A FUZZY APPROACH FOR EARLY HUMAN ACTION DETECTION

EKTA VATS

FACULTY OF COMPUTER SCIENCE AND INFORMATION
TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

A FUZZY APPROACH FOR EARLY HUMAN ACTION DETECTION

EKTA VATS

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: (I.C./Passport No.:)

Registration/Matrix No.:

Name of Degree:

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Field of Study:

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:
Designation:

ABSTRACT

Early human action detection is an important computer vision task with a wide spectrum of potential applications. Most existing methods deal with the detection of an action after its completion. Contrarily, for early detection it is essential to detect an action as early as possible. Therefore, this thesis develops a solution to detect ongoing human action as soon as it begins, but before it finishes.

In order to perform early human action detection, the conventional classification problem is modified into frame-by-frame level classification. There exists well-known classifiers such as Support Vector Machines (SVM), K-nearest Neighbour (KNN), etc. to perform action classification. However, the employability of these algorithms depends on the desired application and its requirements. Therefore, selection of the classifier to employ for the classification task is an important issue to be taken into account. The first part of the thesis studies this problem and fuzzy Bandler-Kohout (BK) sub-triangle product (subproduct) is employed as a classifier. The performance is tested for human action recognition and scene classification. This is a crucial step as it is the first attempt of using fuzzy BK subproduct for classification.

The second part of this thesis studies the problem of early human action detection. The method proposed is based on fuzzy BK subproduct inference mechanism and utilizes the fuzzy capabilities in handling the uncertainties that exist in the real-world for reliable decision making. The fuzzy membership function generated frame-by-frame from fuzzy BK subproduct provides the basis to detect an action before it is completed, when a certain threshold is attained in a suitable way. In order to test the effectiveness of the proposed framework, a set of experiments is performed for few action sequences where the detector is able to recognize an action upon seeing ~32% of the frames.

Finally, the proposed method is analyzed from a broader perspective and a hybrid technique for early anticipation of human action is proposed. It combines the benefits of computer vision and fuzzy set theory based on fuzzy BK subproduct. The novelty lies in the construction of a frame-by-frame membership function for each kind of possible movement, taking into account several human actions from a publicly available dataset. Furthermore, the impact of various fuzzy implication operators and inference structures in retrieving the relationship between the human subject and the actions performed is discussed. The existing fuzzy implication operators are capable of handling only two-dimensional data. A third dimension 'time' plays a crucial role in human action recognition to model the human movement changes over time. Therefore, a new space-time fuzzy implication operator is introduced, by modifying the existing implication operators to accommodate time as an added dimension. Empirically, the proposed hybrid technique is efficiently able to detect an action before completion and outperform the conventional solutions with good detection rate. The detector is able to identify an action upon viewing ~23% of the frames on an average.

ABSTRAK

Pengesanan awal kelakuan manusia merupakan satu tugas visi komputer yang penting kerana ianya mempunyai aplikasi-aplikasi berpotensi luas. Kebanyakan kaedah-kaedah yang sedia ada hanya mengesan kelakuan manusia setelah kelakuan tersebut telah lengkap. Sebaliknya, ia adalah penting bagi mengesan kelakuan manusia secepat mungkin. Oleh yang demikian, tesis ini membentuk satu penyelesaian baru untuk mengesan kelakuan manusia, sebaik sahaja ia bermula, tetapi sebelum kelakuan tersebut disempurnakan.

Dalam usaha untuk melaksanakan pengesanan awal kelakuan manusia, masalah klasifikasi konvensional diubah suai ke masalah klasifikasi bingkai demi bingkai (frame-by-frame level classification). Kini, wujud pengelas terkenal seperti Mesin Vector Sokongan (Support Vector Machine, SVM), K-Neighbour terdekat (K-nearest Neighbour, KNN), dan lain-lain, untuk melaksanakan pengelasan. Walau bagaimanapun, keberkesanan algoritma-algoritma ini bergantung kepada aplikasinya dan syaratnya. Oleh itu, pemilihan pengelas untuk tugas pengelasan merupakan isu penting yang perlu diprihatin. Bahagian pertama tesis ini mengkaji masalah tersebut dan menggunakan Bandler-Kohout kabur dengan Produk sub-segi tiga (fuzzy Bandler-Kohout sub-triangle product, atau ringkasannya fuzzy BK subproduct) sebagai pengelas. Prestasi pengelas tersebut diuji dalam pengiktirafan kelakuan manusia dan klasifikasi tempat (scene). Ini adalah satu langkah penting kerana ia adalah percubaan pertama menggunakan fuzzy BK subproduct untuk pengelasan.

Bahagian kedua tesis ini mengkaji masalah pengesanan awal kelakuan manusia. Kaedah yang dicadangkan adalah berdasarkan mekanisma inferens daripada fuzzy BK subproduct dan menggunakan keupayaan kabur (fuzzy capabilities) dalam menangani ketidakpastian yang wujud di dunia sebenar untuk membuat keputusan yang lebih tepat.

Fungsi keahlian kabur (fuzzy membership function) dihasilkan frame-by-frame dari fuzzy BK subproduct memberi asas yang diperlukan untuk mengesan sesuatu tindakan sebelum ia selesai, apabila ambang (threshold) tertentu dicapai dengan cara yang sesuai. Untuk menguji keberkesanan bagi kaedah yang dicadangkan, eksperimen dilakukan untuk beberapa kelakuan manusia yang mana pengesan dapat mengenali kelakuan tersebut apabila melihat 32% daripada keseluruhan bingkai (frames). Akhirnya, kaedah yang dicadangkan dianalisis dari perspektif yang lebih luas dan satu teknik hibrid untuk jangkaan awal kelakuan manusia adalah dicadangkan. Ia menggabungkan manfaat visi komputer dan teori set kabur berdasarkan fuzzy BK subproduct. Kebaharuannya terletak pada pembinaan fungsi keahlian frame-by-frame untuk setiap jenis pergerakan yang mungkin, dengan mengambil kira beberapa kelakuan manusia dari dataset umum.

Tambahan pula, kesan pelbagai pengendali implikasi kabur dan struktur inferens dalam mendapatkan semula hubungan antara subjek manusia dan kelakuan yang dilakukan telah dibincangkan. Pengendali implikasi kabur yang sedia ada hanya mampu mengendalikan data dalam dua dimensi. Dimensi ketiga, 'masa', memainkan peranan yang penting bagi mengiktiraf tindakan manusia untuk pemodelan bagi perubahan pergerakan manusia dari semasa ke semasa. Oleh itu, satu pengendali implikasi kabur berdasarkan ruang-masa (space-time) diperkenalkan, dengan mengubah pengendali implikasi sedia ada untuk menampung masa sebagai dimensi tambahan. Secara empirik, teknik hibrid yang dicadangkan adalah cekap dan dapat mengesan tindakan sebelum lengkap dan mengatasi penyelesaian konvensional dengan kadar pengesanan yang baik. Pengesan tersebut dapat mengenal pasti sesuatu tindakan setelah melihat 23% daripada keseluruhan bingkai secara purata.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Chan Chee Seng for being an incredible mentor. I would like to express my heartiest gratitude to him for introducing me to the interesting field of computer vision and fuzzy set theory, guiding me tirelessly in my research, and for the strong support throughout my candidature. Without his constant support, this thesis wouldn't have been completed successfully. I am deeply grateful to Dr. Chan, the Faculty of Computer Science and Information Technology, and the High Impact Research (HIR) grant for providing me with the much appreciated financial support during my degree.

I would also like to thank Lim Chee Kau and Lim Chern Hong for their co-authorship. I am grateful to both of them for the cooperation and helpful suggestions. I thank all my colleagues and ex-colleagues in the Center of Image and Signal Processing for a very pleasant and friendly working environment.

Special thanks goes to my father Neeraj Kumar Vats, mother Neetu Vats and sister Vandita for the unconditional support. My words fall short to express how lucky I am to have a beautiful family like ours. Your blessings, love and care have been my constant motivation towards fulfilling my dreams and aspirations in life. Lastly, my profound gratitude to my loving husband Prashant for always standing by my side and keeping faith in me. Thank you for the unflagging support through both the highs and lows of my life during my PhD.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

2D	:	Two-dimensional.
3D	:	Three-dimensional.
ARMA	:	Autoregressive-moving-average.
BK	:	Bandler-Kohout.
BoW	:	Bag of Words.
CV	:	Computer Vision.
CWW	:	Computing with Words.
FCM	:	Fuzzy c-means.
FIS	:	Fuzzy Inference Structure.
FVQ	:	Fuzzy Vector Quantization.
HiL	:	High-level.
HMA	:	Human Motion Analysis.
HMM	:	Hidden Markov Model.
KNN	:	K-nearest Neighbour.
LoL	:	Low-level.
MiL	:	Mid-level.
MMED	:	Max Margin Early Event Detector.
NTtoD	:	Normalised Time to Detect.
pLSA	:	probabilistic Latent Semantic Analysis.
QNT	:	Qualitative Normalized Template.
SIFT	:	Scale Invariant Feature Transform.
SOSVM	:	Structured Output SVM.
subproduct	:	Sub-Triangle Product.
SVM	:	Support Vector Machines.

CHAPTER 1: INTRODUCTION

Temporally changing events surround us in daily life, such as the temperature variations over time, fluctuating stock prices, and the changing human behavior. Monitoring the temporally varying human behavior is an important task in the Computer Vision (CV) community where researchers aim at analyzing the time series data constituting the sequences of actions observed over time. A temporal event is time bounded and has a duration, whereas early detection refers to detecting an event as soon as possible i.e. after it starts but before it finishes. In this thesis, the human behavior is studied in the context of analyzing and interpreting human movements over time (Human Motion Analysis (HMA)), with the aim of detecting human action early.

HMA has been a popular research topic that encompasses many domains such as biology (Bobick, 1997; Troje, 2002), psychology (Barclay, Cutting, & Kozlowski, 1978; Blake & Shiffrar, 2007), multimedia (Kirtley & Smith, 2001), etc. In the CV community, HMA has been an active research area over years due to the advancement in video camera technology and the availability of more sophisticated CV algorithms. The real-time applications of HMA include video surveillance (Hatakeyama, Mitsuta, & Hirota, 2008; Popoola & Wang, 2012), health-care monitoring (Anderson, Keller, Skubic, Chen, & He, 2006; Sanchez-Valdes, Alvarez-Alvarez, & Trivino, 2015; Anderson, Luke, et al., 2009b), sport analysis (Rodriguez, Ahmed, & Shah, 2008a; Yeguas-Bolivar, Muñoz-Salinas, Medina-Carnicer, & Carmona-Poyato, 2014), etc.

However, early human action detection has not received much attention in the recent past despite of the fertile potential applications such as criminal attack detection, risk of elderly patients' fall detection, affective human-robot interaction, etc. Most of the methods (C. H. Lim, Vats, & Chan, 2015) deal with detection of the action after its completion. Figure 1.1 explains the scenario of the state-of-the-art methods. For early detection, it is

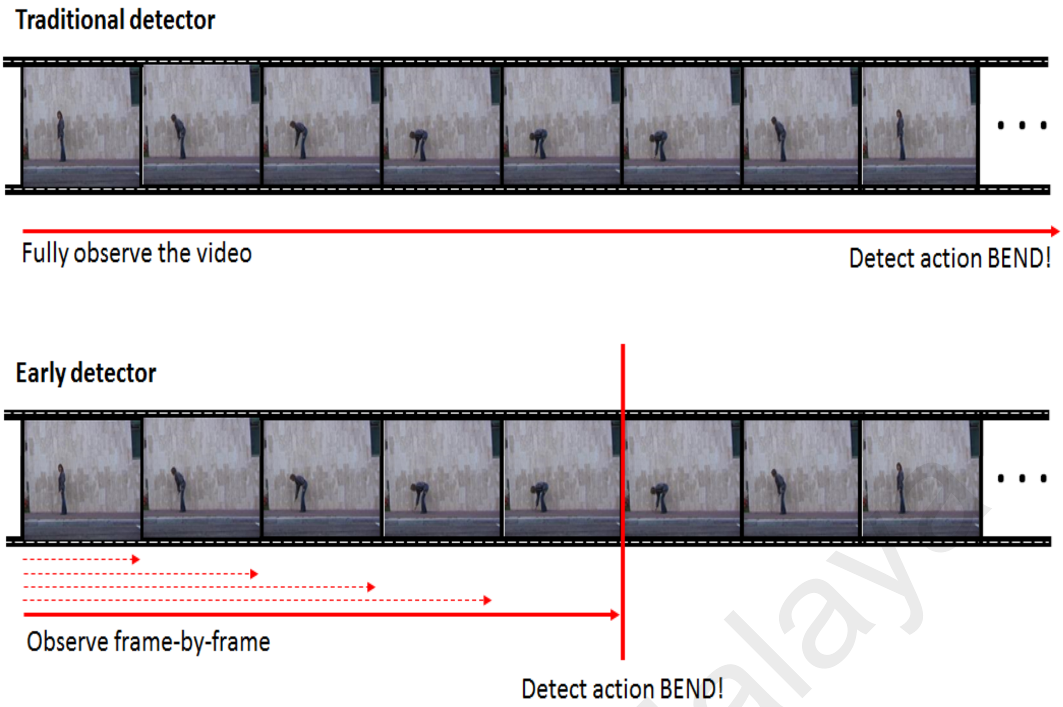
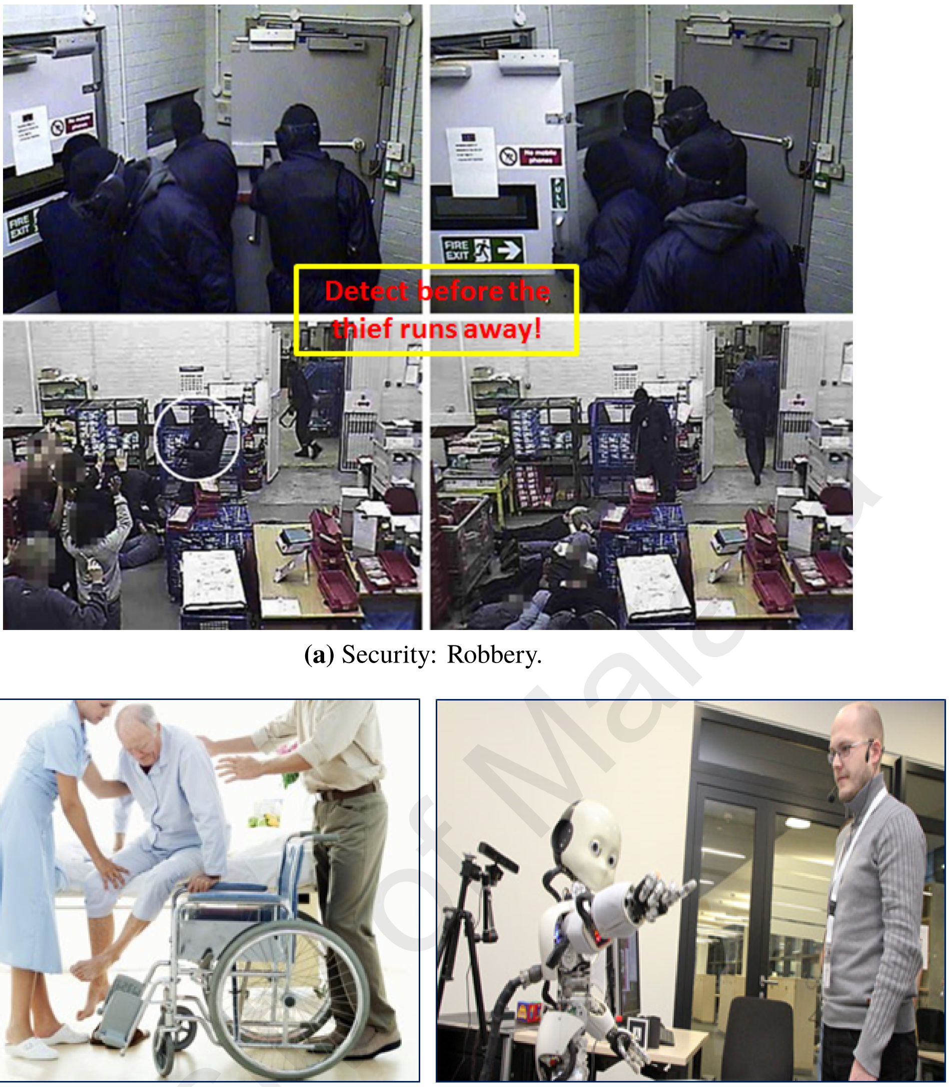


Figure 1.1: Traditional detector versus early detector. The traditional detector detect an action after fully observing the video, whereas the early detector detects an action by observing the video frame-by-frame, such that it able to detect an action before its completion.

essential to detect an action as soon as possible by making observations frame-by-frame (Ryoo, 2011; G. Yu, Yuan, & Liu, 2012; Ryoo, Fuchs, Xia, Aggarwal, & Matthies, 2014; K. Li & Fu, 2012; Hoai & De la Torre, 2012). Figure 1.1 illustrates the difference between traditional detector and the early detector, using an example of ‘bend’ action. By definition, the traditional detector performs action classification after fully observing the video, whereas the early detector aims at detection of an action by observing the video frame-by-frame, such that it able to detect an action before its completion.

1.1 Motivation

The motivation behind early human action detection is driven by the need to detect an action as soon as possible, before it finishes. To see why it is important to detect an action before it is completed, consider the following three concrete examples (as illustrated in Figure 1.2) with reference to the real-world applications:



(**a**) Security: Robbery.

(**b**) Health-care: Elderly patients' fall detection.

(**c**) Robotics: Affective computing.

Figure 1.2: Examples of real-world applications where early human action detection is needed. Image source: *http://images.google.com.*

(a) Security: Consider a surveillance scenario, where recognizing the fact that certain objects are missing after they have been stolen may not be meaningful (Ryoo, 2011). The system could be more useful if it is able to prevent the theft and catch the thieves by predicting the ongoing stealing activity as early as possible based on live video observations.

(b) Health-care: Consider an example of elderly care system. It is crucial to accurately and rapidly detect the elderly patients' fall, so that necessary medical help can be provided in a timely manner before it becomes life threatening (Anderson et al., 2006;

Anderson, Luke, Skubic, et al., 2008). Hence, early detection of elderly patients' fall is very important.

(c) Robotics: Consider an example of building a robot that can affectively interact with a human (Hoai & De la Torre, 2012, 2014). An important characteristic of such robot is its ability to rapidly and accurately detect a human emotion by observing facial expressions, and therefore generate appropriate response with time. The imitation response of the robot should be in synchronization with the current behavior of the human. This means that it is important for the robot to detect facial expression changes of the human, e.g., smiling, frowning, anger or disgust even before they are completed. Therefore, early detection of human behavior is important for affective communication between a robot and a human.

Most of the methods (C. H. Lim et al., 2015) perform after-the-fact detection, where action classification is performed after fully observing the video. However, even if the system detects the action (e.g. crime or patients' fall, etc.), it may be too late to prevent it. Therefore, early detection is required.

1.2 Objectives of Study

This study aims at developing an algorithm for early human action detection. To achieve this goal, efforts are channeled to the following:

(a) The first objective is to select a classifier for human action classification. Therefore, fuzzy Bandler-Kohout (BK) Sub-Triangle Product (subproduct) (Bandler & Kohout, 1980a) is employed as a classifier. The performance is tested for HMA (Three-dimensional (3D) data) and scene classification (Two-dimensional (2D) data).

(b) The second objective is to train a detector to recognize human action as early as possible, without fully observing an action video. The aim is to identify an action upon viewing minimum possible number of frames, and outperform the conventional solutions with good detection rate.

(c) The third objective is to introduce a new space-time fuzzy implication operator, with application in HMA. This is because a third dimension 'time' is not taken into account in the existing fuzzy implication operators, that play a crucial role in a HMA system in order to model human movement changes over time.

In the following section, challenges faced in the research community and the problem formulation are discussed that serve as the main motivation behind this study in order to achieve the research aims and objectives.

1.3 Challenges and Problem Formulation

As previously discussed, monitoring the temporally varying human behavior is an important task, and has been widely studied in literature (C. H. Lim et al., 2015). However, early human action detection has not received the much needed attention despite of the potential applications in the field of security, health-care, etc. The main problem is that most of the methods (C. H. Lim et al., 2015) deal with the detection of action after its completion, and for early detection it is essential to detect an action as soon as possible by making observations frame-by-frame, as illustrated in Figure 1.1. In this thesis, this issue is addressed and an algorithm is proposed to detect ongoing human action early by training a detector capable of detecting a human action seeing minimum possible number of frames. Therefore, the conventional classification problem is modified into frame-by-frame level classification to perform early detection.

However, early human action detection is a daunting task given the vast amount of uncertainties involved therein. Figure 1.3 illustrates the possible uncertainties that may

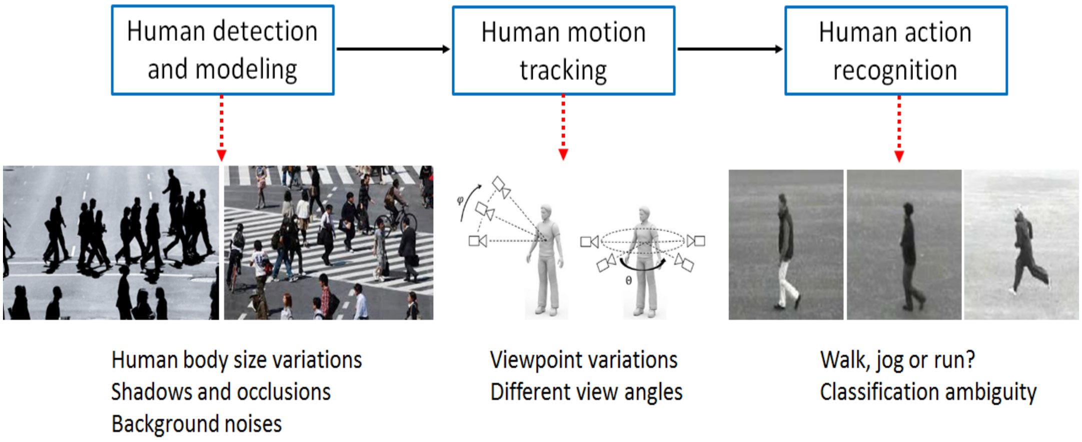


Figure 1.3: Several sources of uncertainties that can exist at each step in a HMA system. For example, human size variations, shadows, occlusions and background noises can affect human detection and modeling process. The performance of human motion tracking algorithms may be affected due to different viewpoint angles. And the classification ambiguity can be a major source of uncertainty while performing human action recognition.

exist at each step in a HMA system. Some of the common sources of uncertainties include background noises, occlusions, human body size variations, different viewpoint or angles, classification ambiguity, etc. An efficient algorithm should be able to handle even the minutest level of uncertainty for reliable decision making as cumulated errors can deteriorate the overall system performance.

There exist some notable works that deal with early human action detection and aim at detecting the unfinished activities, e.g. Ryoo (2011); G. Yu et al. (2012); Ryoo et al. (2014); K. Li and Fu (2012); Hoai and De la Torre (2012). However, despite of the advantages these methods offer, they lack in the ability to handle issues such as uncertainty, imprecision and vagueness. An important reason behind this problem is that their classification results are binary. This means that an action can belong to a single class only at a time. Nonetheless, fuzzy approaches are known to offer an effective solution and allows an action to belong to multiple classes. This is achieved by assigning a degree of belongingness to a human action using the fuzzy membership function, and the fuzzy rules. This work proposes a fuzzy approach for early human action detection.

From the literature review by C. H. Lim et al. (2015), it is found that there exist a number of fuzzy approaches for HMA. In this work, fuzzy BK subproduct approach is selected due to its flexibility and efficacy to be employed in real-world applications (C. K. Lim & Chan, 2015; Bui & Kim, 2006; Groenemans, Van Ranst, & Kerre, 1997; Vats, Lim, & Chan, 2012), and its capability to imitate the natural human behavior, i.e. modus-ponen way (C. K. Lim & Chan, 2011). Modus-ponen refers to the interpretation of available information while solving real-life problems. For example, if *A* implies *B*, and *A* is asserted to be true, therefore *B* must be true. Nonetheless, fuzzy BK subproduct does not require defining rules for inference, and hence is computationally inexpensive. Rather it is based on the study of relationship between two sets, where if there exists an intermediate set which is in relation with both the sets, then the indirect relationship can be established.

Using fuzzy BK subproduct inference mechanism, the detector is trained and used separately for each of the target action classes. The challenge is to study the indirect relationship between the human subject and the action being performed in the video. This can be achieved by modeling the frame-by-frame arrival of data, and subsequently performing action classification on the basis of the membership function values generated from fuzzy BK subproduct.

In general, the CV methods and fuzzy approaches do not behave in a conflicting manner, rather compliment one another (C. H. Lim et al., 2015). The fusion of these techniques towards performing human action recognition as early as possible can be achieved through proper hybridization. To this end, the relationship between a human subject and the action being performed is studied using fuzzy BK subproduct, efficiently integrated with CV techniques including feature extraction and motion tracking to perform human action recognition effectively. The fuzzy membership function provides the basis to detect an action before it is completed when a certain threshold is attained in a suitable way.

A solution for early human action detection is intended that is closest to natural human perception. The novelty lies in the hybrid based learning formulation to train the early detector such that once the detector has been trained, it can be flexibly used in several ways depending upon the application.

