lawn 52

Fig. 3.10 MLB example captions 54
Fig. 3.11 Actions in the moments dataset, left to right and top to bottom: Aiming, Applauding, Buying, Chopping, Dining, Hugging, Reading, Telephoning, Vacuuming, Wrestling, Writing, Yawning 55
Fig. 3.12 HACS classes: Top row-negatives, bottom row positive. Actions from right to left: Swimming rock climbing, skipping rope 56

List of Tables

Table 1.1 Recent survey on different human activity recognition 3
Table 2.1 Different features extracted from sensory data 21
Table 2.2 Difference between deep features and handcrafted features 24
Table 2.3 Confusion matrix 30
Table 3.1 Early datasets evolution 41
Table 3.2 Modern datasets evolution 43
Table 3.3 Evolution of egocentric datasets 45
Table 5.1 Results analysis of several researches on Activity Daily Livings (ADLs) 84
Table 5.2 Results analysis of several researches on abnormal/suspicious human and human 88
Table 5.3 Results Analysis of Abandoned and stolen object detection and tracking 92
Table 5.4 Results analysis of violence and vandal activity detection 95
Table 5.5 Results analysis of several researches on fall detection 99
Table 5.6 Results analysis of detection, recognition and monitoring of fire, flames or smoke 103
Table 5.7 Results analysis of several researches on anomaly detection and behavior analysis of crowd 105
Table 5.8 Results analysis of detection, recognition of human activities in traffic control system 109
Table 5.9 Results analysis of detection and recognition of activities in parking space 113
Table 5.10 Results analysis of detection and recognition of child activities 115

Introduction 1

1.1 Background

HADR is one of the significant research areas in the fields of computer vision and artificial intelligence. Generally, human activities vary from simple to complex. Due to the high criticality of human actions and behaviour, it becomes cumbersome for laymen or human experts to continuously observe social activities to fulfil the requirements in applications like security surveillance, health monitoring, in-house activity monitoring, traffic control systems, public transport security issues, guidance for physically handicapped people, etc. So, an intelligent-based human activity recognition system is highly required. We can observe the journey of research in this field, starting from a particular activity's detection to several activities' recognition and behaviour analysis. The activities can be of three types: normal activity, abnormal activity, and suspicious activity. Normal activities are the usual activities performed by humans in their day-to-day lives. Detecting and recognising activities of daily living (ADLs) like sitting, walking, running, climbing, jumping, hand waving, talking, etc.; detection of child activities like crawling, moving, wiggling, toddling, rolling, sitting down, standing up, standing still, climbing down, etc.; activity control in smart home applications are some of the examples of research applications of Normal-HADR. Abnormal events are those activities that are generally uncommon or rare. Some of the applications of Abnormal-HADR are detecting fall activities, fire monitoring, illegal parking activity recognition, abnormal crowd activity monitoring, health monitoring, elderly care, etc. But suspicious activities are those that are doubtful, illegal, harmful, distrustful, or dishonest and also involve crime-like activities. Examples of such suspicious-HADR are: detection of violence or vandal activities; suspicious human (wearing a helmet and continuously peeping) detection in ATMs; detection of illegal border crossing activities; abandoned object detection in the railways, airports, shopping malls, etc.; recognition of threats like activities in streets, malls, highways, railway stations, etc.

S. K. Tripathy et al., *Advances in Human Activity Detection and Recognition (HADR) Systems*, Synthesis Lectures on Computer Science,
https://doi.org/10.1007/978-3-031-51660-3_1

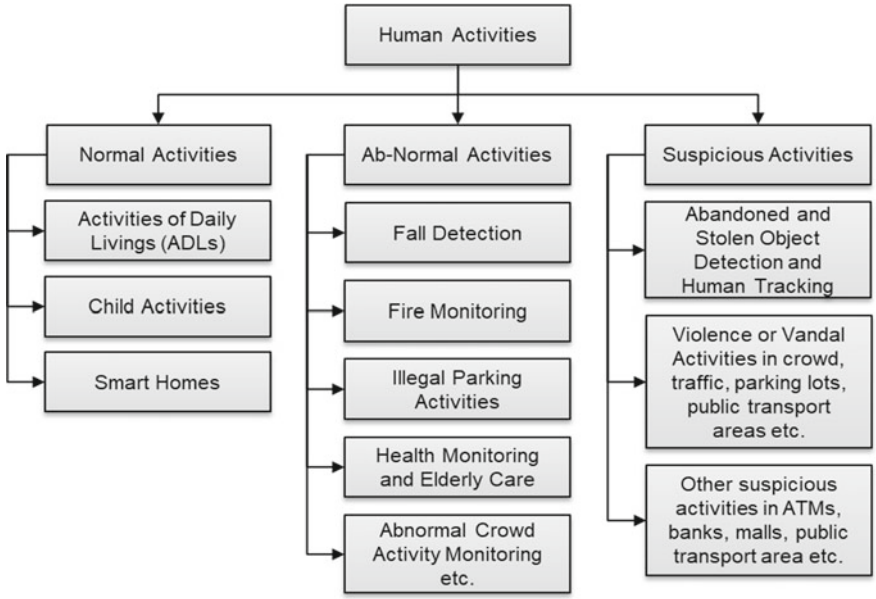

Fig. 1.1 Different types of human activities

In the literature, researchers interchangeably used abnormal and suspicious activities. The following Fig. 1.1 shows some examples of different kinds of activities. In our survey paper, we focused on ten types of applications for HADR.

HADR can be implemented by using security surveillance cameras or any vision-based cameras, mobile phones, or any embedded device consisting of inertial sensors like accelerometers, gyroscopes, and magnetometers, or even by using any multimodal sensors like temperature, pressure, EEG (electroencephalogram), ECG (electrocardiogram), vibration sensors, GPS (Global Positioning System), etc. In our survey, we considered applications that use these sensors for activity detection and recognition purposes.

Researchers have proposed several benchmark processes to recognise human activities, and they have also dealt with several situations that became bottlenecks in implementing the approaches. Different surveys on several types of activity recognition can be found in the literature. The following Table 1.1 shows different surveys made on human activity recognition. By observing different researches made on HADR, we designed a framework that describes a step-by-step process to implement several sensor-based HADR. The following Fig. 1.2 shows the basic work-flow diagram of HADR.

At first, different sensors like CCTV cameras, Radio Frequency Identification Radar (RFID) sensors, pressure sensors, temperature sensors, smartphones, or smart handbands embedded with the accelerometer, gyroscope, etc. are used to collect data. Then, the HADR system will perform a low-level task on the collected data. Such low-level jobs