1.4 Contributions

The main contributions of this thesis are highlighted in Figure 1.4, and are as follows:

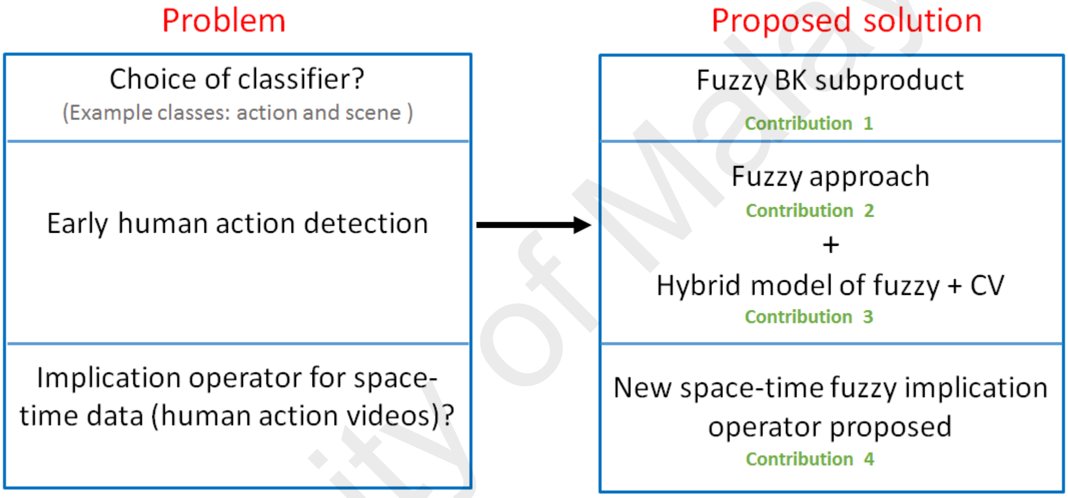

Problem
Choice of classifier?
(Example classes: action and scene)
Early human action detection
Implication operator for space-time data (human action videos)?
Proposed solution
Fuzzy BK subproduct
Contribution 1
Fuzzy approach
Contribution 2
+
Hybrid model of fuzzy + CV
Contribution 3
New space-time fuzzy implication operator proposed
Contribution 4

Figure 1.4: The main problems addressed in this thesis along with the proposed solutions.

Contribution 1: Firstly, this thesis addresses the most fundamental problem of selecting a classifier to employ for the classification task. As a solution, fuzzy BK subproduct is used as a classifier. In order to demonstrate the capability of fuzzy BK subproduct in handling both 3D video data and 2D image data, its performance is tested for HMA and scene classification.

Experimental results on standard public datasets demonstrate the effectiveness of fuzzy BK subproduct in performing HMA and scene classification. This is the first attempt of using fuzzy BK subproduct as a classifier, and the research work is accepted for publication in the proceedings of IEEE International Conference on Fuzzy Systems

(FUZZ-IEEE 2015), held in Istanbul, Turkey, and in the Journal of Intelligent and Fuzzy Systems (2015).

Contribution 2: Secondly, this thesis proposes a novel framework to detect human action early based on fuzzy BK subproduct inference mechanism by utilizing the fuzzy capabilities in handling the uncertainties that exist in the real-world for reliable decision making. Frame-by-frame action classification is performed for early detection where the fuzzy membership function generated from fuzzy BK subproduct provides the basis to detect an action before it is completed when a certain threshold is attained in a suitable way. In order to test the effectiveness of the proposed framework, a set of experiments is performed for few action sequences where the aim of the detector is to recognize an action upon seeing minimum number of frames possible.

To the best of my knowledge, there does not exist any work with the application of fuzzy BK subproduct approach for human action recognition. This is the first work in the fuzzy community dealing with early human action detection. This work is accepted for publication in the proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2015), held in Istanbul, Turkey.

Contribution 3: Thirdly, the proposed framework is analyzed from a broader perspective where it can be represented as a hybrid model of CV and fuzzy set theory based on fuzzy BK subproduct. Hybrid techniques address issues such as uncertainty, vagueness or imprecision to a considerable extent by exploiting the strengths of one technique to alleviate the limitations of another (Acampora, Foggia, Saggese, & Vento, 2012; Hosseini & Eftekhari-Moghadam, 2013).

To this end, the proposed solution is the synergistic integration of CV solutions and fuzzy set theory where the relationship between a human subject and the action being performed is studied using fuzzy BK subproduct, efficiently integrated with CV techniques including feature extraction and motion tracking to perform human action recognition effectively. The novelty lies in the construction of a frame-by-frame membership function for each kind of possible movement, taking into account several human actions from a publicly available dataset. Another issue addressed by the proposed method is to handle the cumulative tracking errors and precision problem. This can be achieved by using a set of overlapped fuzzy numbers known as fuzzy qualitative quantity space, where individual distance among them is defined by a preselected metric (H. Liu & Coghill, 2005). A solution for early human action detection closest to natural human perception is intended. The contribution lies in the hybrid based learning formulation to train the early detector such that once the detector has been trained, it can be flexibly used in several ways according to different types of application.

Empirically, the proposed hybrid technique can efficiently detect a human action before completion and outperform the conventional solutions with good detection rate. The detector aims at identifying an action upon viewing minimum number of frames for test data under the experimental settings. This work is accepted for publication in Applied Soft Computing (2015).

Contribution 4: Finally, a study is performed on the impact of various fuzzy implication operators and the inference structures in retrieving the relationship between the human subject and the action. The existing fuzzy implication operators are capable of handling 2D data only. However, a third dimension ‘time’ plays a crucial role in human action recognition to model human movement changes over time. Therefore, a new space-time fuzzy implication operator is introduced, by modifying the existing implication operators to accommodate time as an added dimension. This work is accepted for publication in the proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2015), held in Istanbul, Turkey, and in Applied Soft Computing (2015).

1.5 Outline of Thesis

This thesis is organized into six main chapters, as described with a brief overview on each as follows:

Chapter 1 presents an overview on HMA and early human action detection in general, while highlighting the motivation and the objectives of the study. Furthermore, the challenges and problem formulation are discussed, followed by the highlights on the main contributions of this thesis.

Chapter 2 reviews the state-of-the-art methods and solutions that are relevant to the problem statement this thesis is addressing. Fuzzy human motion analysis is reviewed in an elaborate manner in order to understand the necessity of employing fuzzy techniques for HMA. Also, the challenges and the current state of the problems are discussed. Furthermore, fuzzy BK subproduct approach is reviewed, followed by the review on the state-of-art methods for early human action detection along with their limitations.

Chapter 3 discusses the most fundamental issue of selecting the classifier to employ for the classification task. As a solution, fuzzy BK subproduct is employed as a classifier, with its employability tested for HMA and scene classification.

Chapter 4 presents a detailed description of the proposed method to detect human action early. The proposed method is based on fuzzy BK subproduct inference mechanism and utilizes the fuzzy capabilities in handling uncertainties that exist in the real-world. It discusses how frame-by-frame action classification is performed, thus enabling early detection. The fuzzy membership function generated from fuzzy BK subproduct provides the basis to detect an action before it is completed when a certain threshold is attained in a suitable way. A set of experiments is performed for few action sequences in order to test the effectiveness of the proposed framework.

Chapter 5 analyzes the the proposed framework from a broader perspective where the novelty lies in the construction of a frame-by-frame membership function for each kind of possible movement, taking into account several human actions from a publicly available dataset. In specific, the main idea behind the proposed framework, i.e. the hybridization of CV and the fuzzy set theory based on fuzzy BK subproduct is discussed and formulated. Furthermore, the impact of various fuzzy implication operators and the inference structures in retrieving the relationship between the human subject and the actions performed is discussed. A new space-time fuzzy implication operator is introduced, with application in HMA. Experimental results are demonstrated to further validate the effectiveness of the proposed hybrid technique to detect a human action early.

Chapter 6 concludes the research work and suggests a number of areas for future investigation.

CHAPTER 2: BACKGROUND RESEARCH

In this section, HMA is first reviewed where the current trends in HMA is studied along with the limitations in terms of the inability to handle the uncertainties that may exist in a real-world. The reason for adopting fuzzy approach in HMA is critically reviewed, and the overall pipeline of HMA is represented in three levels: Low-level (LoL), Mid-level (MiL) and High-level (HiL) HMA. Furthermore, BK subproduct approach is reviewed with highlights on its applications, followed by a review on the state-of-the-art methods for early human action detection and their limitations. In general, the overall background research is conducted as presented in Figure 2.1.

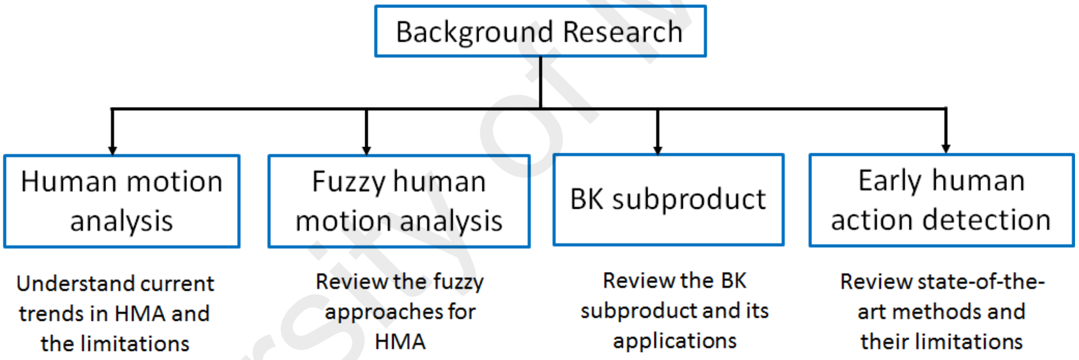

Background Research
Human motion analysis
Fuzzy human motion analysis
BK subproduct
Early human action detection
Understand current trends in HMA and the limitations
Review the fuzzy approaches for HMA
Review the BK subproduct and its applications
Review state-of-the-art methods and their limitations

Figure 2.1: Overall representation of the background research conducted.

2.1 Human Motion Analysis

Human motion analysis (HMA) refers to the analysis and interpretation of human movements over time. HMA has been studied extensively in the CV literature for decades due to its increasing demand and advancement in camera technology. Here, HMA concerns with the detection, tracking and human action recognition, and more generally the understanding of human behaviors from image sequences involving humans. Amongst all, video surveillance is one of the most important real-time applications (Hu, Tan, Wang, & Maybank, 2004; Ko, 2008; Haering, Venetianer, & Lipton, 2008; I. S. Kim, Choi, Yi,

(a) Madrid train bombing

(b) London bombing

(c) Boston marathon bombing

Figure 2.2: (a) *Madrid train bombing (March 11, 2004):* 191 people were killed, and 1,800 others were injured in the Madrid commuter rail network bombing attack, (b) *London bombing (July 7, 2005):* A series of co-ordinated suicide attacks happened in the central London during the morning rush hour, where the civilians were targeted using the public transport system, (c) *Boston marathon bombing (April 15, 2013):* During the Boston Marathon, two pressure cooker bombs exploded, that killed three people and injured 264 others. Image source: *http://images.google.com*, information source: *http://en.wikipedia.org/*.

Choi, & Kong, 2010; Popoola & Wang, 2012). The need for video surveillance systems can be well described using the example of popular bombing tragedies, such as the Madrid, London and Boston marathon bombing tragedies, happened in 2004, 2005 and 2013 respectively, as illustrated in Figure 2.2. The tragedies would not have been critical had there been an intelligent video surveillance system installed that can automatically detect the abnormal human behavior in the public areas. Moreover, if the video surveillance system was trained to detect the event early, the situation could have been possibly controlled in a timely manner.

Table 2.1: Highlight on the survey papers on HMA (1994 till present).

Survey paper	Author	Title	Description	Year
Aggarwal, Cai, Liao, and Sabata (1994)	J.K. Aggarwal, Q. Cai, W. Liao & B. Sabata	Articulated and elastic non-rigid motion: a review	This is the earliest survey on HMA, and discusses different methods used in the articulated and non-rigid human body motion.	1994
Cédras and Shah (1995)	C. Cedras & M. Shah	Motion-based recognition: a survey	This paper reviews several methods for motion extraction. The main focus is on action recognition, body parts recognition and body configuration estimation.	1995
Aggarwal and Cai (1997)	J.K. Aggarwal & Q. Cai	Human motion analysis: a review	This paper focuses on the analysis of human body parts motion, human tracking from a single view or multiple camera perspectives, and human activities recognition from video.	1997
Gavrila (1999)	D.M. Gavrila	The visual analysis of human movement: a survey	Various methodologies for visual analysis of human movements are discussed that are grouped into 2D and 3D approaches.	1999
Pentland (2000)	A. Pentland	Looking at people: sensing for ubiquitous and wearable computing	The state-of-the-art of "looking at people" have been reviewed with focus on surveillance monitoring and person identification.	2000
Moeslund and Granum (2001)	T.B. Moeslund & E. Granum	A survey of computer vision-based human motion capture	This paper surveys the computer vision-based human motion capture, and presents a general view on the taxonomy of system functionalities: initialization, tracking, pose estimation and recognition.	2001

Table 2.1 (continued): Highlight on the survey papers on HMA (1994 till present).

Survey paper	Author	Title	Description	Year
L. Wang, Hu, and Tan (2003)	L. Wang, W. Hu & T. Tan	Recent Developments in Human Motion Analysis	Three major issues in human motion analysis have been discussed i.e. human detection, tracking and activity understanding.	2003
Hu, Tan, et al. (2004)	W. Hu, T. Tan, L. Wang & S. Maybank	A survey on visual surveillance of object motion and behaviors	This paper surveyed the recent developments in visual surveillance of object motion and behaviors in dynamic scenes, and analyzed potential research directions.	2004
Moeslund, Hilton, and Krüger (2006)	T. B. Moeslund, A. Hilton, & V. Kruger	A survey of advances in vision-based human motion capture and analysis	The recent trends in video-based human motion capture and analysis have been discussed.	2006
Poppe (2007)	R. Poppe	Vision-based human motion analysis: An overview	This paper presents an overview on HMA with two phases: modeling and estimation. Modeling deals with the construction of likelihood function, and estimation aims at finding the most likely pose given the likelihood surface.	2007
Turaga, Chellappa, Subrahmanian, and Udrea (2008)	P. Turaga, R. Chellappa, V. Subrahmanian & O. Udrea	Machine recognition of human activities: A survey	The problem of representation, recognition and human activity learning from video have been addressed.	2008

Table 2.1 (continued): Highlight on the survey papers on HMA (1994 till present).

Survey paper	Author	Title	Description	Year
Ji and Liu (2010)	X. Ji & H. Liu	Advances in view-invariant human motion analysis: A review	The recognition of actions and poses have been emphasized with main focus on human detection, view-invariant pose representation and estimation, and behavior understanding.	2010
Poppe (2010)	R. Poppe	A survey on vision-based human action recognition	This paper presents an overview on the recent advances in vision-based human action recognition. The challenges faced have been addressed, along with a discussion on the limitations of the state-of-the-art methods.	2010
Candamo, Shreve, Goldgof, Sapper, and Kasturi (2010)	J. Candamo, M. Shreve, D. Goldgof, D. Sapper, & R. Kasturi	Understanding transit scenes: A survey on human behavior-recognition algorithms	Automatic behavior recognition techniques have been surveyed in this paper, with main focus on human activity surveillance in transit applications.	2010
Aggarwal and Ryoo (2011)	J. K. Aggarwal & M. S. Ryoo	Human activity analysis: A review	This paper discusses the methodologies developed for simple human actions and the high-level human activities.	2011
Weinland, Ronfard, and Boyer (2011)	D. Weinland, R. Ronfard & E. Boyer	A survey of vision-based methods for action representation, segmentation and recognition	This work focused on the methods for classifying full body motions e.g. kicking, punching and waving. Furthermore, categorized them according to spatial and temporal structure of actions, action segmentation from an input stream of visual data and view-invariant representation of actions.	2011

Table 2.1 (continued): Highlight on the survey papers on HMA (1994 till present).

Survey paper	Author	Title	Description	Year
Holte, Tran, Trivedi, and Moeslund (2011)	M.B. Holte, T.B. Moeslund, C. Tran & M.M. Trivedi	Human action recognition using multiple views: A comparative perspective on recent developments	This paper presents a comparative study on the recent multi-view 2D and 3D approaches for HMA.	2011
Lara and Labrador (2013)	O. Lara & M. Labrador	A survey on human activity recognition using wearable sensors	Human activity recognition is surveyed based on the wearable sensors. Several systems were qualitatively evaluated in terms of recognition performance, energy consumption, and flexibility etc.	2013
L. Chen, Wei, and Ferryman (2013)	L. Chen, H. Wei & J. Ferryman	A survey of human motion analysis using depth imagery	This paper presents a review on the use of depth imagery for human activity analysis (e.g. the Microsoft Kinect).	2013
Cristani, Raghavendra, Del Bue, and Murino (2013)	M. Cristani, R. Raghavendra, A. Del Bue & V. Murino	Human behavior analysis in video surveillance: A social signal processing perspective	This paper reviews the automated surveillance of human activities from the social signal processing perspective. For example, facial expressions and gazing, vocal characteristics, body posture and gestures, etc.	2013
Chaquet, Carmona, and Fernández-Caballero (2013)	J. M. Chaquet, E. J. Carmona & A. F.-Caballero	A survey of video datasets for human action and activity recognition	A detailed survey of the important video-based human activity and action recognition datasets have been presented.	2013

Table 2.1 (continued): Highlight on the survey papers on HMA (1994 till present).

Survey paper	Author	Title	Description	Year
G. Guo and Lai (2014)	G. Guo & A. Lai	A survey on still image based human action recognition	A comprehensive survey of the research works on still image-based action recognition is conducted.	2014
Gowsikhaa, Abirami, and Baskaran (2014)	D. Gowsikhaa, S. Abirami & R. Baskaran	Automated human behavior analysis from surveillance videos: a survey	Presents a survey on research on human behavior analysis from surveillance videos, with a scope of analyzing the capabilities of the state-of-art methodologies with special focus on semantically enhanced analysis.	2014
Rautaray and Agrawal (2015)	S. S. Rautaray & A. Agrawal	Vision based hand gesture recognition for human computer interaction: a survey	Provides an analysis of existing literature related to gesture recognition systems for human computer interaction by categorizing it under different key parameters.	2015
Dawn and Shaikh (2015)	D. D. Dawn & S. H. Shaikh	A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector	Presents a comprehensive review on STIP-based methods for human action recognition.	2015
T. Li et al. (2015)	T. Li, H. Chang, M. Wang, B. Ni, R. Hong & S. Yan,	Crowded Scene Analysis: A Survey	Provides a survey on the state-of-the-art techniques for crowd scene analysis.	2015
C. H. Lim et al. (2015)	C. H. Lim, E. Vats & C. S. Chan	Fuzzy human motion analysis: A review	Presents a survey of fuzzy set oriented methods for human motion analysis	2015

Table 2.2: Criterion on which the survey papers on HMA from 1994 till present emphasized on. (A '-' indicates that the topic has not been discussed comprehensively in the corresponding paper, but possibly touched indirectly in the contents.)

Year	Survey paper	Human detection	Motion tracking	Behavior understanding	Multi-view	Feature extraction	Datasets	Application
1994	Aggarwal et al. (1994)	-	✓	✓	-	✓	-	-
1995	Cédras and Shah (1995)	-	✓	✓	-	✓	-	-
1997	Aggarwal and Cai (1997)	✓	✓	✓	✓	✓	-	-
1999	Gavrila (1999)	✓	✓	✓	✓	✓	-	✓
2000	Pentland (2000)	✓	✓	✓	-	✓	-	-
2001	Moeslund and Granum (2001)	✓	✓	✓	-	✓	-	✓
2003	L. Wang et al. (2003)	✓	✓	✓	✓	-	-	✓
2004	Hu, Tan, et al. (2004)	✓	✓	✓	✓	-	-	✓
2006	Moeslund et al. (2006)	✓	✓	✓	✓	-	-	-
2007	Poppe (2007)	✓	✓	-	-	✓	-	-
2008	Turaga et al. (2008)	✓	-	✓	-	✓	-	✓

Table 2.2 (continued): Criterion on which the survey papers on HMA from 1994 till present emphasized on. (A '-' indicates that the topic has not been discussed comprehensively in the corresponding paper, but possibly touched indirectly in the contents.)

Year	Survey paper	Human detection	Motion tracking	Behavior understanding	Multi-view	Feature extraction	Datasets	Application
2010	Ji and Liu (2010)	✓	-	✓	✓	-	✓	-
2010	Poppe (2010)	-	-	✓	✓	✓	✓	-
2010	Candamo et al. (2010)	✓	✓	✓	-	-	-	-
2011	Aggarwal and Ryoo (2011)	-	-	✓	-	✓	✓	✓
2011	Weinland et al. (2011)	✓	-	✓	✓	✓	✓	-
2011	Holte et al. (2011)	-	-	✓	✓	✓	✓	-
2013	Lara and Labrador (2013)	-	-	✓	-	✓	✓	-
2013	L. Chen et al. (2013)	✓	✓	✓	-	-	✓	-
2013	Cristani et al. (2013)	✓	✓	✓	-	-	-	✓
2013	Chaquet et al. (2013)	-	-	-	-	-	✓	-
2014	G. Guo and Lai (2014)	✓	-	✓	✓	✓	✓	✓

Table 2.2 (continued): Criterion on which the survey papers on HMA from 1994 till present emphasized on. (A '-' indicates that the topic has not been discussed comprehensively in the corresponding paper, but possibly touched indirectly in the contents.)

Year	Survey paper	Human detection	Motion tracking	Behavior understanding	Multi-view	Feature extraction	Datasets	Application
2014	Gowsikhaa et al. (2014)	✓	✓	✓	-	-	✓	✓
2015	Rautaray and Agrawal (2015)	✓	✓	✓	-	✓	-	✓
2015	Dawn and Shaikh (2015)	✓	-	✓	-	✓	✓	-
2015	T. Li et al. (2015)	✓	✓	✓	-	✓	✓	-
2015	C. H. Lim et al. (2015)	✓	✓	✓	✓	✓	✓	✓

As highlighted in Table 2.1, the significance and popularity of HMA attracted several researchers and hence a number of survey papers have been published in the literature. The earliest survey paper was by Aggarwal et al. (1994), that focused on different methods employed in the articulated and non-rigid human body motion. An overview on the motion extraction methods using the motion capture systems was presented in Cédras and Shah (1995). This survey was focused mainly on action recognition, individual body parts recognition, and body configuration estimation. A similar taxonomy was used in Aggarwal and Cai (1997), where different labels were assigned for the three classes, and the classes were further divided into subclasses yielding a more comprehensive taxonomy. An interesting survey was conducted by Gavrila (1999), where the applications of visual analysis of human movements was reviewed. Their taxonomy covered the 2D and 3D approaches with and without the explicit shape models.

The most recent papers include Rautaray and Agrawal (2015); Dawn and Shaikh (2015); T. Li et al. (2015); C. H. Lim et al. (2015). Rautaray and Agrawal (2015) provided an analysis of existing literature related to gesture recognition systems for human computer interaction by categorizing it under different key parameters. A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector was presented in Dawn and Shaikh (2015). The state-of-the-art techniques for crowd scene analysis were reviewed in T. Li et al. (2015). And lastly, C. H. Lim et al. (2015) presented a survey on the fuzzy set oriented methods for HMA. Table 2.1 and 2.2 summarizes the available survey papers on HMA from 1994 till present, and the criterion on which these papers emphasized.

In general, three main steps are involved in a HMA system: human detection and modeling, human motion tracking, and human action recognition. As illustrated in Figure 1.3, there may exist uncertainties at each step in a HMA system. For example, while performing human detection and modeling, there may exists background noise, shadows, occlusions etc. that can affect the detection accuracy. Also, humans differ in their body sizes, and therefore proper generalization on the human body size variation is required. This can otherwise affect the process of building human model for further processing. Nonetheless, uncertainties at this level can affect the feature extraction process that serves as the prerequisite for human motion tracking and action recognition.

Furthermore, a sophisticated human motion tracking system should be well-trained to handle the uncertainties such as viewpoint variations. This means that since human can perform an action irrespective of the current position, angles, etc., therefore HMA system should be able to handle the variations in the camera viewpoints. If such uncertainties are not taken into account, they can affect the overall system performance.

Another source of uncertainty that can affect the HMA system is the classification ambiguity or vagueness to accurately detect an action due to high degree of similarities amongst different action classes. For example, in Figure 1.3, it is difficult to distinguish between 'walk', 'jog' and 'run' actions due to similar characteristics. The main reason behind this problem is the binary classification output enforced on the system, where an action can belong to one class only at a time, with zero tolerance to uncertainty.

An efficient algorithm should be able to handle even the minutest level of uncertainty for a reliable decision making as the cumulated errors can deteriorate the overall system performance. Fuzzy set theory (Zadeh, 1965) has inherent capability in handling the uncertainties, and therefore can help in dealing with the above discussed limitations of the conventional HMA system. Hence, this gave rise to a new research direction - "fuzzy HMA", as reviewed in the following section.

2.2 Fuzzy Human Motion Analysis

Before reviewing the fuzzy set oriented approaches for HMA, the main advantages of using fuzzy approach for HMA is required to be discussed. Some important factors are identified that make fuzzy approaches successful in improving the overall system performance. These include, firstly, the ability of the fuzzy approaches to assign soft boundary instead of hard labels. Secondly, the linguistic support provided by the fuzzy approaches to represent the measurement boundaries. Lastly, the flexibility of the fuzzy system to adapt to various system designs. These important factors are discussed as follows:

(a) Soft boundary assignment:

Human reasoning is a mysterious phenomenon that scientists are trying to simulate with machines in the past few decades. With the knowledge that "soft" boundaries exist in concepts formation of human beings, fuzzy set theory (Zadeh, 1965) has emerged as one of the most important methodologies in capturing human motion. In general, fuzzy approach assigns "soft" boundaries, or in other words perform "soft labeling", where a subject can be associated with many possible classes with a certain degree of confidence. As such, the fuzzy representation is more beneficial than the ordinary (crisp) representations. This is because it can represent not only the information stated by a well-determined real interval, but also the knowledge embedded in the soft boundaries of the interval. Thus, it removes, or largely weakens the boundary interpretation problem achieved through the description of a gradual rather than an abrupt change in the degree of membership, closer to how humans make decisions and interpret things in the real world.

This is also supported by a few notable literary works. For example, Bezdek (1992) in their review on computing with uncertainties emphasized on the fact that the integration of fuzzy models always improve the computer performance in pattern recognition problems.

Similarly, Huntsberger, Rangarajan, and Jayaramamurthy (1986); Yager (2002) presented a survey on how to effectively represent the uncertainties using the Fuzzy Inference Structure (FIS). Nevertheless, there are a few studies reported on the type-2 FIS in this regards. H. Wu and Mendel (2002); D. Wu and Mendel (2007) explained on how to design an interval type-2 FIS using the uncertainty bounds, and introduced the measurement of uncertainty for interval type-2 fuzzy sets using the information such as centroid, cardinality, fuzziness, variance and skewness. A comprehensive review on handling the uncertainties in pattern recognition using the type-2 fuzzy approach was provided by Zeng and Liu (2006).

(b) Linguistic support:

Another worth highlighting aspect of human behavior is the way they interpret things in the natural scenarios. Human beings mostly employ words in reasoning, arriving at conclusions expressed as words from the premises in a natural language, or having the form of mental perceptions. As used by humans, words have fuzzy denotations. Therefore, modeling the uncertainties in a natural format for humans (i.e. linguistic summarizations) can yield more succinct description of human activities. Inspired from this, HMA can be modeled efficiently by representing an activity in linguistic terms. This concept was initiated by Zadeh (1996), where words can be used in place of numbers for computing and reasoning (like done by human), commonly known as Computing with Words (CWW).

In CWW, a word is viewed as a fuzzy set of points drawn together by similarity, with the fuzzy set playing the role of a fuzzy constraint on a variable. There are two major imperatives for CWW (Zadeh, 1996). Firstly, CWW is necessary when the available information is too imprecise to be justified using numbers. Secondly, when there is a tolerance for imprecision that can be exploited to achieve tractability, robustness, low solution cost, and better rapport with reality. This concept of using CWW i.e. linguistic support to represent the measurement boundaries can be applied in real-world scenarios. For example, consider the human activities: walking and running, which can be inferred using a simple cue i.e. the speed of a person. Different levels of speed can be modeled by using the linguistic terms such as ‘very slow’, ‘slow’, ‘moderate’, ‘fast’, and ‘very fast’, instead of representing in numerical terms. The use of linguistic terms provide the capability to perform human like reasoning such as the feasibility of defining rules for the inference process. With the integration of the linguistic support in the FIS, the computational complexity of the numeric labeling and the imprecision problem in the interpretation stage are also suppressed. Furthermore, the linguistic terms are more understandable where they mimic how human interpret things and make decisions.

The concept of linguistic support is rooted in several papers. For example, in Zadeh (1973) the concept of a linguistic variable and the granulation was introduced. Besides that, Zadeh (1996) discussed the role played by fuzzy logic in CWW and vice-versa. An interesting work by Rubin (1999) defined CWW as a symbolic generalization of fuzzy logic. Recently, several papers have been published that utilized the concept of linguistic summarization in the fuzzy system, and have been successfully applied in the real-world applications. For example, the works by Anderson, Luke, et al. (2009a); Trivino and van der Heide (2008); Kacprzyk and Yager (2001); Anderson, Keller, Anderson, and Wescott (2011); Wilbik, Keller, and Alexander (2011); Wilbik and Keller (2013), where a complete sentence is preferable as an output, instead of numerical data or a crisp answer like in a conventional decision making systems. For instance, “the resident has fallen in the living room and is down for a long time”. Such succinct linguistic summarization output is more understandable and closest to the natural answer.

(c) Flexibility of the fuzzy system:

Another advantage of the fuzzy approach, especially those that utilize the knowledge based system (fuzzy rules) such as the FIS, is that they possess the flexibility and feasibility to adapt to various system designs. The conventional approaches designed their algorithms to be well-fitted to solve solely some specific problems with low or no extendibility. The world is changing rapidly with the headway of technologies. The flexibility to adapt to such changes is one of the major concerns for a good and long lasting system. Fortunately, the fuzzy approaches allow the alterations to serve the purpose. In addition, the alterations can be made easily on the knowledge base by designing the fuzzy rules.

The knowledge base that comprises of all the rules is considered as the most crucial part of a decision making system where it functions as the “brain” of the overall system. As human growth together with knowledge is capable of making better decisions, similarly if a decision making system is provided with sophisticated knowledge, it can deal with the problems in a better manner. The FIS consists of a knowledge base where it can store a number of conditional “IF-THEN” rules that are used for the reasoning process in a specific problem domain. These rules are easy to write, and as many rules as necessary can be supplied to describe the problem adequately. For example, consider the problem of identifying different human activities e.g. running. Rules can be designed to infer the running activity using a simple cue (speed), as following:

Rule 1: IF (speed is FAST) THEN (person is RUNNING)

Rule 2: IF (speed is MODERATE) THEN (person is NOT RUNNING)

However, in real-world scenarios, various factors can affect the speed of a person such as the height, body size, etc. Therefore, in order to make the system closer to natural solution, these rules are needed to be modified accordingly. Intuitively, if one may observe the running styles of a tall person and a shorter person, due to difference in the step size of their feet, the taller person tends to run with a faster speed as compared to the shorter person, running with moderate speed. However, both are performing the running activity, but with different rules. This situation can be modeled by modifying the “Rule 2” as follows:

Rule 2.1: IF (HEIGHT is TALL) & (SPEED is MODERATE) THEN (person is NOT RUNNING)

Rule 2.2: IF (HEIGHT is SHORT) & (SPEED is MODERATE) THEN (person is RUNNING)

Similarly, the body size can also affect the speed of a person, and can be modeled using flexible fuzzy rules that can be easily added, modified or deleted according to the objective of the system.

In a conventional FIS, most of these rules are built with the help of human expert knowledge. For example, such experts can be doctor, police, forensic expert or researcher, etc. The information that they provide is considered to be the most reliable one as they build it based on their real life experiences and historical analysis. However, human intervention in an intelligent system is becoming a threat due to the heuristic and subjectivity of human decisions. Therefore, automated learning systems have emerged and widely employed in the research society, encouraging learning and generation of fuzzy rules automatically. Several works in the literature have reported efficient methods for the automatic generation of the fuzzy rules such as L.-X. Wang and Mendel (1992); Rhee and Krishnapuram (1993); X. Wang, Wang, Xu, Ling, and Yeung (2001); T. W. Cheng, Goldgof, and Hall (1995); Cordón, Herrera, and Villar (2001).

For example, L.-X. Wang and Mendel (1992) proposed a method for generating the fuzzy rules by learning from examples, more specifically by the numerical data. Similarly, Rhee and Krishnapuram (1993) presented an alternative method to generate the fuzzy rules automatically from the training data with their rules defined in the form of possibility, certainty, gradual, and unless rules. A new approach called the fuzzy extension matrix was proposed in X. Wang et al. (2001), which incorporated the fuzzy entropy to search for the paths, and generalized the concept of the crisp extension matrix. Their method is capable of handling the fuzzy representation and tolerating the noisy or missing data. Fuzzy c-means (FCM) and its variants (e.g. multi-stage random sampling) with fast performance have also been adopted in the fuzzy rule generation, such as the work by T. W. Cheng et al. (1995). Apart from that, there are works reported in the fuzzy rule generation incorporated with other machine learning techniques. Mitra and Hayashi (2000) provided an exhaustive survey on the neuro-fuzzy rule generation algorithms, while Cordón et al. (2001) presented an approach to automatically learn the fuzzy rules by incorporating the genetic algorithm.

2.2.1 Overall taxonomy of fuzzy HMA

The human motion can be conceptually classified into three broad levels: Low-Level (LoL), Mid-Level (MiL), and High-Level (HiL) HMA. LoL HMA is the background (or foreground) subtraction which contributes in the pre-processing of the raw images to discover the areas of interest, such as the human region. MiL HMA is the object tracking, and this level prepares data for activity recognition. HiL HMA deals with the human behavior understanding, where the main aim is the classification of human motion patterns into various activity classes. Figure 2.3 represents the general taxonomy of fuzzy HMA and the fuzzy approaches that are most commonly employed in the literature.

2.2.2 Low-level HMA

Human detection is the enabling step in almost every low-level vision-based HMA system before the higher level of processing steps such as tracking and behavior understanding

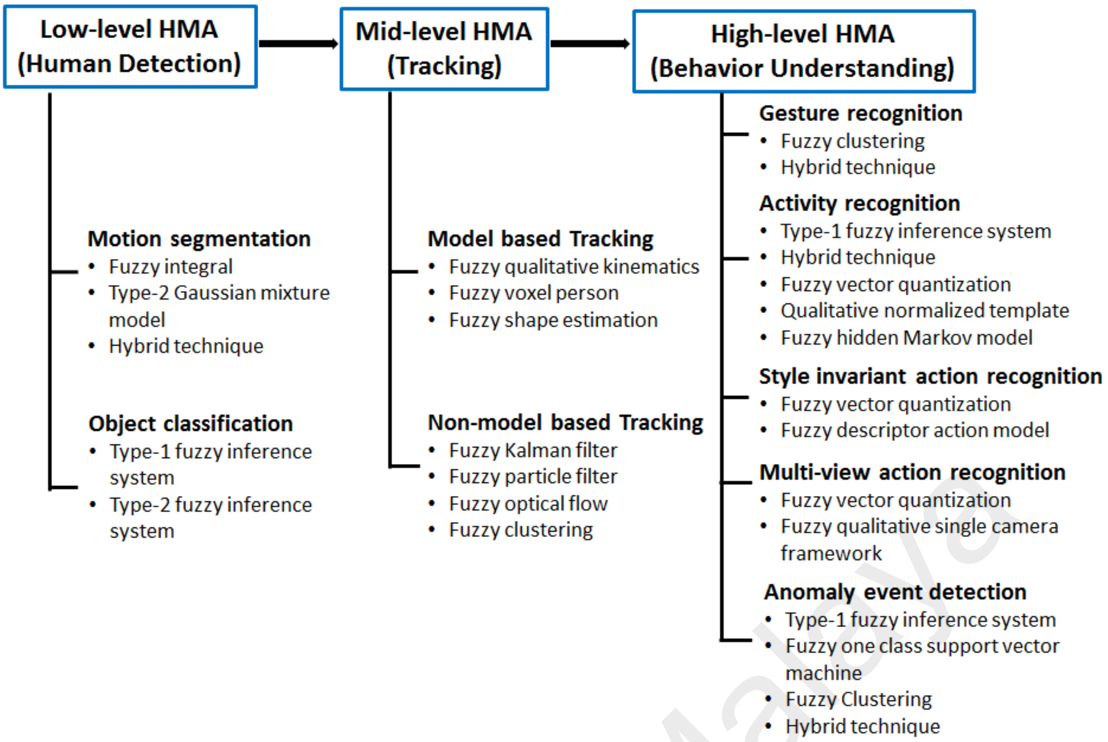

Low-level HMA
(Human Detection)
Mid-level HMA
(Tracking)
High-level HMA
(Behavior Understanding)
Motion segmentation
• Fuzzy integral
• Type-2 Gaussian mixture model
• Hybrid technique
Object classification
• Type-1 fuzzy inference system
• Type-2 fuzzy inference system
Model based Tracking
• Fuzzy qualitative kinematics
• Fuzzy voxel person
• Fuzzy shape estimation
Non-model based Tracking
• Fuzzy Kalman filter
• Fuzzy particle filter
• Fuzzy optical flow
• Fuzzy clustering
Gesture recognition
• Fuzzy clustering
• Hybrid technique
Activity recognition
• Type-1 fuzzy inference system
• Hybrid technique
• Fuzzy vector quantization
• Qualitative normalized template
• Fuzzy hidden Markov model
Style invariant action recognition
• Fuzzy vector quantization
• Fuzzy descriptor action model
Multi-view action recognition
• Fuzzy vector quantization
• Fuzzy qualitative single camera framework
Anomaly event detection
• Type-1 fuzzy inference system
• Fuzzy one class support vector machine
• Fuzzy Clustering
• Hybrid technique

Figure 2.3: The general taxonomy of fuzzy HMA. It is represented into three broad levels: Low-level, Mid-level and High-level HMA, along with the fuzzy approaches that are most commonly employed in the literature.

can be performed. Technically, human detection aims at locating and segmenting the regions bounding the people from rest of the image. This process usually involves motion segmentation and object classification.

2.2.2.1 Motion segmentation

Motion segmentation aims at separating the moving objects from the natural scene. The extracted motion regions are vital for the next level of processing. For example, it relaxes the tracking complexity as only the pixels with changes are considered in the process. However, some critical situations in the real-world environment such as the illumination changes, dynamic scene movements (e.g. rainy weather, waving tree, rippling water etc.), camera jittering, and shadow effects make it a daunting task. Herewith, the fuzzy approaches are reviewed that addressed the background subtraction problems.

Background subtraction is one of the popular motion segmentation algorithms that has received much attention in the HMA system. This is due to the usefulness of its output that is capable of preserving the shape information, and helps in extracting motion and contour information (Bobick & Davis, 2001; Weinland, Ronfard, & Boyer, 2006; Lewandowski, Makris, & Nebel, 2010). In general, background subtraction is to differentiate between the image regions which have significantly different characteristics from the background image (normally denoted as the background model). A good background subtraction algorithm comprises of a background model that is robust to the environmental changes, but sensitive to identify all the moving objects of interest. There are some fuzzy techniques that endowed this capability in the background subtraction which are highlighted in Table 2.3.

2.2.2.2 Object classification

The outcome from the motion segmentation usually result in a rough estimation of the moving targets in a natural scene. These moving targets in a natural scene can be shadow, vehicle, flying bird, etc. Before the region is further processed at the next level, it is very important to verify and refine the interest object by eliminating the unintended objects. There exists some fuzzy approaches in the literature that are significant for human object classification. For example, Type-1 FIS and Type-2 FIS, as highlighted in Table 2.4, along with the problems and the notion of using fuzzy approaches.

2.2.3 Mid-level HMA

Human detection is followed by tracking of human movements over time. Human motion tracking is an important step as it prepares data for higher level tasks such as anomaly event detection, human activity recognition, etc. The aim is to reliably and efficiently track the object of interest (e.g. human) from a video. In general, motion tracking is classified as model based and non-model based tracking.

Table 2.3: A summary of research works in motion segmentation (LoL HMA) using fuzzy techniques.

LoL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Motion segmentation	Several sources of uncertainty that affect the background model or the foreground object includes the illumination changes, camera jitter, dynamic scene movements and shadows.	H. Zhang and Xu (2006); El Baf, Bouwmans, and Vachon (2008b, 2008a); Balcilar and Sonmez (2013)	The crisp decision problem can be relaxed using the fuzzy aggregation method that allows information fusion from a variety of sources.	Fuzzy integral
	The distribution in an ordinary GMM based background modeling method is not reflected accurately using insufficient and noisy training data.	El Baf, Bouwmans, and Vachon (2008c, 2009); Bouwmans and El Baf (2009); Zhao, Bouwmans, Zhang, and Fang (2012)	Instead of crisp values, the uncertainty in GMM is bounded with interval mean and standard deviation. In order to handle higher level of uncertainty, type-2 fuzzy set is utilized.	Type-2 fuzzy GMM
	In the background subtraction algorithms, the problem is to determine the optimum parameters for fuzzy inference process e.g. membership function, threshold values, etc.	Lin, Chung, and Sheu (2000); Maddalena and Petrosino (2010); Z. Li, Liu, and Zhang (2012); Calvo-Gallego, Brox, and Sánchez-Solano (2013); Shakeri, Deldari, Foroughi, Saberi, and Naseri (2008)	A hybrid technique developed with the integration of fuzzy and machine learning techniques are found to efficiently learn the optimum parameters with good system performance.	Hybrid technique

Table 2.4: A summary of research works in object classification (LoL HMA) using fuzzy techniques.

LoL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Object classification	The confusion between the human and non-human objects, and the unintended objects attached to the human region causes the uncertainty in the classification tasks.	Mahapatra, Mishra, Sa, and Majhi (2013); See, Lee, and Hanmandlu (2005); Chowdhury and Tripathy (2014); X. Chen, He, Anderson, Keller, and Skubic (2006); X. Chen, He, Keller, Anderson, and Skubic (2006)	Type-1 FIS is able to model the uncertainty in the features data as the membership function, and perform inference using the fuzzy rules to achieve better classification results.	Type-1 FIS
	Type-1 FIS is insufficient in handling higher level of uncertainties, and therefore can result in misclassification of objects, further affecting the silhouette extraction process.	Yao, Hagras, Al Ghazzawi, and Alhaddad (2012)	Type-2 fuzzy set can handle higher dimensions of uncertainty, and helps in obtaining smoother classification results.	Type-2 FIS

2.2.3.1 Model based tracking

In model based tracking, human body models are adopted to model the complex, non-rigid structure of the human body. The human body models include the stick figures, 2D and 3D motion description models, etc. For example, as presented in Y. Guo, Xu, and Tsuji (1994); Leung and Yang (1995); Iwai, Ogaki, and Yachida (1999); Silaghi, Plänkers, Boulic, Fua, and Thalmann (1998); Niyogi and Adelson (1994); Ju, Black, and Yacoob (1996); Rohr (1994); Wachter and Nagel (1997); Rehg and Kanade (1995); Kakadiaris and Metaxas (1996). The human body is represented as a combination of line segments or sticks connected by joints in the stick figure model (Y. Guo et al., 1994; Leung & Yang, 1995; Iwai et al., 1999; Silaghi et al., 1998). The 2D models use 2D ribbons or blobs to represent the human body (Leung & Yang, 1995; Niyogi & Adelson, 1994; Ju et al., 1996). While the 3D models represent the human body in much more detailed manner by using spheres, cones, ellipses, cylinders, etc., as presented in Rohr (1994); Wachter and Nagel (1997); Rehg and Kanade (1995); Kakadiaris and Metaxas (1996).

However, human motion tracking is not an easy task due to the complex non-rigid structure of the human body that consist of a number of joints and each body part has the freedom to move in several directions. This can result in self-occlusions of the body parts, and the issue can be handled well using the 3D models. Furthermore, other factors that can affect the tracking performance includes cluttered background, monotone clothes, illumination changes, etc. that even 3D models fail to handle (Ning, Tan, Wang, & Hu, 2004). As a solution, the fuzzy techniques such as the fuzzy qualitative kinematics, the fuzzy voxel person, and the fuzzy shape estimation are employed in the model based tracking algorithms to address the problem statements (Table 2.5).

Table 2.5: A summary of research works in model based tracking (MiL HMA) using fuzzy techniques.

MiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Model based tracking	The kinematic chain represented in a crisp manner suffers from the precision problem, and the cumulative errors can directly affect the performance of the tracking process.	H. Liu, Brown, and Coghill (2008b); Chan and Liu (2009); H. Liu, Coghill, and Barnes (2009); H. Liu, Brown, and Coghill (2008a); Chan, Liu, Brown, and Kubota (2008)	In the kinematic chain representation, the uncertainty can be handled in a natural way by integrating the fuzzy set theory with the fuzzy qualitative reasoning. The resultant fuzzy qualitative kinematics approach provides a solution to the precision problem by eliminating the hard boundary assignment in the measurement space that can tolerate the offset errors.	Fuzzy qualitative kinematics
	Due to the object's position and the camera location, the information collected using crisp voxel person model can be imprecise and inaccurate. Crisp approach works fine in multi-camera environment, but it is not feasible due to high cost and limited space.	Anderson, Luke, et al. (2009b); Anderson, Luke III, Stone, and Keller (2009); Anderson, Luke, et al. (2009a)	Fuzzy voxel person is able to model different types of uncertainties associated with the construction of the voxel person by using the membership functions, employing only a few cameras and a minimal prior knowledge about the object.	Fuzzy voxel person

Table 2.5 (continued): A summary of research works in model based tracking (MiL HMA) using fuzzy techniques.

MiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Model based tracking	In shape based (blob) tracking, the imperfect image segmentation techniques result in multiple blobs generation for a single object because of the image irregularities, shadows, occlusions, etc. While in the multiple object tracking, recovering from the overlapping regions is a big challenge.	García, Molina, Besada, Portillo, and Casar (2002); Garcia, Patricio, Berlanga, and Molina (2011)	FIS is applied to perform the fuzzy shape estimation to achieve a better tracking performance by taking into account the uncertainty in shape estimation. If the shape is uncertain, the tracking will be locked and it will be recovered once the confidence becomes higher. This is to prevent the tracking errors caused by the uncertain shapes.	Fuzzy shape estimation

2.2.3.2 Non-model based tracking

In non-model based tracking, the randomly dispersed points are used to represent the objects. The association between these points depend on the object's characteristics and behavior. However, it is a complex task due to the presents of occlusions, new object entries, misdetections, etc. that can generate permanent tracking errors. The fuzzy approaches explicitly handles the uncertainties involved in establishing point correspondence between the object motions, and are commonly employed in the non-model based object tracking. For example, the fuzzy Kalman filter, fuzzy particle filter, fuzzy optical flow and fuzzy clustering. A summary of research works in non-model based tracking using fuzzy techniques is presented in Table 2.6.

2.2.4 High-level HMA

The final aim of the HMA system is to perform human behavior understanding. In this section, the employability of the fuzzy techniques to perform human behavior understanding is studied with main focus on hand gesture recognition, activity recognition, style invariant action recognition, multi-view action recognition, and anomaly event detection.

2.2.4.1 Hand gesture recognition

The aim of a gesture recognition system is to recognize meaningful expressions of the human motion. The applications of gesture recognition include sign language recognition, medical rehabilitation, virtual reality, etc. (Lyons, Budynek, & Akamatsu, 1999). Gesture recognition is important for building efficient and intelligent human-computer interaction applications (Y. Wu & Huang, 1999) where the system can be controlled from a distance without screen touching, or cursor movements.

However, there exists several sources of uncertainties and ambiguities that can affect a gesture recognition system. For example, dynamic lighting conditions, complex back-

Table 2.6: A summary of research works in non-model based tracking (MiL HMA) using fuzzy techniques.

MiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Non-model based tracking	Conventional Kalman filter algorithms suffer from the divergence problem and it is difficult to model the complex dynamic trajectories.	Hu, Tan, et al. (2004); I. S. Kim et al. (2010); Ko (2008); Aggarwal and Cai (1997); Gavrila (1999); G. Chen, Xie, and Shieh (1998); Kobayashi, Cheok, Watanabe, and Munekata (1998)	Fuzzy Kalman filters are capable of solving the divergence problem by incorporating the FIS, and are more robust against the streams of random noisy data inputs.	Fuzzy Kalman filter
	Particle filters suffer from the tradeoff between the accuracy and computational cost as its performance usually relies on the number of particles. This means more number of particles will improve the accuracy, but at the same time increases the computational cost.	Chan and Liu (2009); Chan et al. (2008); H. Wu, Sun, and Liu (2008); Yoon, Cheon, and Park (2013); Kamel and Badawy (2005); Y.-J. Kim, Won, Pak, and Lim (2007)	The fuzzy particle filter effectively handles the system complexity by compromising the low number of particles that were used while retaining the tracking performance.	Fuzzy particle filter
	Random noises in optical flow field due to the sources of disturbances in a natural scene (e.g. dynamic background) affects the tracking performance.	Bhattacharyya, Maulik, and Dutta (2009); Bhattacharyya and Maulik (2013)	In order to efficiently filter the random noises in the optical flow field, fuzzy hostility index has been employed.	Fuzzy optical flow

Table 2.6 (continued): A summary of research works in non-model based tracking (MiL HMA) using fuzzy techniques.

MiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Non-model based tracking	Multiple object tracking using hard clustering algorithms (e.g. K-means) involve high computational cost, and they are incapable of handling problems such as occlusions and pervasive disturbances.	Xie, Hu, Tan, and Peng (2004)	FCM tracking algorithm uses soft computing techniques and offers more meaningful and stable performance. Faster processing speed is obtained with the integration of component quantization filtering with FCM.	Fuzzy clustering

grounds, deformable human limb shapes, etc. Also, “pure” gestures are seldom elicited, as human normally demonstrate “blends” of these gestures (Mitra & Acharya, 2007). Among several solutions available to tackle the issue, fuzzy clustering algorithms and the hybrid technique of machine learning with fuzzy set theory are often employed, and help in achieving better system performance, as illustrated in Table 2.7.

2.2.4.2 Activity recognition

Activity recognition is an important task in the HiL HMA system. The goal of activity recognition is to autonomously analyze and interpret the ongoing human activities and their context from the video data. For example, in the surveillance systems for detecting suspicious actions, or in sports analysis for monitoring the correctness of the athletes' postures.

In recent times, the fuzzy approaches such as type-1 FIS, fuzzy HMM, and hybrid techniques have proved to be beneficial in human activity recognition, with capability of modeling the uncertainties in the feature data. Nonetheless, Fuzzy Vector Quantization (FVQ) and Qualitative Normalized Template (QNT) provide the capability to handle the complex human activities occurring in our daily life, such as walking followed by running, then running followed by jumping, or a hugging activity where two or more people are involved. Table 2.8 discusses the applications of the fuzzy techniques in performing activity recognition, highlighting the problem statements and the reason for employing fuzzy approaches.

2.2.4.3 Style invariant action recognition

Style invariant action recognition refers to recognizing the actions of different people executed in various styles. In general, people differ from one another in their styles of performing an action because of the physical differences that refers to size variations,

Table 2.7: A summary of research works in hand gesture recognition (HiL HMA) using fuzzy techniques.

HiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Hand gesture recognition	The crisp clustering algorithms often produce ineffective clustering results due to sources of uncertainties such as dynamic lighting conditions, complex backgrounds, deformable human limbs' shape, etc.	J. Wachs, Kartoun, Stern, and Edan (2002); J. P. Wachs, Stern, and Edan (2005); X. Li (2003); Verma and Dev (2009)	FCM relaxes the uncertainties involved in the gesture learning and recognition using soft computing technique. Furthermore, FCM reduces the errors as a result of crisp decisions and enhances the overall system performance.	Fuzzy clustering
	In gesture recognition algorithms, the problem is to determine the optimum parameters for fuzzy inference process e.g. membership function, threshold values, etc.	Al-Jarrah and Halawani (2001); Binh and Ejima (2005); Várkonyi-Kóczy and Tusor (2011)	Integration of the fuzzy approaches with machine learning algorithms help in learning the important parameters for the fuzzy system adaptively based on the training data.	Hybrid technique

Table 2.8: A summary of research works in activity recognition (HiL HMA) using fuzzy techniques.

HiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Activity recognition	The performance of a human activity recognition system can be affected due to the presence of uncertainty in the feature data.	Le Yaouanc and Poli (2012); Yao, Hagras, Alhaddad, and Alghazzawi (2014)	FIS is tolerant to the vague feature data and efficiently distinguishes the human motion patterns using flexible fuzzy rules and membership function.	Type-1 FIS
	Sometimes it is difficult to find the optimum membership function values and define the fuzzy rules in the FIS.	Acampora et al. (2012); Hosseini and Eftekhari-Moghadam (2013)	Hybridization of fuzzy logic and machine learning techniques produces optimum membership function and fuzzy rules for human activity recognition.	Hybrid technique
	It is difficult to model the complex activities and continuous human movements over time. And most of the state-of-the-art algorithms assume simple and uniform activities.	Gkalelis, Tefas, and Pitas (2008)	Complex and continuous human movements can be flexibly and efficiently modeled by integrating FVQ with FCM.	FVQ

Table 2.8 (continued): A summary of research works in activity recognition (HiL HMA) using fuzzy techniques.

HiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Activity recognition	The sophisticated tracking algorithms employed for human action recognition often suffer from the trade-off between accuracy and the computational cost.	Chan and Liu (2009); Chan et al. (2008); Chan, Liu, and Lai (2010)	Using QNT, a fuzzy motion template, the complexity involved in the representation of human joints can be relaxed, and as a result complex activity recognition can be performed efficiently.	QNT
	The uncertainties that can exist in the training stage is not handles by the conventional HMM model, and hence degrades the classification performance.	Mozafari, Charkari, Boroujeni, and Behrouzifar (2012)	Fuzzy HMM model applies soft computing in the training stage, and therefore effectively enhances the classification performance of even the similar actions e.g. “walk” and “run”.	Fuzzy HMM

appearances, postures, etc., or the dynamic differences that includes speed variations, motion patterns, etc. (C. H. Lim & Chan, 2013). In order to model the style variations, several notable works have been reported in the literature that utilizes the fuzzy techniques, and have been listed in Table 2.9.

2.2.4.4 Multi-view action recognition

Multi-view action recognition refers to performing an action irrespective of camera viewing angles. In real-world, a human can perform an action at any angle, with no restriction on being frontal parallel to the camera. However, most of the existing works fix the camera angles, and limit the camera view. The problem of view independent action recognition has received much attention in the CV community. Some of the noteworthy works include Ji and Liu (2010); Weinland et al. (2006); Lewandowski et al. (2010). Nonetheless, fuzzy techniques such as the FVQ, and fuzzy qualitative reasoning have been applied in the literature to perform multi-view action recognition, and a summary of research works is presented in Table 2.10.

2.2.4.5 Anomaly event detection

Anomaly detection is important in our daily life. It deals with the problem of discovering patterns in the input data that do not conform to the expected behavior. This can help in inferring abnormal human behavior such as an action or an activity that is not following the routine or different from the normal behavior (Hu, Tan, et al., 2004; Kratz & Nishino, 2009; S. Wu, Moore, & Shah, 2010). For example, in video surveillance scenario, to automatically detect the criminal acts. To understand better, a summary of research works on anomaly event detection using the fuzzy approaches is shown in Table 2.11.

From the extensive review on fuzzy HMA, it can be seen that there exists a number of fuzzy approaches to deal with a given problem statement and sources of uncertainties

Table 2.9: A summary of research works in style invariant action recognition (HiL HMA) using fuzzy techniques.

HiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Style invariant action recognition	People differ from one another in their styles of performing an action, and therefore same action can be performed in different styles by different people. This can lead to difficulty in action learning and recognition.	Iosifidis, Tefas, and Pitas (2011, 2012a)	A person specific fuzzy movement model can be employed that is trained using FVQ to perform style invariant action recognition.	FVQ
	An ordinary descriptor vector can contain only a single value in each vector dimension. This limits its capability to model different styles of human actions.	C. H. Lim and Chan (2013)	Fuzzy descriptor action mode allows each vector dimension to accommodate a set of possible descriptor values, and hence able to model different action styles in a single underlying fuzzy action descriptor.	Fuzzy descriptor action model

Table 2.10: A summary of research works in multi-view action recognition (HiL HMA) using fuzzy techniques.

HiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Multi-view action recognition	A human can perform an action at any angle, with no restriction on being frontal parallel to the camera.	Iosifidis, Tefas, and Pitas (2012a); Iosifidis, Tefas, Nikolaidis, and Pitas (2012); Iosifidis, Tefas, and Pitas (2013, 2012b)	In order to perform view-invariant action recognition, FVQ can be utilized. A multi-view fuzzy motion model can be constructed by utilizing FVQ to generate posture patterns.	FVQ
	View invariant action recognition is assumed to be impractical in the real-world.	C. H. Lim and Chan (2013)	Fuzzy qualitative framework can be used to perform multi-view action recognition within single camera efficiently.	Fuzzy qualitative single camera framework

Table 2.11: A summary of research works in anomaly event detection (HiL HMA) using fuzzy techniques.

HiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Anomaly event detection	The problem in the existing methods is the difficulty to extend the framework to deal with new issues and provide support for new activities.	Anderson et al. (2006); Anderson, Luke, et al. (2009b, 2009a); Anderson, Luke, Keller, and Skubic (2008)	As a solution, FIS can be efficiently used to model the falling activity by utilizing the fuzzy rules that can be easily modified, added, or removed according to the given situation.	Type-1 FIS
	In a fall detection system, the classification performance is highly affected by the imperfect training data.	M. Yu, Naqvi, Rhuma, and Chambers (2011)	In order to indicate the significance of each training sample, FOCSVM is used that assigns a membership degree to each training data. Hence, good decision boundaries and high accuracy is obtained under a training dataset with outliers.	FOCSVM

Table 2.11 (continued): A summary of research works in anomaly event detection (HiL HMA) using fuzzy techniques.

HiL processing	Problem statements / Sources of Uncertainty	Research papers	Why use fuzzy?	Technique
Anomaly event detection	A single camera environment specific elderly fall detection system provides limited information to infer the anomalous human behavior.	Wongkhuenkaew, Auephanwiriyakul, and Theera-Umpon (2013)	In order to learn the multi-prototype action classes in the multi-camera environment, fuzzy clustering algorithms such as FCM, Gustafson and Kessel clustering, or Gath and Geva clustering have been employed along with Hu moment invariant features and principle component analysis.	Fuzzy clustering
	In anomaly event detection algorithms, the problem is to determine the optimum parameters for fuzzy inference process e.g. membership function, threshold values, etc.	Z. Wang and Zhang (2008); Juang and Chang (2007); Hu, Xie, Tan, and Maybank (2004)	Integration of the fuzzy approaches with machine learning algorithms allows the learning of optimum fuzzy membership functions and fuzzy rules that can adapt to newly encountered problems.	Hybrid technique

depending upon the different levels of HMA task to be handled. In this thesis, fuzzy BK subproduct is employed for human action classification which is the first attempt in the community.

2.3 BK Subproduct

BK subproduct is a study of compositions of relations between sets first proposed by Bandler and Kohout (1980a). It can be defined in terms of crisp relations as well as fuzzy relations. To make the review as self-contained as possible, the discussion starts with an overview on BK subproduct.

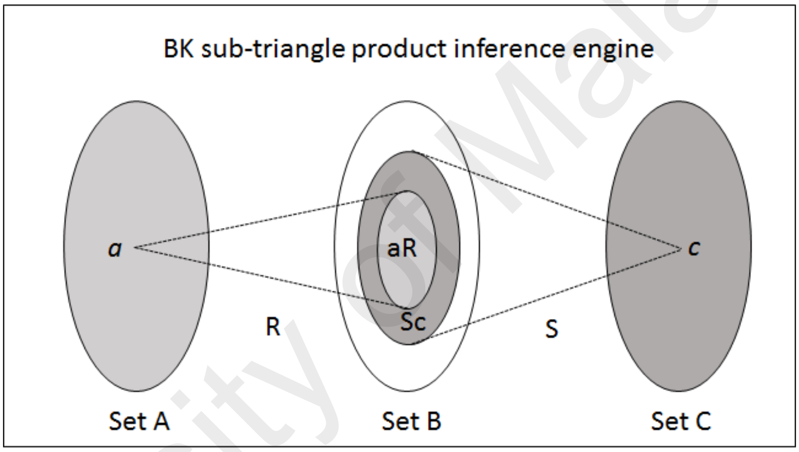


Figure 2.4: Overview of BK subproduct: element a in set A is in relation with element c in set C if its image under R (aR) is a subset of image Sc.

2.3.1 Overview on BK subproduct

Bandler and Kohout proposed that the relationship between two indirectly associated sets can be studied with BK relational product which defines the relationship between the elements within the two indirectly associated sets as the overlapping of their images in a common set. Figure 2.4 gives an overview of BK subproduct for crisp relations.

Let us assume that there exist three sets: set $A = \{a_i | i = 1, \cdots, I\}$, set $B = \{b_j | j = 1, \cdots, J\}$ and set $C = \{c_k | k = 1, \cdots, K\}$. If a relation R is defined between A and B such that $R \subseteq \{(a, b) | (a, b) \in A \times B\}$, and a relation S is defined between B and C such that

$S \subseteq \{(b, c)|(b, c) \in B \times C\}$, then BK subproduct can be defined as:

$$R \triangleleft S = \{(a, c)|(a, c) \in A \times C \text{ and } aR \subseteq Sc\} \tag{2.1}$$

BK subproduct finds all (a, c) couples such that the image of a under relation R in B (aR) is among the subset of c under the converse relation of S in B (S^c), as illustrated in Figure 2.4. For example, let A is a set of patients, B is a set of signs and symptoms and C is a set of diseases. For a patient a, relation R provides the signs and symptoms that are found on patient a (aR), while S gives the signs and symptoms that characterizes a disease c, it can be concluded that the patient a might be suffering from the disease c.

Though BK subproduct is very useful, it suffers from vagueness and uncertainty issues that exist in the real-world, and therefore Bandler and Kohout (1980a) extended the crisp BK subproduct to fuzzy BK subproduct to cope with these situations. Observing Eq. 2.1, it can be seen that $aR \subseteq Sc$ is the main element to retrieve the relationship between a and c. Therefore, the fuzzy subsethood measure was developed in Bandler and Kohout (1980b) based on the fuzzy implication operators '→', as shown in Table 2.12.

Let P and Q be the fuzzy subsets in the universe X, such that $x \in X$. Then the possibility that P is a subset of Q is given as:

$$\pi(P \subseteq Q) = \bigwedge_{x \in X} (\mu_P(x) \rightarrow \mu_Q(x)) \tag{2.2}$$

where $\bigwedge$ represents the arithmetic mean in mean criterion or the infimum operator in harsh criterion; $\mu_P(x)$, and $\mu_Q(x)$ represents the membership function of x in P and Q respectively; while → is the fuzzy implication operator. Utilizing Eq. 2.1 and 2.2, BK subproduct as the composition of relations between $a_i \in A$ and $c_k \in C$ is defined as

Table 2.12: Fuzzy implication operators, and their respective symbols and definitions.

Name	Symbol	Definition
Standard Sharp	$r \rightarrow_{S\#} s$	$\begin{cases} 1 & \text{iff } r \neq 1 \text{ or } s = 1 \\ 0 & \text{otherwise} \end{cases}$
Standard Strict	$r \rightarrow_{S} s$	$\begin{cases} 1 & \text{iff } r \leq 1 \\ 0 & \text{otherwise} \end{cases}$
Standard Star	$r \rightarrow_{S*} s$	$\begin{cases} 1 & \text{iff } r \leq s \\ s & \text{otherwise} \end{cases}$
Gaines 43	$r \rightarrow_{G43} s$	$\min(1, \frac{r}{s})$
Modified Gaines 43	$r \rightarrow_{KD} s$	$\min(1, \frac{r}{s}, \frac{1-r}{1-s})$
Kleene-Dienes	$r \rightarrow_{KD} s$	$\max(s, 1-r)$
Reichenbach	$r \rightarrow_{R} s$	$1 - r + rs = \min(1, 1 - r + s)$
Łukasiewicz	$r \rightarrow_{Ł} s$	$\min(1, 1 - r + s)$
Early Zadeh	$r \rightarrow_{EZ} s$	$(r \wedge s) \vee (1 - r)$

follows (Bandler & Kohout, 1980a):

$$R \triangleleft S(a,c) = \bigwedge_{b \in B} (R(a,b) \rightarrow S(b,c)) \tag{2.3}$$

where, $R(a,b)$ represents the membership function of the relation R between a and b; and $S(b,c)$ represents the membership function of the relation S between b and c.

Studies by Kohout and Bandler (1992); C. H. Lim and Chan (2012) also found that among all the fuzzy implication operators, Reichenbach fuzzy implication operator gives the expected values in the subsethood measurement. However, in De Baets and Kerre (1993), it was found that even if $a = a'$ has no image under relation R in set B, a' is still in relation $R \triangleleft S$ with all $c \in C$, because $\emptyset \subseteq Sc$. This limitation was studied and improved by reinforcing the non-emptiness condition:

$$R \triangleleft_K S = \{(a,c)|(a,c) \in A \times C \ \text{ and } \emptyset \subseteq aR \subseteq Sc\} \tag{2.4}$$

$$R \triangleleft_K S(a,c) = \min\Big(\bigwedge_{b \in B} (R_{ab} \to S_{bc}), \bigvee_{b \in B} \tau(R_{ab}, S_{bc})\Big) \tag{2.5}$$

where $\vee$ is the supremum operator and τ is the t-norm. To apply Eq. 2.5 in real-world applications, operators such as $\wedge$, $\vee$ and the t-norm must be defined, where Yew and Kohout (1996); Meng (1997) developed it into a list of inference structures. The study in C. K. Lim and Chan (2011) found that not all of these inference structures are reliable and proposed that the inference structures $K7$ and $K9$ deliver good performance. The definitions of $K7$ and $K9$ are as follows:

$$K_7 : \ R \triangleleft_{K7} S(a,c) = \min\Big(\frac{1}{J}\sum_{b \in B}(R_{ab} \to S_{bc}), \mathrm{OrBot}(\mathrm{AndBot}(R_{ab}, S_{bc}))\Big) \tag{2.6}$$

$$K_9 : \ R \triangleleft_{K9} S(a,c) = \min\Big(\mathrm{AndTop}(R_{ab} \to S_{bc}), \mathrm{OrBot}(\mathrm{AndBot}(R_{ab}, S_{bc}))\Big) \tag{2.7}$$

where AndTop, AndBot and OrBot are the logical connectives defined as follows:

$$\mathrm{AndTop}(p,q) = \min(p,q) \tag{2.8}$$

$$\mathrm{AndBot}(p,q) = \max(0, p+q-1) \tag{2.9}$$

$$\mathrm{OrBot}(p,q) = \min(1, p+q) \tag{2.10}$$

2.3.2 Applications

BK subproduct is a flexible approach that can be applied in real-life applications. For example, in developing the inference engine for several applications such as:

(a) Medical expert system: BK relational products has been successfully applied in Meng (1997), where the interval-valued inference structures were developed for medical diagnosis. The comparative studies were conducted in a medical expert system where the prime focus was on the identification of body systems.

(b) Information retrieval: The relational product architectures for information processing has been presented in Kohout and Bandler (1985). In this work, several examples of relational expressions demonstrated the strength of unification of relational representations in the field of information processing.

(c) Autonomous underwater vehicles' path navigation: An obstacle avoidance technique for autonomous underwater vehicles based on BK-products of fuzzy relation has been presented in Bui and Kim (2006) where the autonomous underwater vehicles are equipped with a looking-ahead obstacle-avoidance sonar. BK-products helps in revealing the characteristics and inter-relationships of the sonar sections. The experimental results demonstrated the capability of the proposed search technique, that employs BK-product of a fuzzy relation, in navigating safely through the obstacle with the optimal path.

(d) Land evaluation: In Groenemans et al. (1997), a fuzzy relational calculus based approach is introduced for land evaluation. The method is based on fuzzy relations between land qualities and land units, and hence describe the land suitability for a particular crop.

(e) Scene classification: A BK subproduct approach for scene classification is presented in Vats et al. (2012); Vats, Lim, and Chan (2015), where the experimental results on indoor and outdoor scene classes demonstrated the benefits of employing BK subproduct for scene classification task. By using fuzzy BK subproduct inference

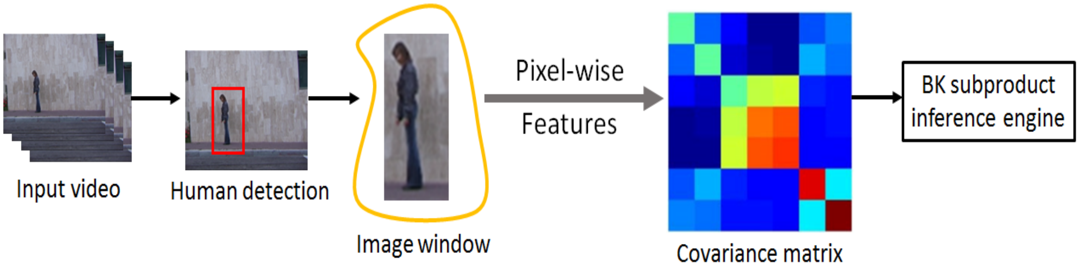


Figure 2.5: Application of fuzzy BK subproduct in human action recognition, illustrated with the help of an example of human motion image.

mechanism, the problem of mutual exclusiveness is solved. Therefore, a scene image can belong to multiple scene classes with a certain degree of belongingness.

In terms of human action recognition, consider an example of a human motion image as illustrated in Figure 2.5, where an actor is performing an action. Given an input video of action sequences, the human object is first detected for each image frame. This is followed by feature extraction. Features are the elements to be modeled and represented in a meaningful manner to signify the action. A popular feature extraction approach is to represent the image window by a covariance matrix of features (Porikli, Tuzel, & Meer, 2006), where the concept of covariance implies how much two variables vary together.

Let f_i denote the features extracted from image frames $i = 1, \cdots, I$ of the video describing the human action a_k, for $k = 1, \cdots, K$ action classes. The features extracted can be associated directly with the pixel coordinates. Therefore, the pixel-wise features $f_i = [x \;\; y \;\; I \;\; I_x \;\; I_y \;\; I_{xx} \;\; I_{yy}]$ can be extracted as represented in Fig. 2.5. By constructing the covariance of different features of a human image window (e.g. color, gradient, motion, edge etc.), the information from the histograms and the appearance models can be extracted. And by using bag of covariance matrices, the detection of actions, poses and shape changes can be taken into account efficiently (Porikli et al., 2006).

To detect human action in a given image, a BK subproduct classifier is first trained. The indirect relationship between the features representing the human image and the action being performed can be deduced using fuzzy BK subproduct. This is conditional on the presence of an intermediate set that is in relation with both f_i and a_k, such as the human body part-based model m_j for $j = 1, \cdots, J$ (where J denotes the number of models), obtained as a result of covariance tracking. Fuzzy BK subproduct classifier is invoked at each candidate image window to determine the target human action. For testing image sequences, this entails finding the features that signify the desired human action. The detection is triggered at frame i when the detector obtains the segment having the highest membership value.

2.3.3 Discussion

There exists several popular methods to perform this task e.g. using the well-known classifiers such as Support Vector Machines (SVM), K-nearest Neighbour (KNN) etc., but the employability of these algorithms depend on the desired application and its requirements. Although BK subproduct is not a popular classifier, but it can be efficiently used for classification tasks with the ability to provide a solution closer to how human interpret a situation in real life. For example, the relationship between a set of features and the action classes can be established if there exists an intermediate element that is in relationship with both, such as a human body model generated as a result of human motion tracking.

Furthermore, the introduction of fuzzy subsethood measure in BK subproduct simplifies the classification process in the sense that the crisp BK subproduct allows an action to belong to a single class only i.e. mutually exclusive classification approach. Whereas, fuzzy BK subproduct provides flexibility where an action can belong to a particular class with a certain degree of belongingness defined using fuzzy membership functions. Therefore, it offers non-mutually exclusive action classification. This is very crucial for early human action detection, because initially there is no information available about the action being performed. As the video progresses, the membership function values generated using fuzzy BK subproduct vary following a certain trend (e.g. monotonically increasing or decreasing), and thus enabling frame-by-frame action classification.

2.4 Early Human Action Detection

Early human action detection refers to detecting an action after it has begun but before it finishes. In real-world environment, it is essential to recognize human action before it is too late such as criminal acts, patients' fall etc. The sooner one can detect the action, the faster one can generate a response. There are several human action recognition methods existing in the CV literature (Cristani et al., 2013), and fuzzy literature (C. H. Lim et al., 2015). Almost all are focused on detecting once an action is completed, whereas for early detection it is necessary to detect partial (i.e. incomplete) actions.

Figure 2.6 highlights the scenario of early human action detection using an example of three common human actions i.e. bend, jump and skip. The aim is to detect the human action as soon as possible. The action video is observed frame-by-frame for early detection, instead of the conventional action classification approach where a video is fully observed to infer an action.

In the following subsections, a review is presented on the learning mechanism for existing early event detector (Hoai & De la Torre, 2012, 2014), followed by a study on the state-of-the-art methods for early human action detection along with their pros and cons.

2.4.1 Review on learning mechanism for early event detectors

For early event detection, partial events are used as positive training examples (Hoai & De la Torre, 2012, 2014), instead of a complete event. For a training set X^i (at time frame i) of length l^i and time $t = 1, 2, \cdots, l^i$, the output of the detector at time t is a partial event represented as:

$$g(X^i_{[1,t]}) = y^i_t = \underset{y \in Y(t)}{\arg\max}\, f(X^i_y) \tag{2.11}$$

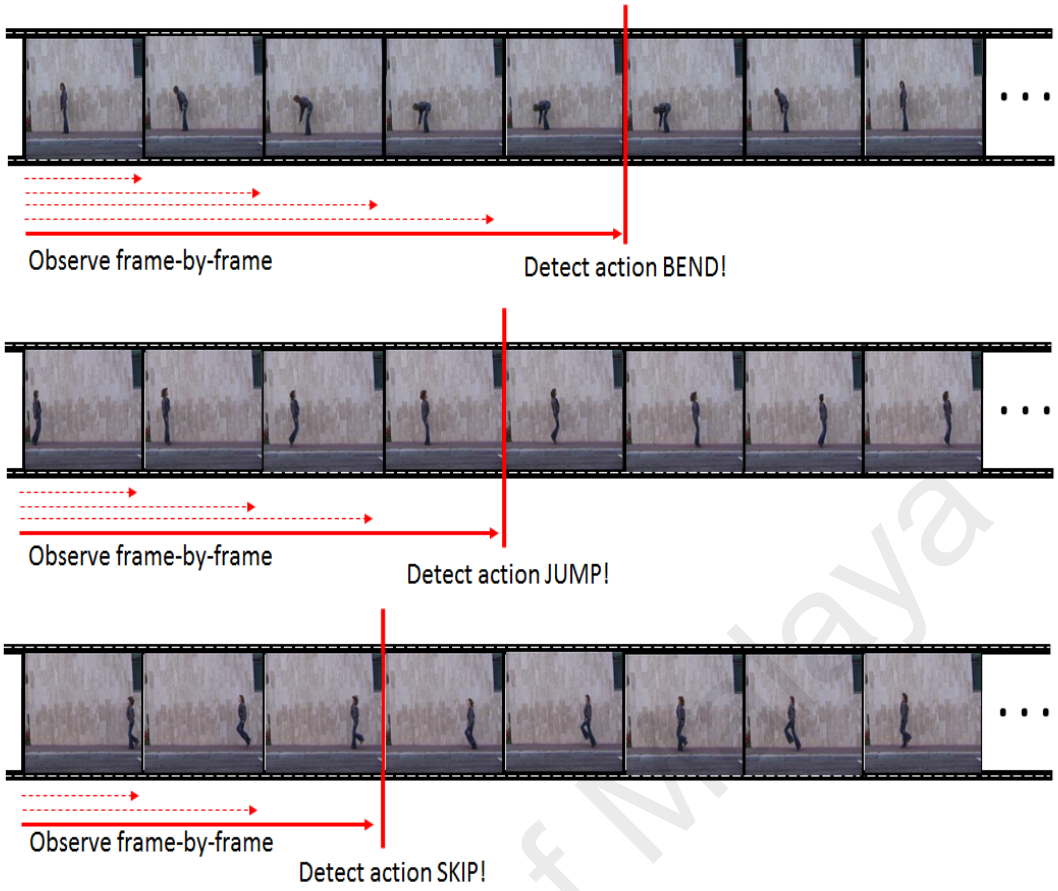


Figure 2.6: An example of early detection of three human actions: bend, jump, and skip. The action video is observed frame-by-frame, and the aim is to detect the action before it is completed.

where, $y_t^i = y^i \cap [1, t]$ is the part of event y^i that has already happened and is possibly empty; $g(X^i_{[1,t]})$ is the output of detector on the subsequence of time series X^i, not the entire set; and $f(X^i_y)$ is the detection score function. It is required that the detector score function is a monotonic and non-decreasing function. This means that the score of the partial event y_t^i should be greater than the score of any segment y ending before the partial event, which has been seen in the past, i.e.

$$f(X^i_{y_t^i}) \geqslant f(X^i_y) + \Delta(y_t^i, y) \forall y \in Y(t) \tag{2.12}$$

where $\Delta(y_t^i, y)$ is the loss of detector for outputting y when the desired output is y_t^i. This is illustrated in Figure 2.7. However, the score of the partial event is not required to be

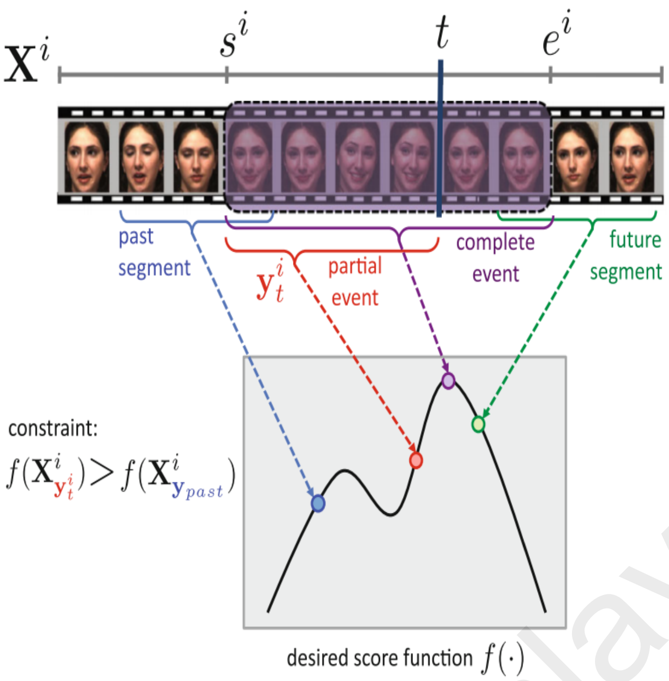


Figure 2.7: The desired score function for early event detection as presented in Hoai and De la Torre (2012, 2014).

greater than the score of a future segment.

The constraint in Eq. 2.12 is enforced for all $t = 1, 2, \cdots, l^i$. The learning formulation for early event detector is obtained as in Hoai and De la Torre (2012):

$$\min_{w,b,\xi^i \geqslant 0} \frac{1}{2}\|w\|^2 + \frac{C}{n}\sum_{i=1}^{n}\xi^i \tag{2.13}$$

so that

$$f(X^i_{y^i_t}) \geqslant f(X^i_y) + \Delta(y^i_t, y) - \frac{\xi^i}{\mu\left(\frac{|y^i_t|}{|y^i|}\right)} \tag{2.14}$$

$$\forall i, \forall t = 1, ..., l^i, \forall y \in Y(t)$$

where w is a weight vector, b is a scalar bias term, C is the cost parameter, and n denotes the number of instances of the training data. This is an extension of Structured Output SVM (SOSVM), with the alteration on setting $t = 1, 2, \cdots, l^i$ instead of $t = l^i$, because partial events are trained instead of a complete event. An additional slack variable ξ^i is added as a rescaling factor for correctly detecting the occurrence of an event at time t.

2.4.2 State-of-the-art methods and limitations

Most of the existing work dealing with early human action detection aims at detecting unfinished activity. Ryoo (2011) proposed the integral Bag of Words (BoW) and dynamic BoW approaches as an extension to the BoW paradigm for early recognition of ongoing human activities, and delivered promising results. However, the model learned for activity recognition may not be representative if the action sequences of the same action class have large appearance variations. Also, it was found to be sensitive to outliers. The solution to these two issues was provided in Cao et al. (2013) where the action models were built by utilizing sparse coding to learn the feature bases, and using the reconstruction error in the likelihood computation.

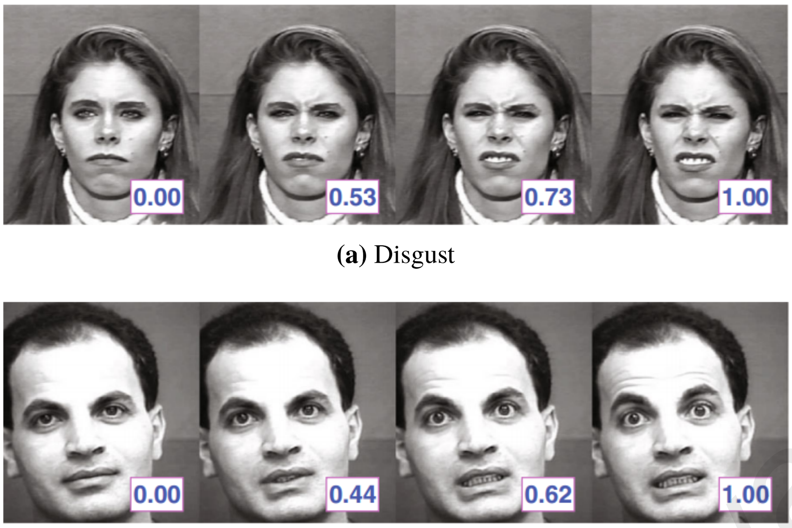
Other limitations of Ryoo (2011) include the assumption made that the activities within the same action class always have identical speed and duration which is not true in most cases. Also, the poor discriminative model generated to describe human action, ignoring the BoW model in spatial-temporal relationships among the interest points. This issue was taken into account in G. Yu et al. (2012) where a spatial-temporal implicit shape model was proposed to model the relationship between the local features, and at the same time predict multiple activities. The method proposed in Kong, Kit, and Fu (2014) incorporated an important prior knowledge that as new observations are available when the action video progresses, the amount of crucial information about the action also increases. However, the methods Ryoo (2011); Cao et al. (2013); G. Yu et al. (2012) did not utilize this prior knowledge. In addition, Kong et al. (2014) modeled the label consistency of segments, which provides discriminative local information, and implicitly captures the context-level information that is useful for predicting action. Moreover, Kong et al. (2014) captured the action dynamics in both global and local temporal scales, unlike Ryoo (2011); Cao et al. (2013) where the dynamics in single scale were captured. Despite of the advantages these methods offer, they lack in the ability to handle the uncertainties that exist in a real-world.

The early recognition of human action for the dynamic first-person videos was studied in Ryoo et al. (2014), where the pre-activity observations were considered that includes the frames ‘before’ the starting time of the activity. However, this work is different from the goal of this thesis. An activity forecasting method was proposed in Kitani, Ziebart, Bagnell, and Hebert (2012), however activity forecasting and early detection differs in the sense that forecasting makes prediction about the future events, whereas early detection interprets the present action, as soon as possible. Autoregressive-moving-average (ARMA) - Hidden Markov Model (HMM) based approach was employed in K. Li and Fu (2012) which integrates both the predictive power of sequential model HMM and the time series model ARMA. Unfortunately, it requires building separate HMMs for each activity and therefore is computationally expensive.

Max Margin Early Event Detector (MMED) was proposed for early events detection in Hoai and De la Torre (2012, 2014) which is based on the SOSVM (Tsochantaridis et al., 2005) and requires extensive labeling on each of the training samples. In terms of timeliness and accuracy, MMED performs efficiently. For example, consider the detection of facial expressions (e.g. disgust and fear) as illustrated in Figure 2.8, MMED fires seeing lesser number of frames as compared to SOSVM. However, early human action detection is a complex task given the vast amount of uncertainty involved therein. An efficient algorithm should be able to handle even the minutest level of uncertainty for a reliable decision making.

2.5 Summary

The conventional CV solutions (Ryoo, 2011; Cao et al., 2013; G. Yu et al., 2012; Kong et al., 2014; Ryoo et al., 2014; K. Li & Fu, 2012; Hoai & De la Torre, 2012, 2014) often fall



(**a**) Disgust

(**b**) Fear

Figure 2.8: From left to right: the onset frame, the frame at which MMED fires (Hoai & De la Torre, 2012, 2014), the frame at which SOSVM fires (Tsochantaridis et al., 2005), and the peak frame. The number in each image represents the corresponding NTtoD.

short of providing an effective solution as they are not robust enough to handle issues such as uncertainty, imprecision and vagueness that arise in a real-world. The fuzzy approaches are well-known to offer an effective solution with the inherent capability of assigning a degree of belongingness to a human action using the fuzzy membership function. The problem of early human action recognition can be efficiently addressed by integrating CV solutions with fuzzy set oriented techniques in a way that the strength of fuzzy set theory can alleviate the limitation of CV solutions.

An extensive review was presented in this section on the state-of-the-art methods and solutions that are relevant to the problem statement this thesis is addressing. First of all, fuzzy human motion analysis was reviewed in an elaborate manner in order to understand the necessity of employing fuzzy techniques for HMA. The first fuzzy human motion analysis review paper (C. H. Lim et al., 2015) in the research community was hence delivered. Secondly, BK subproduct approach was reviewed, with highlight on its applications. Lastly, the state-of-art methods for early human action detection were reviewed, along with the discussion on the challenges and the current state of the problems.

CHAPTER 3: FUZZY BK SUBPRODUCT - A CLASSIFIER

The most fundamental problem is the selection of classifier to employ for the classification task. As a solution, in this work fuzzy BK subproduct is employed as a classifier because of the advantages it offers in terms of providing a solution closest to natural human perception. However, this is the first attempt of using fuzzy BK subproduct as a classifier. Therefore, its performance is validated in terms of handling both 3D and 2D data. For 3D video data, its performance is tested for HMA (section 3.1). For the 2D image data, an example of outdoor and indoor scene images are taken into account, and the employability of fuzzy BK subproduct is tested for scene classification (section 3.2).

3.1 Human Motion Analysis

The aim of a HMA system is to model human movement changes with respect to time. It involves feature extraction in the LoL processing steps and human motion tracking in the intermediate levels. In general, the relationship between the features representing the human body and the human action being performed is required to be modeled. This section discusses the capabilities of fuzzy BK subproduct in representing the indirect relationship between the features and the action to perform HMA.

A general framework of HMA processing using fuzzy BK subproduct approach is represented in Figure 3.1. As illustrated, the interest is to study the indirect relationship between the human, and the actions being performed by modeling the inference mechanism in a way capable of reliable decision making. The aim is to find the relationship between set A (consisting features f) and set C (consisting actions a), if there exists an intermediate set B (e.g. human body model m) which is in relation with both A and C.

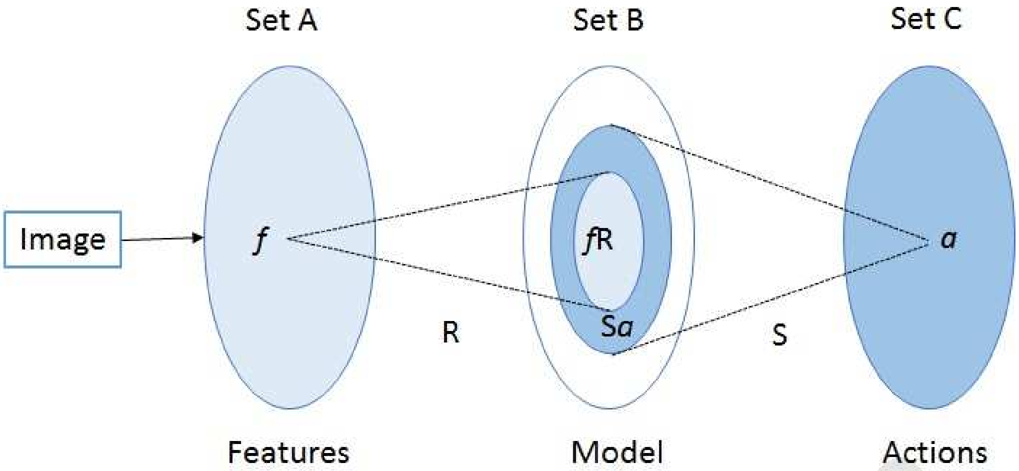


Figure 3.1: Fuzzy BK subproduct approach for HMA.

3.1.1 Proposed methodology

The overall pipeline for the proposed methodology is highlighted in Figure 3.2. The three important steps include feature extraction, covariance tracking, and human action classification, and will be discussed step-by-step in the following subsections.

3.1.1.1 Feature extraction

For each image frame, firstly a feature image is constructed following the method in Porikli et al. (2006). For an image I, let F be the $W \times H \times d$ dimensional feature image (RGB), such that

$$F(x, y) = \Phi(I, x, y) \tag{3.1}$$

where the function Φ can be any mapping, e.g. color, image gradients, edge magnitude or orientation, etc. Let $\{f_i\}_{i=1..I'}$ be the d-dimensional feature vectors inside a rectangular window R' (where $R' \subset F$). A feature vector f_i is constructed using two types of mapping, i.e. spatial attributes based mapping that is obtained from the pixel coordinates values, and the appearance attributes based mapping (e.g. color, gradient, infrared, etc.). The feature vector forms the set A in the proposed method, and is denoted as set $A = \{f_i | i = 1, \cdots, I'\}$

.

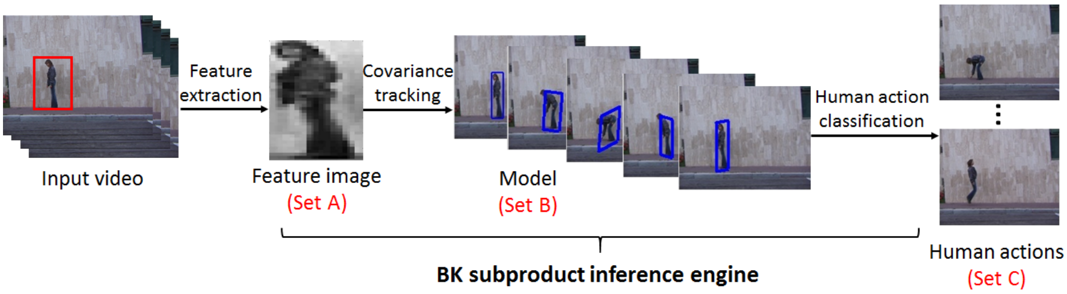


Figure 3.2: Overall pipeline for fuzzy BK subproduct approach towards HMA.

3.1.1.2 Covariance tracking

For a given object region R', a $d \times d$ covariance matrix of features $C_{R'}$ is then computed as the model of human object:

$$C_{R'} = \frac{1}{MN} \sum_{i=1}^{MN} (f_i - \mu_{R'})(f_i - \mu_{R'})^T \tag{3.2}$$

where $\mu_{R'}$ is the vector of the means of the corresponding features for the points in region R'. The covariance matrix is basically a symmetric matrix where the diagonal represents the variance of each feature in the image, and the non-diagonal values represent their representative correlations. The reason for using covariance matrices as region descriptors is that covariance matrix proposes an efficient way of combining multiple features without the need to normalize features or blend weights, along with the advantage of scale invariance property (Porikli et al., 2006). In general, a single covariance matrix extracted from a region is sufficient to perform matching of the region in multiple views and poses. Also, the noise in the images is filtered out with average filter during the covariance computation. In the current image frame, the region having the minimum covariance distance from the model is found, and assigned as the estimated location (covariance tracking). In order to adapt to variations, a set of previous covariance matrices is kept, and an intrinsic mean using Lie algebra is extracted. The study in Porikli et al. (2006) presents a detailed view on the Lie algebra based covariance tracking method.

In order to yield more succinct description of human body, the human body is segmented into three parts: head, torso and leg. The covariance tracking is performed on these parts, resulting in a part based model m, that forms the set B in the proposed method. It is denoted as set $B = \{m_j | j = 1, \cdots, J\}$. The models are defined using the user knowledge (researcher), utilizing the knowledge on various human actions. In general, set B constitutes three main models (i.e. $J = 3$):

(i) Distance obtained by modeling the head movements from start to end frame.

(ii) Distance obtained by modeling the position changes of the human body from the origin (first frame).

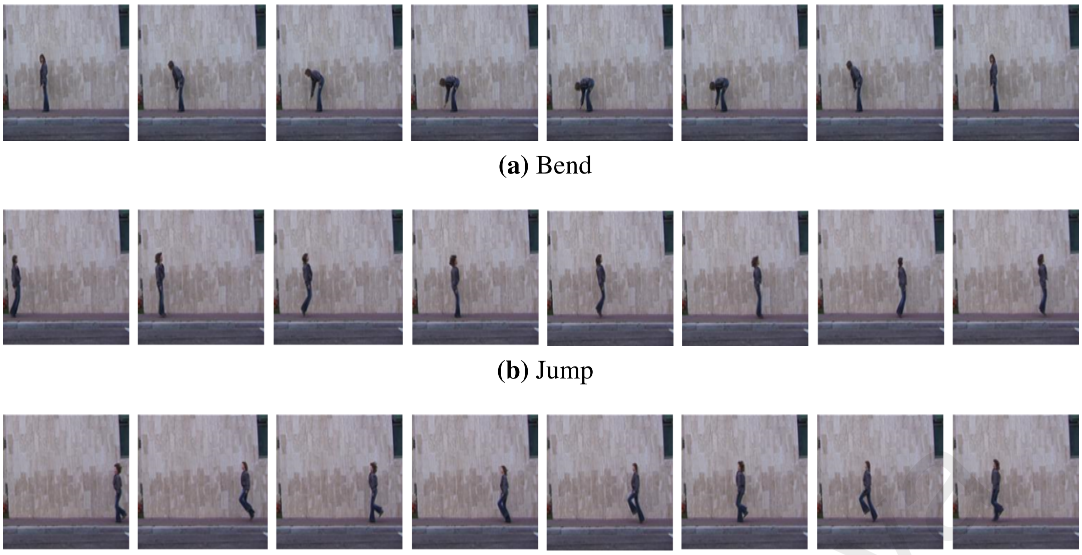
(iii) Distance between both legs.

The relationship R is therefore derived between the features and the part-based human body models by normalizing the covariance tracking results obtained using the min-max normalization method.

3.1.1.3 Human action classification

The main aim is to perform human action classification. Therefore, let set $C = \{a_k | k = 1, \cdots, K\}$ constitute the action being performed by a human (e.g. bend, jump and skip). A (image features) has no direct relation with C (as there is no information about which action is being performed and by whom). However, if there exists an intermediate set B (model), which is in relation with both A and C, the indirect relationship between A and C can be derived using fuzzy BK subproduct (Eq. 3.3).

$$BK: \ R \triangleleft_{BK} S(a,c) = \frac{1}{J} \sum_{b \in B} (R(a,b) \rightarrow S(b,c)) \tag{3.3}$$



(a) Bend

(b) Jump

(c) Skip

Figure 3.3: Example of image frames from the Weizmann human actions dataset (Gorelick et al., 2007).

Referring to Eq. 3.3, the relation between A and B is defined by relation R; and S defines the converse relation between B and C. BK subproduct gives all (f, a) couples such that the image of f under relation R in B is among the subset of a under Sa in B, as illustrated in Figure 3.1. Therefore, Eq. 3.3 can be re-written as:

$$BK:\ R \triangleleft_{BK} S(f, a) = \frac{1}{J} \sum_{m \in B} (R(f, m) \rightarrow S(m, a)) \tag{3.4}$$

where, $R(f, m)$ is the membership function of the relation R between f and m; $S(m, a)$ is the membership function of the relation S between m and a. The membership function values generated from Eq. 3.4 are modeled for HMA.

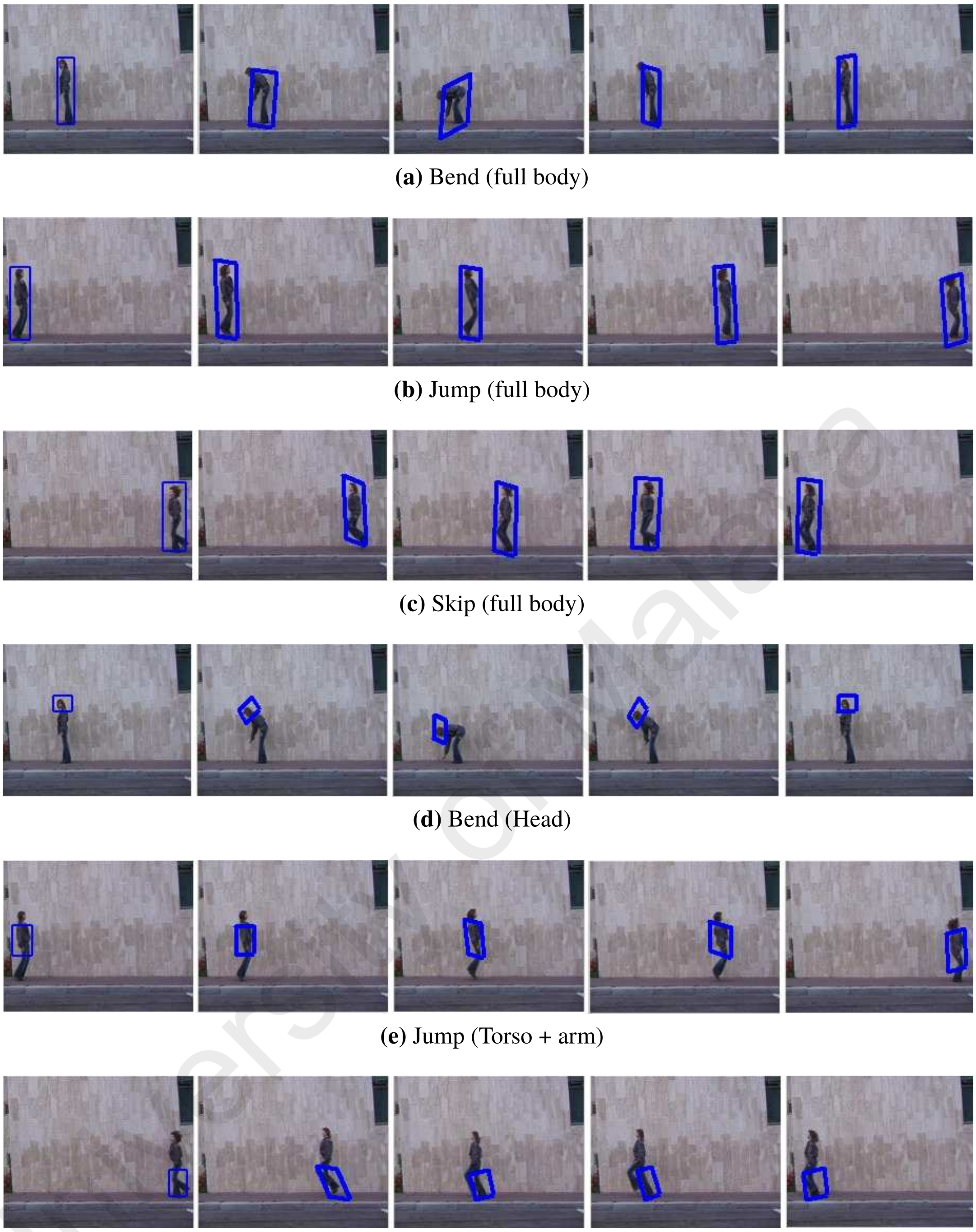
3.1.2 Validation

In order to test the efficiency of fuzzy BK subproduct to perform HMA, experiments are performed on the Weizmann human actions dataset (Gorelick et al., 2007). Weizmann human actions dataset is a database of 90 low-resolution (180 x 144, deinterlaced 50 fps) action video sequences. It presents nine different people where each actor performs ten natural actions that include run, walk, skip, jumping-jack (jack), jump-forward-on-two-legs (jump), jump-in-place-on-two-legs (pjump), gallop sideways (side), wave-two-hands (wave2), wave-one-hand (wave1), and bend. From the dataset, three actors, and three actions (a_1 = bend, a_2 = jump, and a_3 = skip) are selected, as presented in Figure 3.3. The main reason behind selecting these three actions is that *bend* is distinctive as compared to the other two; and *jump* and *skip* are quite similar in their movement patterns.

The following pseudo-code represents the implementation of fuzzy BK subproduct as a classifier for human action recognition in a step-by-step manner:

```
Step 1: Input an action video.
Step 2: Perform human detection.
Step 3: Segment the human body into three parts: head, torso+arm and leg.
Step 4: Perform feature extraction by constructing covariance matrix of features.
Step 5: Feature image is constructed. Save in Set A.
Step 6: Perform part-based covariance tracking.
Step 7: Human body part-based models are obtained. Save in Set B.
Step 8: Normalize the results obtained using min-max normalization. Save as membership function R.
Step 9: Obtain the converse relation S between actions and models by normalizing the tracking results using min-max normalization.
Step 10: Call BK subproduct inference engine utilizing R and S.
Step 11: Output the membership degree.
Step 12: Done.
```

The preprocessing of images, feature extraction and covariance tracking are performed referring the method in Porikli et al. (2006). Their method is modified to generate part-based human body model, and the covariance tracking is performed on each body



(a) Bend (full body)

(b) Jump (full body)

(c) Skip (full body)

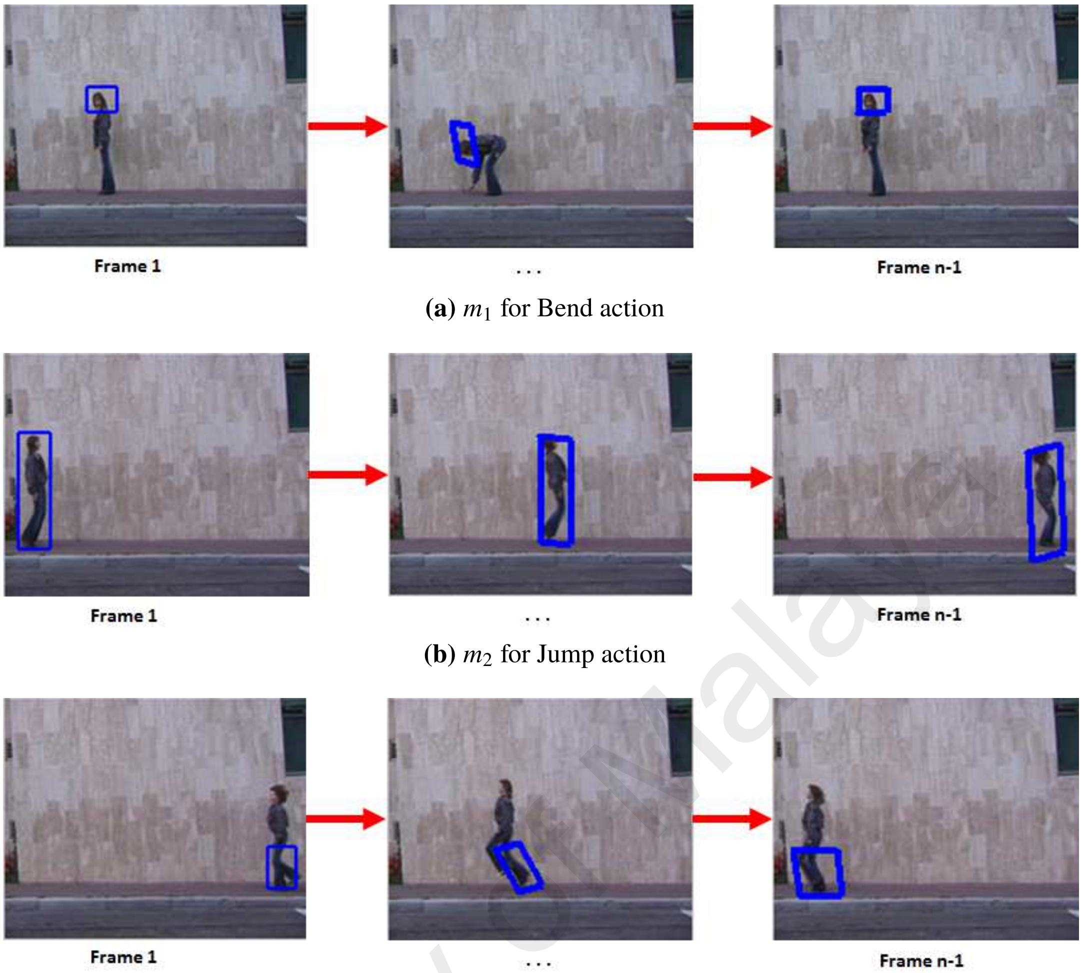
(d) Bend (Head)

(e) Jump (Torso + arm)

(f) Skip (Leg)

Figure 3.4: Sample human motion tracking results for three different action sequences. (a) - (c) gives the tracks for full body, while (d) - (f) highlights the tracking results for the body parts: head, torso+arm, and leg respectively, represented using blue colored bounding box.

part. Sample human motion tracking results are presented in Figure 3.4, where the tracking results for body parts: head, torso+arm, and leg are highlighted (using blue colored bounding box). The tracking results demonstrate the capability of the tracking algorithm



(a) m_1 for Bend action

(b) m_2 for Jump action

(c) m_3 for Skip action

Figure 3.5: Set B defining the three models used, where m_1: models the changes in the head positions with time from start to end frame; m_2: models the position changes of the human body from the origin (first frame); m_3: models the distance between both legs.

to adapt to the undergoing object deformations and appearance changes. No assumption was made on the measurement noise and the motion of the objects tracked. Hence, the tracking results demonstrate the efficiency of the covariance tracking algorithm in modeling the movements of different human body parts over time, with remarkable detection accuracy, and its tolerance to the background noise.

Furthermore, for the set B, the euclidean distance is computed by modeling position changes of the human body. More specifically, the changes in the head positions with respect to time from start to end frame (m_1), the position changes of the human body from the origin i.e. the first frame (m_2), and the distance between both legs (m_3). Figure

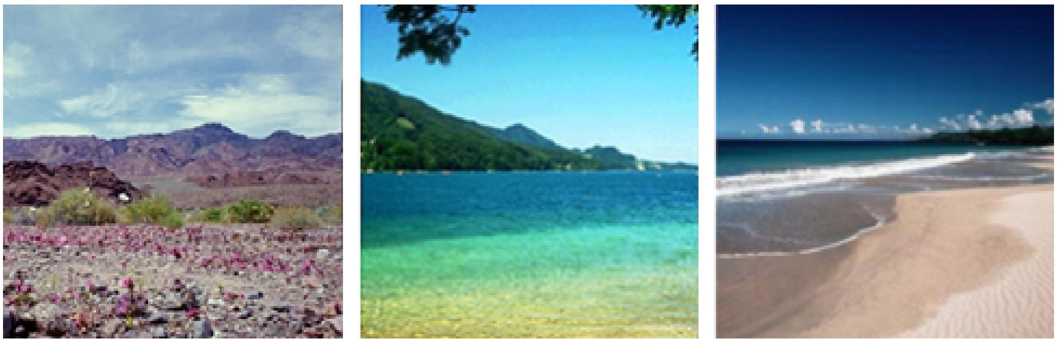
3.5 represents the results for set B. Hence, three models are generated that are crucial to establish the relationship between the features and the action.

However, in this section, the results till the tracking stage are highlighted. Because, the next chapter studies fuzzy BK subproduct approach for HMA in detail, where the conventional classification problem is modified into frame-by-frame level classification to perform early human action detection. The results are presented in section 4.3.

3.2 Scene Classification

Understanding and interpreting a natural scene is a challenging task in the CV community because of the variability, ambiguity, illumination and scale conditions that can exist in the scene images. A scene composed of several objects often is organized in an unpredictable layout. A set of perceptual dimensions - naturalness, openness, roughness, expansion and ruggedness was presented in Oliva and Torralba (2001) to represent the dominant spatial structure of a scene. SVM classifier was employed with Gaussian kernel to classify the scene classes. While Bosch, Zisserman, and Muñoz (2006) proposed probabilistic Latent Semantic Analysis (pLSA) based method incorporated with the KNN classifier. Inspired from Bosch et al. (2006), Fei-Fei and Perona (2005) proposed a Bayesian hierarchical model for learning natural scene categories. Furthermore, Kumar and Hebert (2003) employed the graphical models for the detection and localization of man-made features in a scene. The concept of occurring frequency of different concepts was used by Vogel and Schiele (2004, 2007) as the intermediate feature for scene classification.

Though all the aforementioned methods have achieved promising results, it is observed that the classification errors often occur when there is an overlap between the scene classes in the selected feature space. The reason is the assumption made that the scene classes are mutually exclusive, where most systems learn patterns from a training set and search similar images. Figure 3.6 explains this scenario. It is unclear in Figure 3.6b



(a) Open Country **(b)** ?? **(c)** Coast

Figure 3.6: Example of ambiguous scene images. Which class does (b) belong? It is not clear that it is an *open country* scene or a *coast* scene and different people may respond inconsistently.

that it is an *open country* scene or a *coast* scene where different people may respond inconsistently. Therefore, it is argued that the scene classes are non-mutually exclusive. This is also stated in the work conducted by the authors in C. H. Lim, Risnumawan, and Chan (2014).

Fuzzy BK subproduct approach is employed in this work to tackle this issue and perform scene classification. A series of CV techniques and online surveys are used to compute the relational products of an image and its scene classes. The proposed classification method is closely related to some of the approximate reasoning methods which have been developed in the recent years, more specifically Barrenechea, Bustince, Fernandez, Paternain, and Sanz (2013); Bustince, Burillo, and Soria (2003). In Barrenechea et al. (2013), a fuzzy reasoning method is presented in which the Choquet integral is used as an aggregation function for the rule-based classification systems. A wide benchmark of numerical datasets are used to test the classification performance. However, their classification results are binary, allowing an element in the dataset to belong to a single class only. Most likely the chances of classification errors occur when there is an overlap between the classes. The proposed method in this thesis deals with the multi-class, multi-label classification problem, wherein the driving force is the non-mutually exclusive scene classes.

Nonetheless, rule-based systems require the expert knowledge in designing the rules for the system. Fuzzy BK subproduct approach is based on the study of the relationship between different fuzzy sets, and hence can provide a better alternative that is closer to the natural solution. A study on the implication operators was presented in Bustince et al. (2003). In this this, the fuzzy implication operators are employed for scene classification.

The closest research is C. H. Lim et al. (2014), where a fuzzy qualitative approach is incorporated to address the problem. However, the proposed method in this thesis is found to be much closer to a natural solution in the sense that fuzzy BK subproduct inference mechanism is a flexible and efficient method (C. K. Lim & Chan, 2015) that can be employed in the real-world scenarios. This is because it imitates how human think in real life, i.e. modus-ponen way (if A implies B, A is asserted to be true, so therefore B must be true.).

3.2.1 Proposed methodology

Let $A = \{a_i | i = 1, \cdots, I\}$ denote a set of scene images, $B = \{b_j | j = 1, \cdots, J\}$ denote a set of features extracted from the image frames, and $C = \{c_k | k = 1, \cdots, K\}$ denote a set of scene classes. A has no direct relation with C. However, if there exists an intermediate set B, which is in relation with both A and C, the indirect relationship between A and C can be derived using fuzzy BK subproduct, along with the combination of $K7$ and $K9$, and this information can be utilized to classify different scene images.

First of all, for each image $a \in A$, several local patches are extracted and represented in terms of 128-dimensional numerical vectors $(V_1, V_2, \cdots, V_{128})$ using Scale Invariant Feature Transform (SIFT) descriptors. This is to find the features that govern the image. With this, each image a is represented by a set of vectors $a' \in A'$, as depicted in Figure 3.7. Instead of using the relation $R \subseteq A \times B$, R is replaced with R', where $R' \subseteq A' \times B$.

After the key information from the images is extracted, k-means clustering is per-

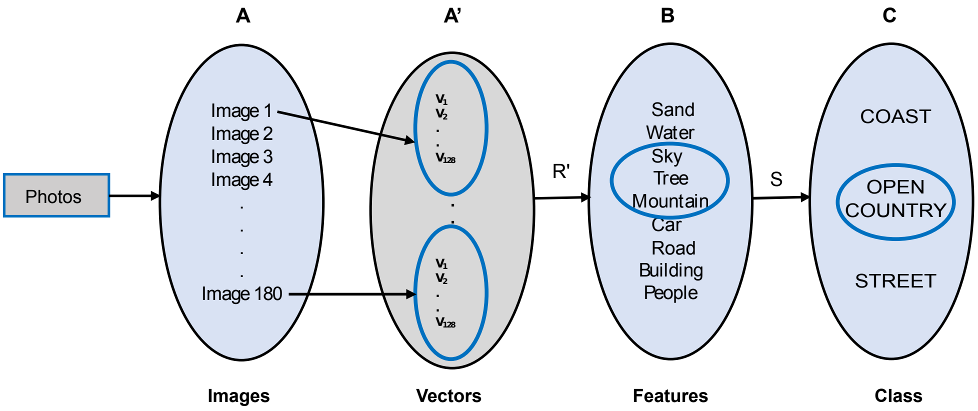


Figure 3.7: An example of fuzzy BK subproduct approach towards scene classification.

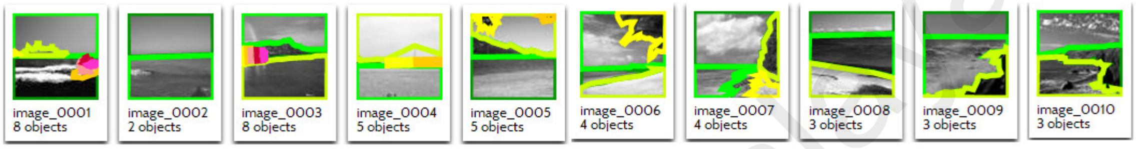
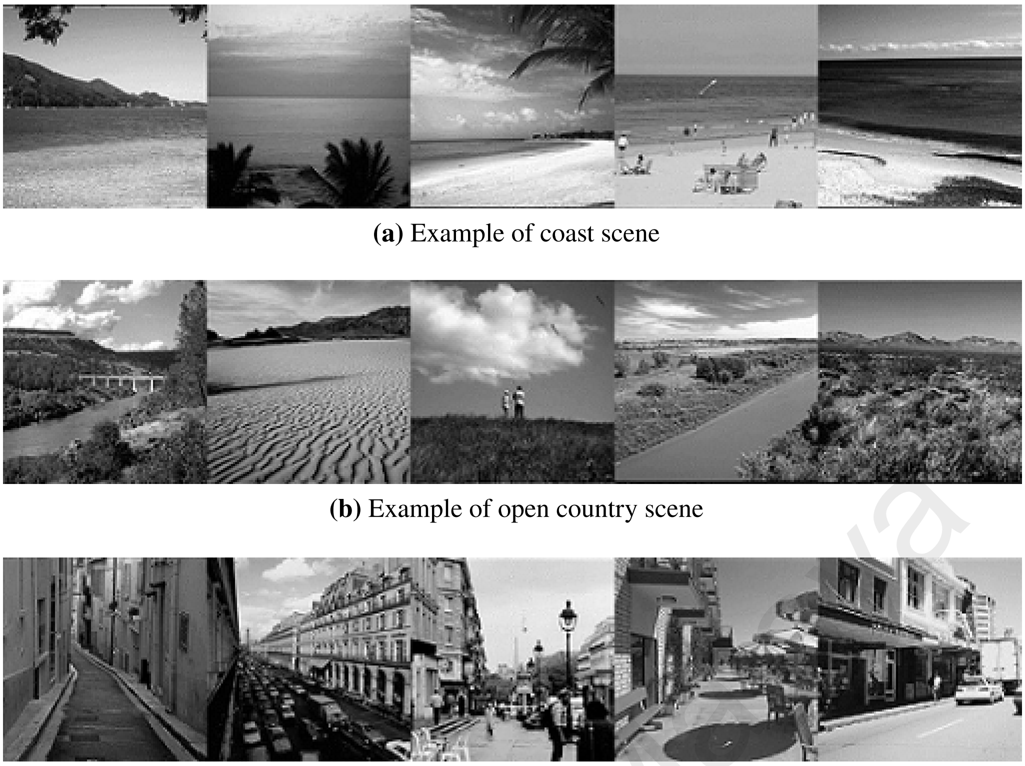


Figure 3.8: An example of the annotated images from *coast* scene employing *Labelme* (Russell et al., 2008).

Table 3.1: Membership Function for Relation R'

Images	Sand	Water	Sky	Tree	Mountain	Vehicle	Road	Building	People
Image 1	0.00	0.45	0.415	0.00	0.003	0.102	0.00	0.000	0.03
Image 2	0.00	0.51	0.49	0.00	0.00	0.00	0.00	0.00	0.00
Image 3	0.00	0.50	0.26	0.00	0.19	0.00	0.00	0.05	0.00
Image 4	0.00	0.42	0.55	0.00	0.03	0.00	0.00	0.00	0.00
Image 5	0.00	0.56	0.254	0.046	0.14	0.00	0.00	0.00	0.00
Image 6	0.24	0.16	0.41	0.19	0.00	0.00	0.00	0.00	0.00
Image 7	0.08	0.32	0.44	0.16	0.00	0.00	0.00	0.00	0.00
Image 8	0.39	0.42	0.19	0.00	0.00	0.00	0.00	0.00	0.00
Image 9	0.00	0.58	0.24	0.00	0.18	0.00	0.00	0.00	0.00
Image 10	0.00	0.38	0.42	0.00	0.20	0.00	0.00	0.00	0.00

formed to group together the similar features found on each image a. However, K-means clustering only provides the information required by the SVM classifier, and the desired result is the linguistic formulation of set B. Therefore, an open annotation tool *LabelMe* (Russell et al., 2008) is used to detect and label the image features and generate set B, as has been illustrated in Figure 3.8. To find the relation R', the membership function values are computed by calculating the distribution of each feature in the image, and normalizing the results based on the total area covered by the feature attribute in the image. A total of



(**a**) Example of coast scene

(**b**) Example of open country scene

(**c**) Example of street scene

Figure 3.9: Example of three scene classes from the Outdoor Scene Recognition (OSR) dataset (Oliva & Torralba, 2001).

nine distinct features namely, *sand*, *water*, *sky*, *tree*, *mountain*, *vehicle*, *road*, *building* and *people* are identified. An example of the membership function R' for ten example images from the *coast* scene is shown in Table 3.1. From here, it can be noticed that the major features that represent *coast* scene are *water* and *sky*; and *sand*, *tree* and *mountain* are the minor features.

Finally, in order to find the relation S, which denotes the membership function between the image features and the scene classes, an online survey is conducted where each subject is given a series of image features for scene classification. This is in contrast to the conventional solutions (Bosch et al., 2006; Fei-Fei & Perona, 2005; Kumar & Hebert, 2003; Oliva & Torralba, 2001; Vogel & Schiele, 2004, 2007) that learned a binary classifier with the assumption that the scene classes are mutually exclusive.

3.2.2 Validation

In order to test the effectiveness of the proposed framework, the public dataset: Outdoor Scene Recognition (OSR) (Oliva & Torralba, 2001) is employed. A total of three scene classes namely, *coast*, *open country* and *street* are used throughout the experiments. Figure 3.9 shows the example of the scene classes in gray scaled. Each scene class has 60 images, and therefore there are 180 images in total. In each scene class, 40 images are used for training and the rest are for testing. The SVM implementation is based on the LIBSVM MATLAB toolbox.

The following pseudo-code represents the implementation of fuzzy BK subproduct classifier for scene classification in a step-by-step manner:

```
Step 1: Input scene images.
Step 2: Perform feature description using SIFT.
Step 3: 128 dimensional feature vectors generated. Save in Set A.
Step 4: Perform K-means clustering.
Step 5: Using LabelMe annotate the scene images to get linguistic description. Save in Set B.
Step 6: Calculate the area of each image feature with respect to whole image, normalize and save as membership function R.
Step 7: Normalize the online survey results to obtain the membership function S.
Step 8: Call BK subproduct inference engine utilizing R and S.
Step 9: Output the membership degree.
Step 10: Done.
```

As mentioned previously, different people tend to respond inconsistently for a given scene image. Therefore, it is possible for an image to belong to multiple classes. Herein, a survey is conducted on 200 people via social networking website to get information

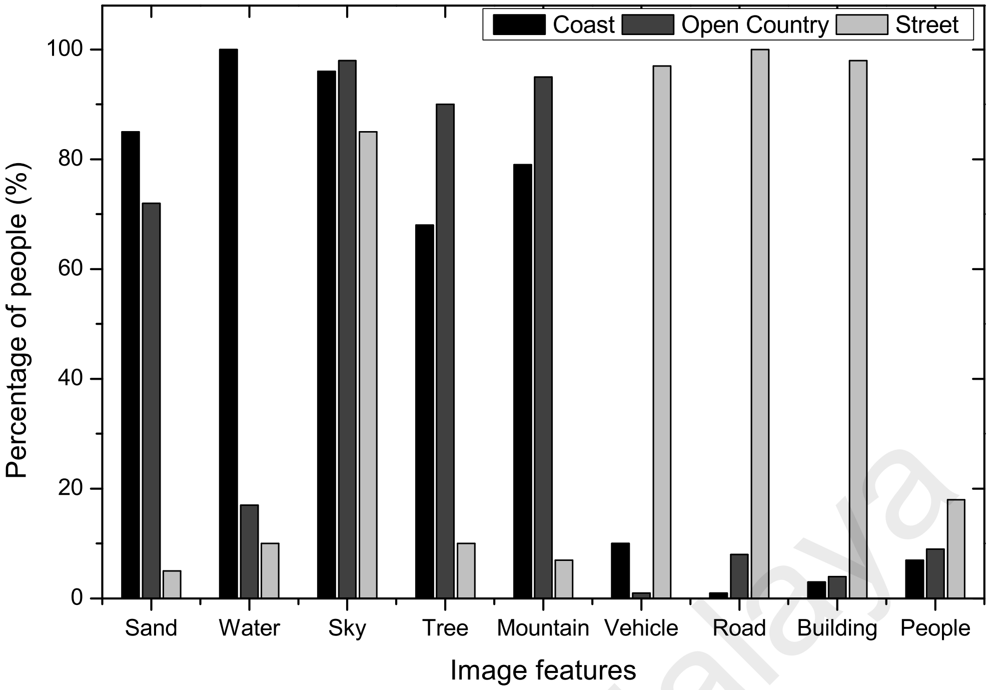


Figure 3.10: Bar chart representing the results from the online survey on 200 people.

Table 3.2: Membership Function for Relation *S*

Features	Coast	Open Country	Street
Sand	0.55	0.35	0.00
Water	1.00	0.10	0.02
Sky	1.00	1.00	0.87
Tree	0.25	0.72	0.17
Mountain	0.45	0.60	0.02
Vehicle	0.12	0.00	0.80
Road	0.00	0.02	0.95
Building	0.05	0.05	1.00
People	0.15	0.10	0.30

on how different people can classify features into various scene classes. Each subject is given a choice of nine image features (*sand*, *water*, *sky*, *tree*, *mountain*, *vehicle*, *road*, *building* and *people*). The outcome obtained is represented in Figure 3.10, where X-axis denotes the image features and Y-axis denotes the total percentage of people. The results are further normalized using min-max normalization method in order to obtain the membership function values for converse relation *S*. On analyzing the bar charts in

Table 3.3: Test results for all the scenes against coast scene class

Coast (Threshold=0.7)	Accept Original BK	Accept K7	Accept K9	Reject Original BK	Reject K7	Reject K9
Coast	20	17	20	0	3	0
Open Country	20	4	14	0	16	6
Street	0	0	5	20	20	15

Table 3.4: Test results for all the scenes against open country scene class

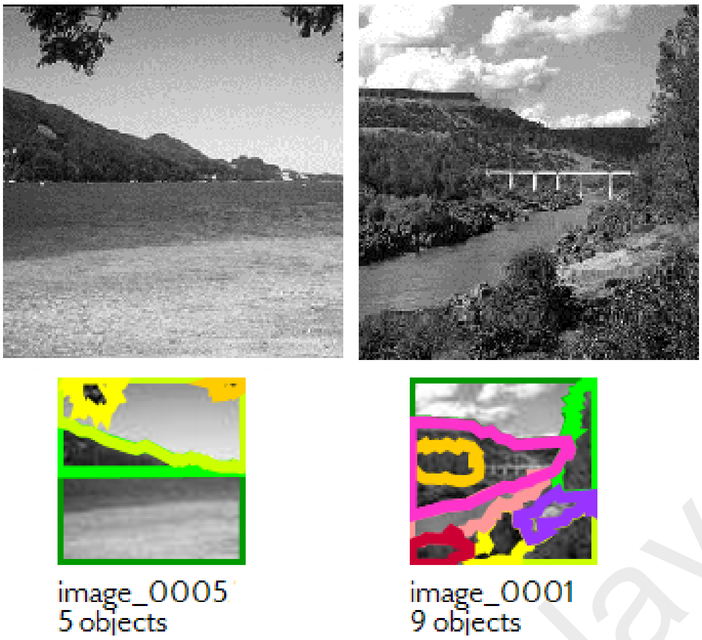
Open Country (Threshold=0.6)	Accept Original BK	Accept K7	Accept K9	Reject Original BK	Reject K7	Reject K9
Coast	20	1	19	0	19	1
Open Country	20	4	19	0	16	1
Street	0	0	4	20	20	16

Table 3.5: Test results for all the scenes against street scene class

Street (Threshold=0.5)	Accept Original BK	Accept K7	Accept K9	Reject Original BK	Reject K7	Reject K9
Coast	0	0	4	20	20	16
Open Country	0	0	1	20	20	19
Street	20	20	20	0	0	0

Figure 3.10, it is observed that the features sky, $tree$, $mountain$ and $sand$ are present in both $open\ country$ and $coast$ scenes in different proportions, making them related to one another with a certain degree. On the other hand, $street$ scene is governed by $vehicle$, $road$ and $building$. The membership function values for relation S from the survey are presented in Table 3.2.

Furthermore, a dynamic threshold value is defined for each of the scene classes, and the number of images accepted or rejected are classified, as in Table 3.3-3.5. The proposed methodology uses a combination of original BK subproduct (fuzzy BK subproduct) along with $K7$ and $K9$ inference structures. From Table 3.3-3.5, it can be observed that BK subproduct has very low discrimination as compared to $K7$ and $K9$ respectively. BK subproduct accepts and rejects all the $coast$ images as $open\ country$ scene class as well



(a) Coast

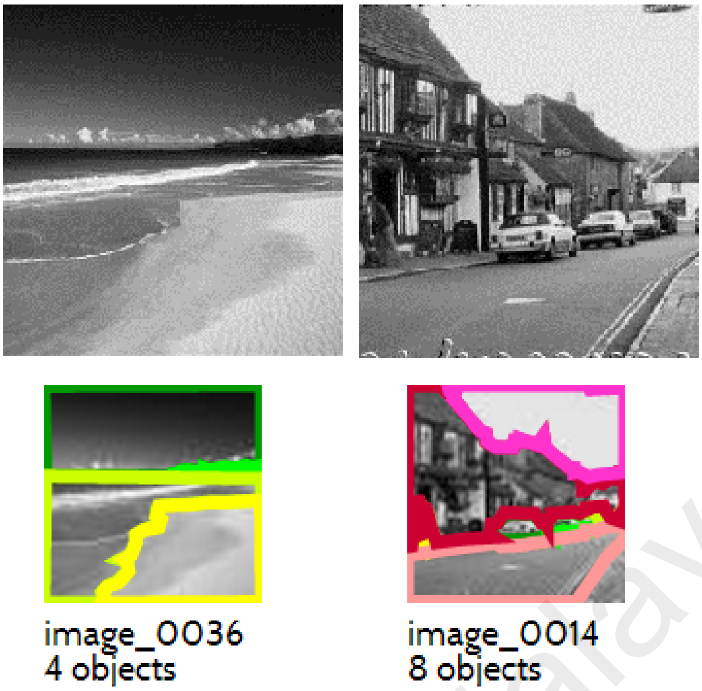
(b) Open country

Figure 3.11: An example of images from coast and open country scene classes with annotated objects.

Table 3.6: Membership function for coast and open country scene classes

Features	Coast	Open Country
Sand	0.000	0.056
Water	0.560	0.122
Sky	0.254	0.281
Tree	0.046	0.348
Mountain	0.140	0.193
Vehicle	0.000	0.000
Road	0.000	0.000
Building	0.000	0.000
People	0.000	0.000

as *street* scene class. For each of the *coast* images to be accepted and rejected as *open country*, this scenario is possible, as proved in Figure 3.11 and Table 3.6, respectively. It can be seen from Figure 3.11 that *coast* and *open country* scenes are very similar in terms of features such as *water*, *sky*, *tree* and *mountain*. Because of the ambiguity it might be difficult for the human to distinguish between the two scene classes. Table 3.6 provides a detailed information about the features and their degree of belongingness to



(a) Coast (b) Street

Figure 3.12: An example of images from coast and street scene classes with annotated objects.

Table 3.7: Membership function for coast and street scene classes

Features	Coast	Street
Sand	0.299	0.000
Water	0.324	0.000
Sky	0.336	0.234
Tree	0.000	0.000
Mountain	0.041	0.000
Vehicle	0.000	0.069
Road	0.000	0.266
Building	0.000	0.419
People	0.000	0.012

the scene classes using the membership function values.

Quantitatively, it has been proven that these two images are correlated, as shown in Table 3.6, since both scene images share some common features such as *water*, *sky*, *tree* and *mountain*. Nonetheless, qualitatively also it has been shown that it is hard for a human to distinguish the scene class of the two scene images as depicted in Figure 3.11. It can either be *coast* or *open country*, different answers will be provided from different human subjects. However, fuzzy BK subproduct also accepts and rejects all the *coast* images as *street* class. From the investigation as shown in Table 3.7, this scenario is impossible as there are no common features that are shared by *coast* and *street*. One of the main reasons that fuzzy BK subproduct is not able to distinguish between the *coast* and *street* class is due to $\emptyset \subseteq Sc$, as identified by C. K. Lim and Chan (2011). It can be noticed that *vehicle*, *road*, *building* and *people* for both *coast* and *street* scene are empty sets.

A further investigation is performed, as shown in Figure 3.12 and Table 3.7. Qualitatively, from Figure 3.12, it is clear that one of the images is *coast* scene, while the other is *street* scene. Table 3.7 also shows quantitatively there are no common features (except the universal feature *sky*) that is shared between *coast* and *street* images.

On comparing the performance of $K7$ and $K9$ inference structures, $K9$ is found to be much more consistent than $K7$ for scene classification. As shown in Table 3.3, $K9$ achieves 100% precision, where it is able to classify all the *coast* images into *coast* class; while $K7$ only achieves 85% accuracy. In Table 3.4, $K9$ presents 95% accuracy compared to 20% by $K7$ in recognizing *open country* images belonging to *open country* class. In Table 3.5, both $K7$ and $K9$ share the same precision results.

One of the main advantages of the proposed approach is that it is able to model the non-mutually exclusive data. It allows an image to belong to multiple classes as opposed to Bosch et al. (2006); Fei-Fei and Perona (2005); Kumar and Hebert (2003); Oliva and Torralba (2001); Vogel and Schiele (2004, 2007), where the classification result is binary. Instead, it classifies each scene image as a combination of different classes using the fuzzy membership function. From Table 3.3, it can be observed that when *open country* scene images are tested against *coast* scene, 14 images are accepted to be *coast* by $K9$. This means an image from *open country* scene class can also belong to a *coast* scene (Figure

3.11).

3.2.3 Performance evaluation

In general, there are several standard evaluation metrics available such as precision, recall, accuracy, F-measure, etc. However, the performance evaluation of the multi-label classification problem (the proposed method) is different from the evaluation of uni-label scene classification problem in the sense that in the multi-label classification the output result can be fully correct, partially correct, or fully incorrect (Boutell, Luo, Shen, & Brown, 2004), hence making the process a little complicated.

For example, say there are three classes $\{c_1, c_2, c_3\}$, and a scene image belongs to c_1, c_2 with a certain degree. Then following results are possible:
c_1, c_2 - fully correct,
c_1 - partially correct, or
c_3 - fully incorrect,
where the results differ from one another in their degree of correctness. In order to evaluate the performance of the proposed scene classification framework, α-evaluation criteria is employed as in (Boutell et al., 2004).

α-evaluation: Let Y_i be the ground truth labels for the test image samples i, and let P_i be the set of prediction labels from the classifier. Then, using α-evaluation, each prediction is given scores using the following formula:

$$\begin{aligned} &score(P_i) = \left(1 - \frac{|\beta M_i + \gamma F_i|}{|Y_i \cup P_i|}\right)^{\alpha} \\ &\forall \alpha \geqslant 0, 0 \leqslant \beta, \beta = 1|\gamma = 1 \end{aligned} \tag{3.5}$$

where, $M_i = Y_i$ - P_i denotes the missed labels, and $F_i = P_i$ - Y_i denotes the false positive labels. The parameters (α, β, γ) allows the false positives and the misses to be penalized differently

Table 3.8: Example of scores as a function of β and γ when the true label is $\{c_1, c_2, c_3\}$, and $\alpha = 1$. c_1 : coast, c_2 : open country and c_3 : street

	Parameter values	Scores
$\alpha = 1$	$\beta = 0.25, \gamma = 1$	0.9000
	$\beta = 1, \gamma = 1$	0.8500
	$\beta = 1, \gamma = 0.25$	0.9125

Table 3.9: Example of α-evaluation scores as a function of α when the true label is $\{c_1, c_2, c_3\}$.

	Parameter values	Scores
$\beta = \gamma = 1$	$\alpha = 0$	1
	$\alpha = 0.25$	0.9602
	$\alpha = 0.50$	0.9220
	$\alpha = 0.75$	0.8852
	$\alpha = 1$	0.8500
	$\alpha = 2$	0.7225
	$\alpha = 10$	0.1969
	$\alpha = \infty$	0

according to the application. Table 3.8 shows the example results after performing α-evaluation on the proposed method, showing how the score varies with different β and γ values. On setting $\beta = \gamma = 1$, simpler formulation is obtained as in Eq. 3.6. Table 3.9 shows some examples of the effect of α on the score.

$$score(P_i) = \left(\frac{|Y_i \cap P_i|}{|Y_i \cup P_i|} \right)^{\alpha} \forall \alpha \geqslant 0 \tag{3.6}$$

Also, to test the feasibility of the proposed method, comparison of fuzzy BK sub-product based scene classification approach is performed with the popular classifiers such as KNN and SVM, as highlighted in Table 3.10. The proposed method supports both multi-label and multi-class classification problem. Multi-label classification refers to the classification problem where multiple target labels are assigned to each data instance.

Table 3.10: Comparison of fuzzy BK subproduct approach based scene classification with other popular classifiers (in terms of scene understanding).

Classifier	Multi-label	Multi-class
KNN	No	Yes
SVM	No	No
Ours	Yes	Yes

Multi-class classification deals with the problem of classifying the data instances into multiple classes. By definition and nature of algorithm, KNN is only multi-class, and SVM is neither multi-label nor multi-class. In terms of overall computational complexity, the proposed method takes $O(NM)$ time where N is the total number of scene classes, and M is the total number of features.

3.3 Summary

As a summary, this section presents the capability of fuzzy BK subproduct to be used as a classifier for 3D video data (HMA) and 2D image data (scene classification). The efficiency in the classification performance delivered by fuzzy BK subproduct is supported with experimental results on the Weizmann human actions dataset and the scene image data (OSR dataset). The advantages of the proposed approach include: the ability to model the non-mutually exclusive data; and the classification results are not binary, instead it classifies each scene image or an action video as a combination of different classes using the fuzzy membership function.

To the best of my knowledge, this is the first attempt of using fuzzy BK subproduct for HMA and scene classification. Most of the fuzzy image processing works have been focusing on applications such as object recognition (DeKruger, Hodge, Bezdek, Keller, & Gader, 2001; Zaki & Abulwafa, 2002), color clustering (Chaira, 2012), edge detection (Bělíček, Kidéry, Kukal, Matěj, & Rusina, 2013), threshold segmentation (Peng, Wang, Pérez-Jiménez, & Shi, 2013), etc. Therefore, this study introduces a new finding where the research community can employ fuzzy BK subproduct approach as a classifier for real-world applications.

CHAPTER 4: EARLY HUMAN ACTION DETECTION

Humans have natural capabilities to perceive and anticipate the actions of other objects they interact with, as well as the happenings in their surrounding. This important aspect of human perception is widely incorporated in the CV systems these days. However, little attention has been given to the problem of early human action detection, which is crucial in several applications ranging from video surveillance to health-care.

Early human action detection refers to anticipating human action as early as possible, i.e. detecting an action after it has begun but before it finishes. In a real-world environment, it is essential to recognize a human action before it is too late such as criminal acts, patients' fall etc. The sooner one can detect an action, the faster one can generate a response. For instance, there is a need to build a system for monitoring the well-being of elderly patients in the hospital. Arguably, a crucial requirement for such a system is its ability to accurately and rapidly detect the patients' fall so that necessary response can be generated in a timely manner. This requires the fall to be detected as soon as possible, before it becomes life threatening and risk the life of the patient.

This section focuses on the early human action detection. The proposed framework is discussed and validated using experiments on real-world human action dataset.

4.1 Introduction

Can an action be detected before it is completed? How many frames are needed to detect an action timely? These are the key requirements for a reliable detector. Figure 4.1 illustrates the idea behind early human action detection. However, the existing detectors are trained to recognize completed action only. They require seeing the entire action video to detect an action. This prevents early detection, as instead partial actions are to be recognized for detecting an action early.

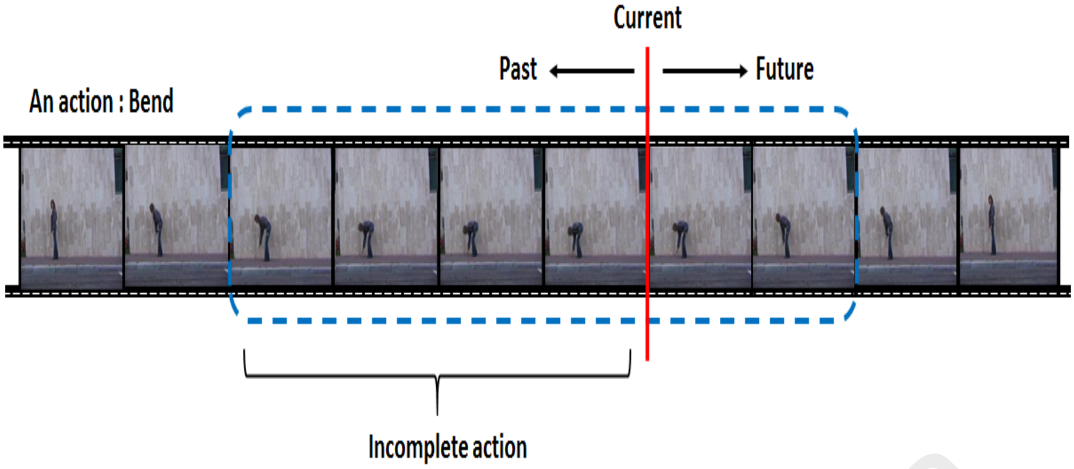


Figure 4.1: Can an action be detected before it is completed? How many frames are needed to detect an action timely? The existing detectors are trained to recognize completed action only. They require seeing the entire action video to detect an action. This prevents early detection, as instead partial actions are to be recognized for detecting an action early.

Therefore, the ultimate goal is to perform early human action detection. However, early human action detection is a daunting task given the vast amount of uncertainty involved therein. An efficient algorithm should be able to handle even the minutest level of uncertainties for reliable and accurate detection. The cumulated errors can further deteriorate the overall system performance. Therefore, a fuzzy approach for early human action detection is proposed.

The section 3.1 studied how fuzzy BK subproduct performs HMA efficiently. The aim is to model fuzzy BK subproduct inference mechanism in a way capable of making decisions as early as possible. This is achieved by modifying the conventional classification problem into frame-by-frame level classification, as illustrated in Figure 4.2. The fuzzy membership function provides the basis to detect an action before it is completed when a certain threshold is attained. Therefore, for a given input video, fuzzy BK subproduct inference engine is invoked at each image frame. The output from each image frame is a membership function value. By modeling the membership function value obtained at each frame, early human action is performed where the detector detects an action when

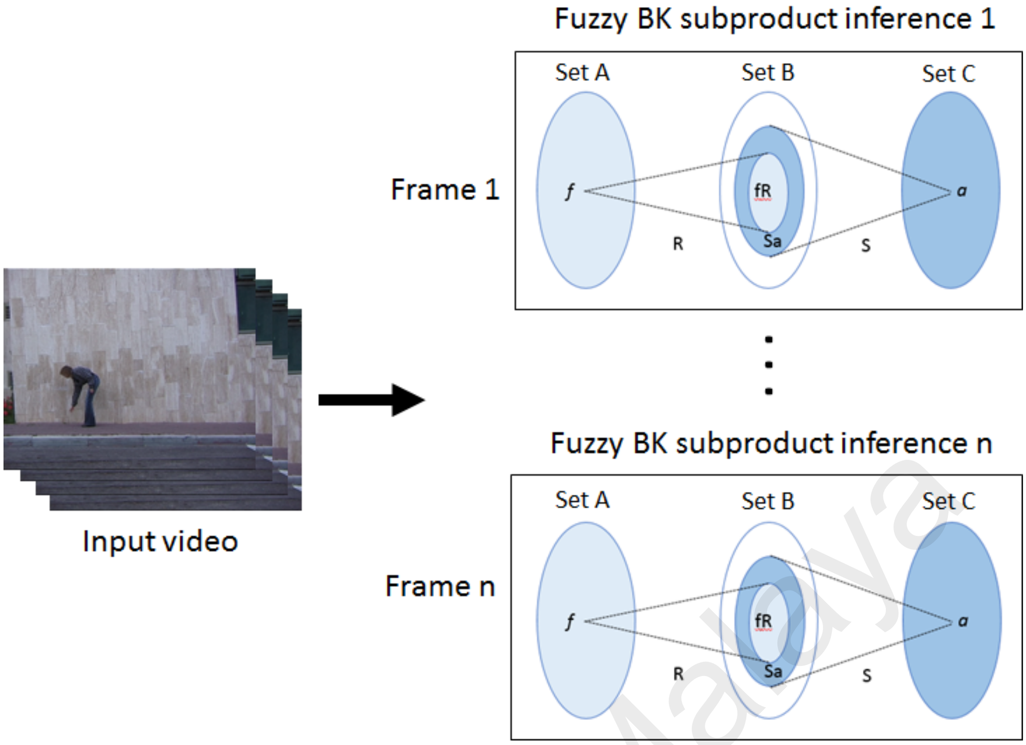

Fuzzy BK subproduct inference 1
Set A
Set B
Set C
Frame 1
f
fR
Sa
a
R
S
Input video
Fuzzy BK subproduct inference n
Frame n

Figure 4.2: Frame-by-frame level classification using fuzzy BK subproduct. The membership function values generated from fuzzy BK subproduct inference engine at each image frame are modeled for early human action detection.

the membership function value exceeds a pre-defined threshold in a suitable way. This is discussed in detail in the following section.

4.2 Proposed Methodology

In this work, an algorithm is proposed for early human action detection that is capable of detecting partial actions, instead of complete action. In specific, the human actions are modeled sequentially frame-by-frame for training fuzzy BK subproduct inference engine, and the detector is learned that is capable of accurately and rapidly performing the classification of the partially observed action sequences. The overall pipeline for the proposed method is highlighted in Figure 4.3.

Fuzzy BK subproduct approach for HMA has been discussed in detail in section 3.1.1. Here, the same methodology is followed which is further extended to perform early detection, as can be observed in Figure 4.3. The modification is done on the classification

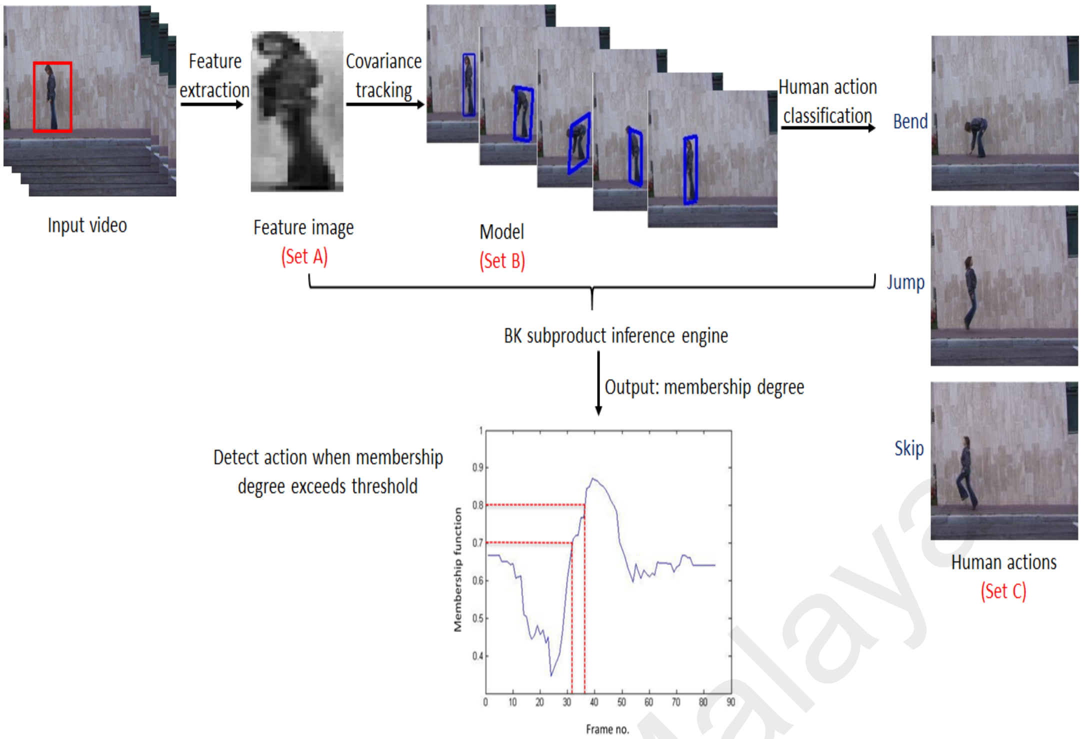


Figure 4.3: Overall pipeline for proposed framework. For a given input video, frame-by-frame BK subproduct inference engine is invoked and action classification is performed. When the membership function values generated from BK subproduct exceeds a certain threshold (e.g. 0.8, 0.7, represented using red dotted lines), the detector detects the action at that particular frame number, enabling early detection.

part where frame-by-frame classification is performed, instead of performing classification after fully observing the video, and thus enabling early human action detection.

As illustrated in Figure 4.3, firstly a feature image is constructed for each image frame following the method in Porikli et al. (2006). A feature vector f_i is constructed using two types of mapping, i.e. spatial attributes based mapping that is obtained from the pixel coordinates values, and appearance attributes based mapping (e.g. color, gradient, infrared, etc.). The feature vector forms the set A in the proposed method, denoted as set $A = \{f_i | i = 1, \cdots, I'\}$.

For a given object region, a covariance matrix of features is then computed as the model of the human object, and the covariance tracking is performed. The human body is then segmented into three parts: head, torso and leg, and covariance tracking is performed on the parts. This results in a part based model m, which forms the set B in the proposed method, denoted as set $B = \{m_j | j = 1, \cdots, J\}$.

The ultimate goal of the proposed method is to perform early human action recognition. Therefore, let set $C = \{a_k | k = 1, \cdots, K\}$ constitutes the actions being performed by human (e.g. bend, jump and skip). The aim is to derive the indirect relationship between between A and C using fuzzy BK subproduct inference mechanism, and further model it to detect an action as early as possible.

To this end, the membership function values generated from fuzzy BK subproduct inference engine at each image frame are modeled for early human action detection. For example, for an action video with n number of frames, invoking fuzzy BK subproduct inference engine for each frame yields a membership function value for each frame as an output. The early detector models the frame-by-frame membership function values generated from fuzzy BK subproduct. An action is triggered when a pre-defined threshold is exceeded monotonically[1]. Even if a single action is being continued, the membership grades are constructed using fuzzy BK subproduct frame-by-frame, and the early detector detects the action in a similar manner. When the membership function value attains the desired threshold value at a certain frame, the detector stops, and triggers the action at that particular frame number.

[1] **Monotonicity requirement:** An important constraint is imposed on early detector function: "monotonicity" requirement, i.e. non-decreasing detection function. This means that the membership degree of a partial action cannot exceed the membership degree of an encompassing partial action. However, the membership degree of a partial action is not required to be greater than that of a future action. Hence, the detector function is desired to be a monotonic and non-decreasing function. Figure 4.4 illustrates the monotonicity requirement for the detector function. This idea is inspired from the work by Hoai and De la Torre (2012), where the monotonicity constraint was imposed on

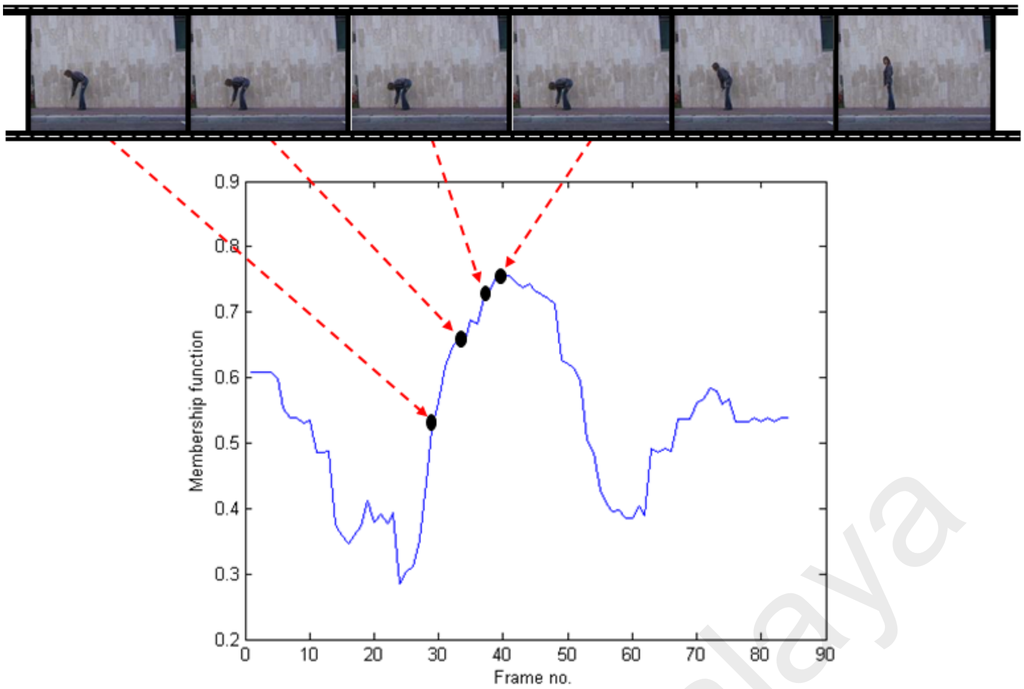


Figure 4.4: [1]Monotonicity requirement for early detection: the membership function of the partial action should always be higher than the membership function of any segment that ends before the partial action.

the detection score function. Here, the early detector is modeled on the basis of the fuzzy membership function values generated from fuzzy BK subproduct inference engine for each image frame.

4.2.1 Learning formulation for early HMA

In this subsection, the learning formulation for early HMA detector will be theoretically justified, and henceforth empirically evaluated in the next section.

Let (X^1, y^1), ... , (X^n, y^n) be the set of series of actions being performed by human and the associated ground truth annotations for the action of interest such that $y^i = [s^i, e^i]$, where s^i denotes the start of the action and e^i denotes the end of the action in the series of actions X^i. Let the length of an action is bounded by l_{min} and l_{max}, and $Y(t)$ denote the set of length-bounded time intervals from the 1^{st} frame to the t^{th} frame in an action video represented as:

$$Y(t) = \{y \in \mathbb{N}^2 | y \subset [1, t], l_{min} \leqslant |y| \leqslant l_{max}\} \cup \{\emptyset\} \tag{4.1}$$

Also, for a series of actions X of length l, let $Y(l)$ denote the set of all possible locations of an action in a video. For an interval $y = [s, e] \in Y(l)$, let X_y denote the subsegment of X from frame s to e inclusive. Then, the output of the detector, which is the segment that has the highest membership value (degree of belongingness to an action) is represented as:

$$D(X) = \underset{y \in Y(t)}{\arg\max}\, \mu(X_y) \tag{4.2}$$

The detector searches the action from l_{min} to l_{max}. If $D(X) = \{\emptyset\}$, it means no action is detected. $\mu(X_y)$ is the membership function representing the membership degree of the segment X_y belonging to the series of actions X^i.

For early human action detection, it is important to model the detector function using partial action frames. This means that the output of the detector on action series X^i at time t is desired to be the partial action, instead of a complete action. Therefore, Eq. 4.2 can be modified to accommodate partial actions as:

$$D(X^i_{[1,t]}) = y^i_t = \underset{y \in Y(t)}{\arg\max}\, \mu(X^i_y) \tag{4.3}$$

where, $D(X^i_{[1,t]})$ denotes the output of the detector on the subsequence of a series of action X^i from the 1^{st} frame to the t^{th} frame only, not the entire X^i.

However, for early human action detection, it is desirable that the membership function $\mu(X^i_y)$ be monotonic and non-decreasing. This means that the membership degree of the partial action y^i_t should always be higher than the membership degree of any segment that ends before the partial action (i.e. seen in the past). And when the membership value exceeds a certain pre-defined threshold monotonically, the detector triggers the occurrence of the action. Therefore, Eq. 4.3 must hold with the desired property, or the monotonicity constraint:

$$\mu(X^i_{y^i_t}) \geqslant \mu(X^i_y)\forall i, \forall t = 1, ..., l^i, \forall y \in Y(t) \tag{4.4}$$

Note that the constraint in Eq. 4.4 is enforced for all $t = 1, 2, \cdots, l^i$, instead of $t = l^i$ because partial actions are trained instead of complete action.

The learning formulation for early human action detection is obtained in Eq. 4.3-4.4, where the membership function $\mu(X^i_y)$ is learned using fuzzy BK subproduct inference engine. The main reason behind using fuzzy BK subproduct for training the detector $D(X)$ is to find a solution closest to how humans anticipate actions in a real-world (i.e. modus ponens way), along with the natural benefits fuzzy sets provide. The trick is to study the indirect relationship between the human subject and the actions being performed in the video. This is achieved by modeling the frame-by-frame arrival of data, and subsequently performing action classification on the basis of the membership function values generated from BK relational products.

As formulated in section 3.1.1, fuzzy BK subproduct inference for HMA is defined as:

$$BK: \ R \triangleleft_{BK} S(f, a) = \frac{1}{J}\sum_{m \in B}(R(f, m) \rightarrow S(m, a)) \tag{4.5}$$

where, $R(f, m)$ is the membership function of the relation R between the features f, and the human body part-based model m; $S(m, a)$ is the membership function of the relation S between m and the human actions a.

Therefore, replacing $R \triangleleft_{BK} S(f, a)$ in Eq. 4.5 with $\mu(X^i_y), \forall i, \forall t = 1, ..., l^i, \forall y \in Y(t)$, yields the desired membership function (Eq. 4.6) required for early human action detection. When the membership function value monotonically exceeds a certain threshold, the detector detects the action.

$$\mu(X_y^i) = R \triangleleft_{BK} S(f,a) = \frac{1}{J}\sum_{m\in B}(R(f,m) \rightarrow S(m,a)) \tag{4.6}$$

4.2.2 Study on the semantic relationship between human and the action

Early human action detection can also be defined in terms of the semantic relationship between human and the action. Given an input set of training series of action sequences X^1, X^2, ..., X^n performed by a human and the associated ground truth annotations y^1, y^2, ..., y^n for the action of interest, it is assumed that each training action sequence contains at most one action of interest, as a training sequence containing several actions can always be divided into smaller subsequences of a single action. Therefore, y^i = [s^i, e^i] consists of two numbers that indicate the start and end of the action in the time series of action X^i respectively. Early human action detection aims at finding the semantics (human - action) in a set of series of actions (X^1, y^1), ..., (X^n, y^n) where $y^i \subset$ [s^i, e^i]. However, the semantics (human - action) remain invariant if all the frames have been used. If so, a Silico DNA based computing is considered to serve the purpose effortlessly. For example, in Ullah, D'Addona, and Arai (2014), a DNA based computing approach for understanding complex shapes have been proposed where the authors have shown that whatever may be the outlook of the image frames, they underlie the same semantics (fern-leaf).

However, the method in Ullah et al. (2014) is applicable only to two-dimensional image data. The proposed method can handle these issues, in the sense that there cannot possibly exist a situation where all the frames have been used to detect an action as then it will be same as the conventional classification problem which requires seeing a complete action. Instead, the early detector is trained to detect partial actions. This means that for an interval $y = [s, e] \in Y(l)$, where $Y(l)$ denote the set of all possible locations of an action in a video, and X_y denote the subsegment of X from frame s to e inclusive, the detector $D(X^i_{[t_0,t]})$ outputs the segment having the highest membership degree of belongingness to an action i.e. $\mu(X^i_y)$, which is a partial segment y^i_t instead of a complete action y.

Furthermore, Eq. 4.4 is modified by adding an additional variable $\Delta(y^i_t, y)$, which is the loss of detector for outputting y when the desired output is y^i_t, and represented as follows:

$$\mu(X^i_{y^i_t}) \geqslant \mu(X^i_y) + \Delta(y^i_t, y), \forall i, \forall t = 1, ..., l^i, \forall y \in Y(t) \tag{4.7}$$

where $\Delta(y^i_t, y)$ handles the exceptional case where all the frames have been used and the detector fails to detect the occurrence of an action before it finishes.

4.3 Validation

In order to test the efficiency of the early human action detector, the experiments are performed on the Weizmann human actions dataset (Gorelick et al., 2007). The experimental set up follows section 3.1.2.

The following pseudo-code represents the implementation of early human action detection using fuzzy BK subproduct in a step-by-step manner:

Step 1: Input an action video.

Step 2: Perform human detection.

Step 3: Segment the human body into three parts: head, torso+arm and leg.

Step 4: Perform feature extraction by constructing covariance matrix of features.

Step 5: Feature image is constructed. Save in Set A.

Step 6: Perform part-based covariance tracking.

Step 7: Human body part-based models are obtained. Save in Set B.

Step 8: Normalize the results obtained using min-max normalization. Save as membership function R.

Step 9: Obtain the converse relation S between actions and models by normalizing the tracking results using min-max normalization.

Step 10: Call BK subproduct inference engine frame-by-frame utilizing R and S.

Step 11: Set the threshold value as the cutting point for the detector.

Step 12: When membership degree exceeds the threshold value, stop.

Step 13: Output the frame number.

Step 14: Done.

The preprocessing of images, feature extraction and covariance tracking are performed using the method in Porikli et al. (2006), modified to generate part-based human body model. Sample human motion tracking results are illustrated in Figure 3.4, and the human body part-based model in Figure 3.5, of section 3.1.2.

Furthermore, the membership function values are generated for the relation between image features and the models (m_1, m_2, m_3), normalizing the results obtained from the covariance tracking. Table 4.1 presents the examples (out of total 576 frames) of the membership function values (R) generated from the one-to-many relationship between set A and set B. The membership function S is obtained by studying the relationship between the models and the actions being performed. Table 4.2 represents the membership degree S generated for the relation between set B and set C, with each model having a degree of belongingness to an action (one-to-many relationship).

Obtaining R and S frame-by-frame, BK subproduct inference engine is invoked, and by empirically formulating Eq. 4.6, human action classification is performed. As partial human actions are modeled instead of the complete actions, the detector is capable of detecting an action before it finishes (i.e. early detection). For the experiments, the action classification performance is tested using three inference structures: original BK (i.e.

Table 4.1: Example of membership degree $R(f,m)$ generated for relation between set A and set B.

Frame no.	m_1	m_2	m_3
1	0.9856	0	0
20	0.7318	0.6055	0
25	0.4296	0.9976	0
30	0.1592	0.6964	0
39	0	0.2604	0
50	0.4001	0.3723	0
55	0.7512	0.2094	0
67	1	0.0358	0
75	0.9852	0.0572	0
152	0.2418	0	0.4615
156	0.3634	0.0494	1
162	0.3465	0.1613	0.7692
172	0.6104	0.3442	0.3077
193	0.8549	0.7118	0
200	0.3603	0.8025	0.9231

Table 4.2: Example of membership degree $S(m,a)$ generated for relation between set B and set C.

Model	Bend	Jump	Skip
m_1	0.80	0.80	0.90
m_2	0.20	0.80	0.90
m_3	0.10	0.30	0.70

fuzzy BK subproduct) which is modified to suit the application requirements (Eq. 4.6), $K7$ (Eq. 2.6) and $K9$ (Eq. 2.7). It is found that overall original BK subproduct performed the best for all the three action classes, as can be seen in Table 4.3. However, for $jump$ and $skip$, originalBK performed better in the initial few frames, and later original BK and $K9$ delivered similar performance. $K7$ performed fairly poorer for all the three action classes.

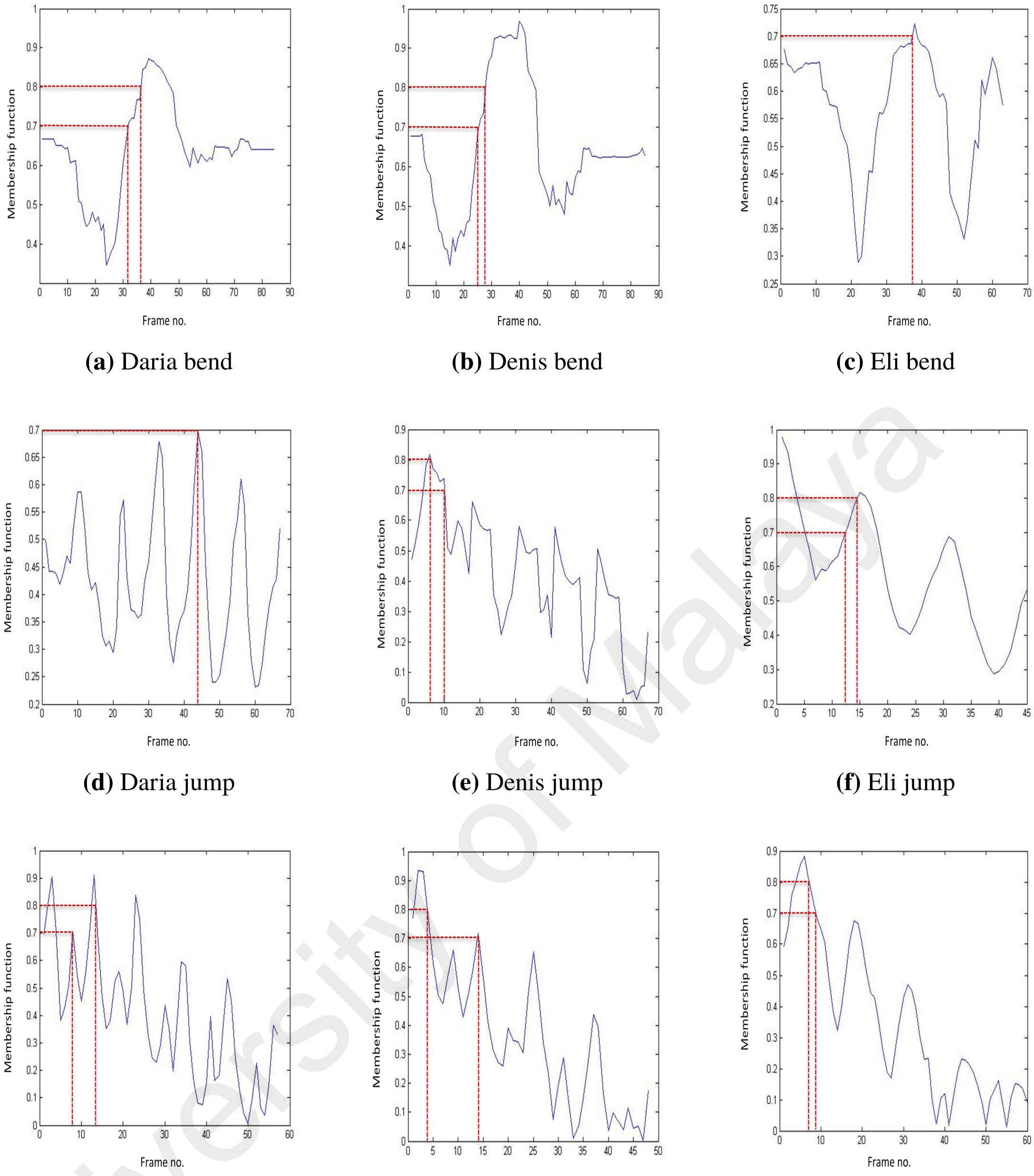
The proposed early detector detects an action when the membership function values (as presented in Table 4.3) exceeds the pre-defined threshold monotonically. For the experiments, the threshold values is set as 0.8 and 0.7. Table 4.4 highlights the results

Table 4.3: Results obtained after applying Original BK subproduct (fuzzy BK subproduct), $K7$ and $K9$ inference structures.

Action	Frame no.	Original BK	K7	K9
BEND	1	0.6774	0	0.0486
	10	0.4847	0	0.2967
	20	0.4234	0.2481	0.2967
	30	0.8798	0	0.2967
	40	0.9684	0	0.0818
	50	0.5255	0	0.2967
	60	0.5742	0.0050	0.2686
	70	0.6246	0.0437	0.1150
	80	0.6257	0.0404	0.1150
	85	0.6281	0	0.1460
JUMP	1	0.4726	0.4166	0.4726
	10	0.7386	0	0.2171
	20	0.5868	0	0.3116
	30	0.4595	0	0.4595
	40	0.2150	0.2150	0.2150
	50	0.0624	0.0624	0.0624
	60	0.1116	0.1116	0.1116
	67	0.2295	0.2295	0.2295
SKIP	1	0.7708	0	0.2245
	10	0.5329	0.0169	0.5329
	20	0.3913	0.3469	0.3913
	30	0.1936	0.1936	0.1936
	40	0.0361	0.0361	0.0361
	48	0.1740	0.1740	0.1740

obtained using original BK, and Figure 5.7 represents the experimental results graphically. Following observations are made from the results obtained from early human action detection on setting different threshold values:

(i) When threshold=0.8, the detector detects the *bend* action for the first actor (Daria) from seeing ~42% of the frames, and for the second actor (Denis) from seeing ~32% of the frames. However, it slightly missed the detection for the third actor (Eli) due to the fewer number of image frames in the action video. For the *jump* action, the detector failed to detect the action for Daria in high threshold. However, for Denis



(a) Daria bend **(b)** Denis bend **(c)** Eli bend

(d) Daria jump **(e)** Denis jump **(f)** Eli jump

(g) Daria skip **(h)** Denis skip **(i)** Eli skip

Figure 4.5: Graphical results for early human action detection for *Bend*, *Jump*, and *Skip* performed by three actors (Daria, Denis and Eli). The threshold values are set as 0.8 and 0.7 (represented using red dotted lines), and the detector detects the action when the membership function value exceeds the threshold monotonically. On an average, the detector is able to detect an action from seeing ~32% of the frames.

and Eli, the detector efficiently detected the action upon seeing ~8% and ~31% of the frames respectively. Lastly, for the *skip* action, the detector is able to detect the action seeing on an average ~14% of the frames.

(ii) When threshold=0.7, the detector is able to detect all the actions performed by all the

Table 4.4: Results for early human action detection.

Actor and action	Total no. of frames	Detect at frame no. (threshold=0.8)	Detect at frame no. (threshold=0.7)	Early detection seeing %age of frames
Daria bend	84	36	31	36.90
Denis bend	85	28	25	29.41
Eli bend	63	-	37	58.73
Daria jump	67	-	44	65.67
Denis jump	67	6	10	14.92
Eli jump	45	14	12	26.66
Daria skip	57	13	8	14.03
Denis skip	48	4	14	29.16
Eli skip	60	7	8	13.33

three actors upon seeing ∼32% of the frames on an average. Therefore, the proposed early human action detector is capable of efficiently detecting an action before it is completed, seeing only a few number of initial frames.

4.4 Summary

Herewith, a framework is proposed for detecting human action as early as possible using fuzzy BK subproduct inference mechanism. Human action classification problem is modified into frame-by-frame level classification to enable early detection. Based on the best of my knowledge and a recent survey paper by C. H. Lim et al. (2015), this is the first work in the fuzzy community dealing with early human action detection. The closest research to this work is MMED proposed in Hoai and De la Torre (2012, 2014). In terms of timeliness and accuracy of detection, MMED outperforms the other algorithms. For human action recognition using Weizmann human actions dataset, MMED requires seeing ∼40% of the action (with a score of 0.7). In this work, the experiments are performed using the same human action dataset, and it is found that the detector significantly outperforms MMED where the detector requires seeing ∼32% of the image frames on an average in an action video (with membership function score of 0.7).

In summary, the proposed method is capable of making reliable early human action detection by modeling the partial actions. The membership values generated from fuzzy

BK subproduct inference engine provides the basis to detect an action before it is completed when a certain threshold is attained. The efficiency in the performance delivered by the early detector is supported with the experimental results, where the detector is able to detect an action from seeing ~32% of the frames on an average.

CHAPTER 5: HYBRID TECHNIQUE FOR EARLY HMA

The proposed early human action detection framework is further analyzed from a broader perspective where it is represented as a hybrid model of CV and fuzzy set theory. Hybrid techniques are well-known in addressing issues such as uncertainty, imprecision and vagueness to a considerable extent by exploiting the strengths of one technique to alleviate the limitations of another (Acampora et al., 2012; Hosseini & Eftekhari-Moghadam, 2013). Therefore, in this work a hybrid technique for early human action detection is proposed as the synergistic integration of CV solutions and fuzzy set theory that is based on fuzzy BK subproduct approach.

In this section, the proposed hybrid technique for early human action detection is discussed in detail, further validated with experimental results on publicly available human action dataset for a variety of action classes. The aim is to carry out reliable early human action detection and infer an action upon observing minimum possible number of image frames.

5.1 Introduction

In general, CV methods and fuzzy approaches do not behave in a conflicting manner, rather compliment one another (C. H. Lim et al., 2015). The fusion of these techniques towards performing human action recognition as early as possible can be achieved through proper hybridization. To this end, the relationship between a human and the action being performed is studied using fuzzy BK subproduct, efficiently integrated with CV techniques including feature extraction and motion tracking to perform human action recognition effectively.

Another issue addressed by the proposed hybrid method is to handle the cumulative tracking errors and precision problem that can affect the overall system performance. A set of overlapped fuzzy numbers known as the fuzzy qualitative quantity space are used as a solution, where individual distance among them is defined by a pre-defined metric (H. Liu & Coghill, 2005). FQS helps in modeling the accumulated tracking errors and precision problem because of the uncertainties arising due to different height, size and step size of each human.

Furthermore, a deep study is performed on the impact of various fuzzy implication operators and inference structures in retrieving the relationship between the human subject and the action. The existing fuzzy implication operators are capable of handling 2D data only. However, a third dimension 'time' plays a crucial role in human action recognition to model human movement changes over time. Therefore, a new space-time fuzzy implication operator is introduced, by modifying the existing implication operators to accommodate time as an added dimension.

It is intended to provide a solution for early human action detection closest to natural human perception. The novelty lies in the hybrid based learning formulation to train the early detector such that once the detector has been trained, it can be flexibly used in several ways according to different types of application.

5.2 Proposed Methodology

In this work, a hybrid technique is proposed for early HMA. The proposed hybrid solution performs hybridization on the generated tracking output and fuzzy BK subproduct. Figure 5.1 highlights the overall pipeline of the proposed methodology. The three main steps involved are: feature extraction, human motion tracking (covariance tracking) and early human action detection (using hybrid technique). Frame-by-frame membership function is constructed for each kind of possible movement, taking into account several human actions from a publicly available dataset. The partial human action is modeled, where the fuzzy membership function provides the basis to detect an action before it is completed.

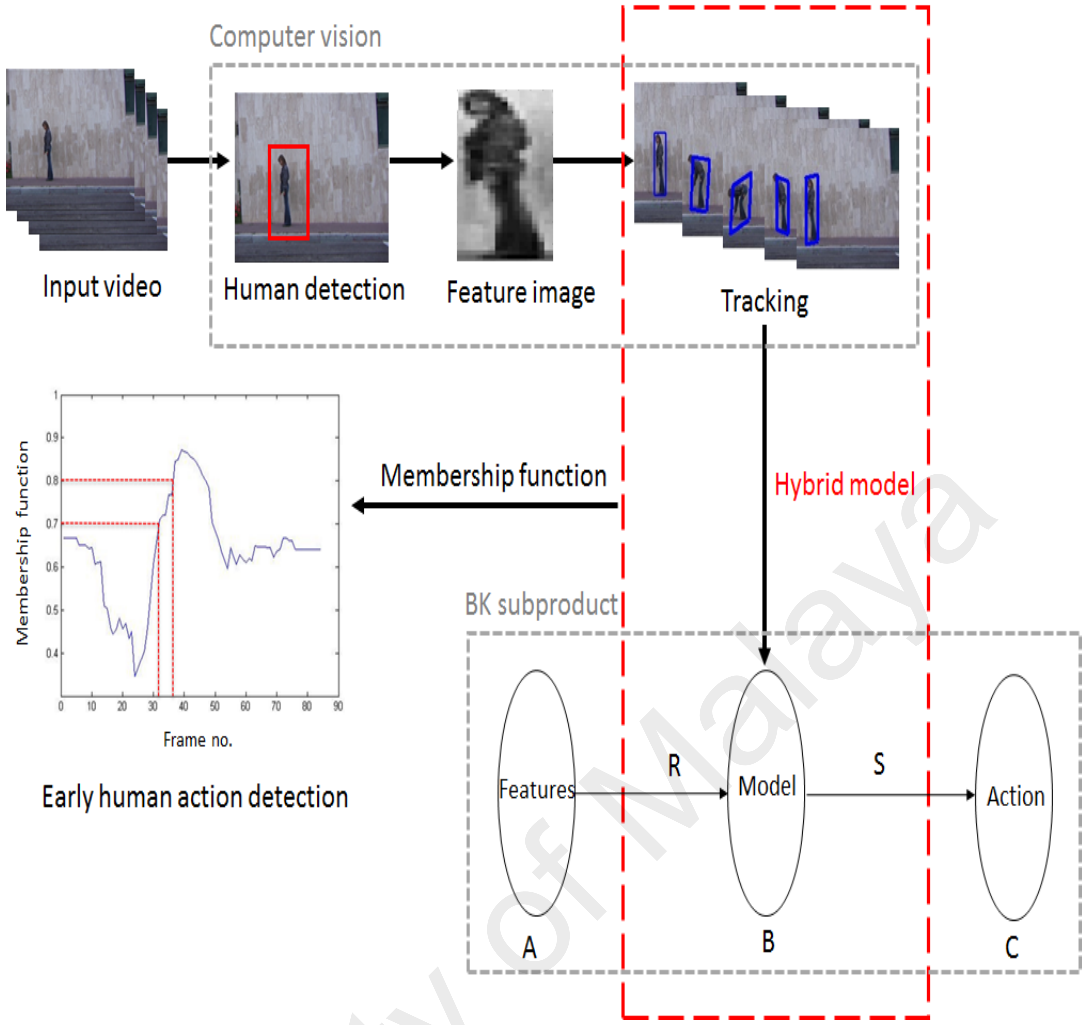


Figure 5.1: Overall pipeline of the proposed hybrid technique. The hybridization is performed on the tracking output from CV solutions and the set B of fuzzy BK subproduct which includes a set of human body part-based models obtained from the human motion tracking. Red colored dotted lines represent the hybridization.

This is achieved when a certain threshold is attained in a suitable way. The overall process is discussed step-by-step in this section as follows.

5.2.1 Feature extraction

Given an input video of action sequences, the object window is represented as a covariance matrix of features following the method in Porikli et al. (2006). This enables capturing the spatial and the statistical properties along with their correlation within the same representation. Figure 5.2 highlights the pixel-wise feature representation, where an object window is represented as a covariance matrix of features.

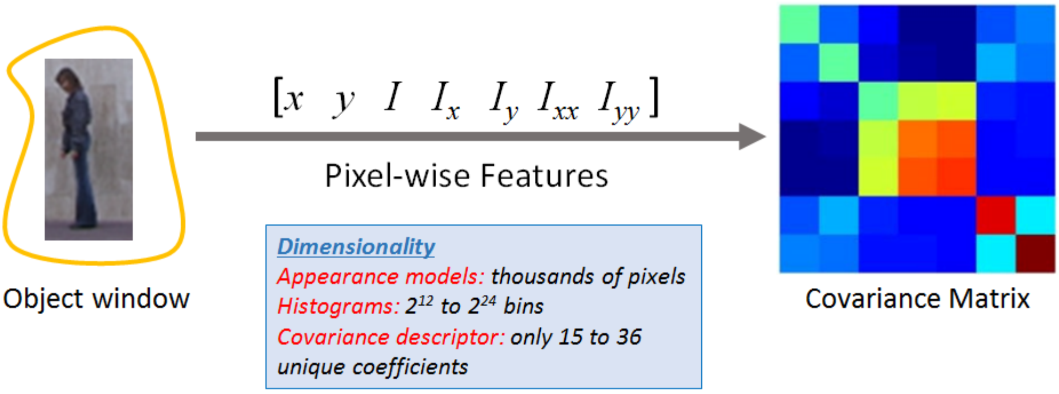

Object window
$[x \;\; y \;\; I \;\; I_x \;\; I_y \;\; I_{xx} \;\; I_{yy}]$
Pixel-wise Features
Dimensionality
Appearance models: thousands of pixels
Histograms: 2^{12} to 2^{24} bins
Covariance descriptor: only 15 to 36 unique coefficients
Covariance Matrix

Figure 5.2: Pixel-wise feature representation of an object window using a covariance matrix of features. In the covariance matrix, color model is used here to represent the object region.

Let F be the $W \times H \times d$ dimensional RGB feature image of an image I, such that $F(x, y) = \Phi(I, x, y)$, where the function Φ can be any mapping such as image gradients, color, edge magnitude or orientation. Let $\{f_i\}_{i=1..I'}$ be the d-dimensional feature vector inside a rectangular window R' where $R' \subset F$. A feature vector f_i is constructed using: (i) spatial attributes based mapping - obtained from the pixel coordinates values, and (ii) appearance attributes based mapping - e.g. gradient, color or infrared. The features extracted may be associated directly with the pixel coordinates ($f_i = [\, x \;\; y \;\; I(x, y) \;\; I_x(x, y) \dots]$), or can be arranged in a radially symmetric relationship ($f_i^r = [\, \|(x', y')\| \;\; I(x, y) \;\; I_x(x, y) \dots]$).

5.2.2 Covariance tracking

Human motion tracking is important in finding the correspondences between the previously detected objects in the current image frame. A common approach in tracking is to employ predictive filtering, where the object's location in the distance calculation and color attributes are used to update the model (Wren, Azarbayejani, Darrell, & Pentland, 1997). When the measurement noise is assumed to be Gaussian, Kalman filter (Kalman, 1960) offers an optimal solution. Whereas, Markovian filters can be applied for tracking when the state space consists of a finite number of states. Another well-known approach is to employ particle filters (Isard & Blake, 1998), which are based on Monte Carlo integration methods. In particle filtering, the current state density (i.e. speed, size, location) is represented using a set of random samples with associated weights. Furthermore, the new density is computed utilizing these samples and weights. However, the main disadvantage of particle filtering is that it is based on random sampling. Therefore, it suffers from the problem of sample degeneracy and impoverishment, especially for higher dimensional representations (Porikli et al., 2006).

In order to find a global optimal solution, the covariance tracking method presented in Porikli et al. (2006) is employed. It is a simple algorithm used to track non-rigid objects using covariance based object description. A model update mechanism is incorporated using Lie algebra (Porikli et al., 2006) to adapt to the undergoing object deformations and appearance changes. Unlike other tracking methods, covariance based tracking does not make any assumption on the measurement noise and the motion of the objects tracked. It has shown remarkable detection accuracy for the moving objects in non-stationary camera sequences. As discussed in the previous chapter, the covariance tracking is performed as follows.

For a given object region R', a $d \times d$ covariance matrix of features $C_{R'}$ is computed as the model of the human object:

$$C_{R'} = \frac{1}{MN} \sum_{i=1}^{MN} (f_i - \mu_{R'})(f_i - \mu_{R'})^T \tag{5.1}$$

where, $\mu_{R'}$ is the vector of the mean of the corresponding features for the points in region R'. A single covariance matrix extracted from a region is sufficient to perform matching of the region in multiple views and poses. In the current image frame, the region that has the minimum covariance distance from the model is located and assigned as the estimated location.

Furthermore, the covariance tracking algorithm is modified to perform part-based human motion tracking. Human body is segmented into three parts: head, torso and leg, and the covariance tracking is performed on each of the part, resulting in a part-based model m. In order to model several distinct human actions, five models are generated:

(i) Head distance - model the head movement from start to end frame.

(ii) Body distance - model the position changes of the human body from the first frame.

(iii) Leg distance - model the distance between both legs.

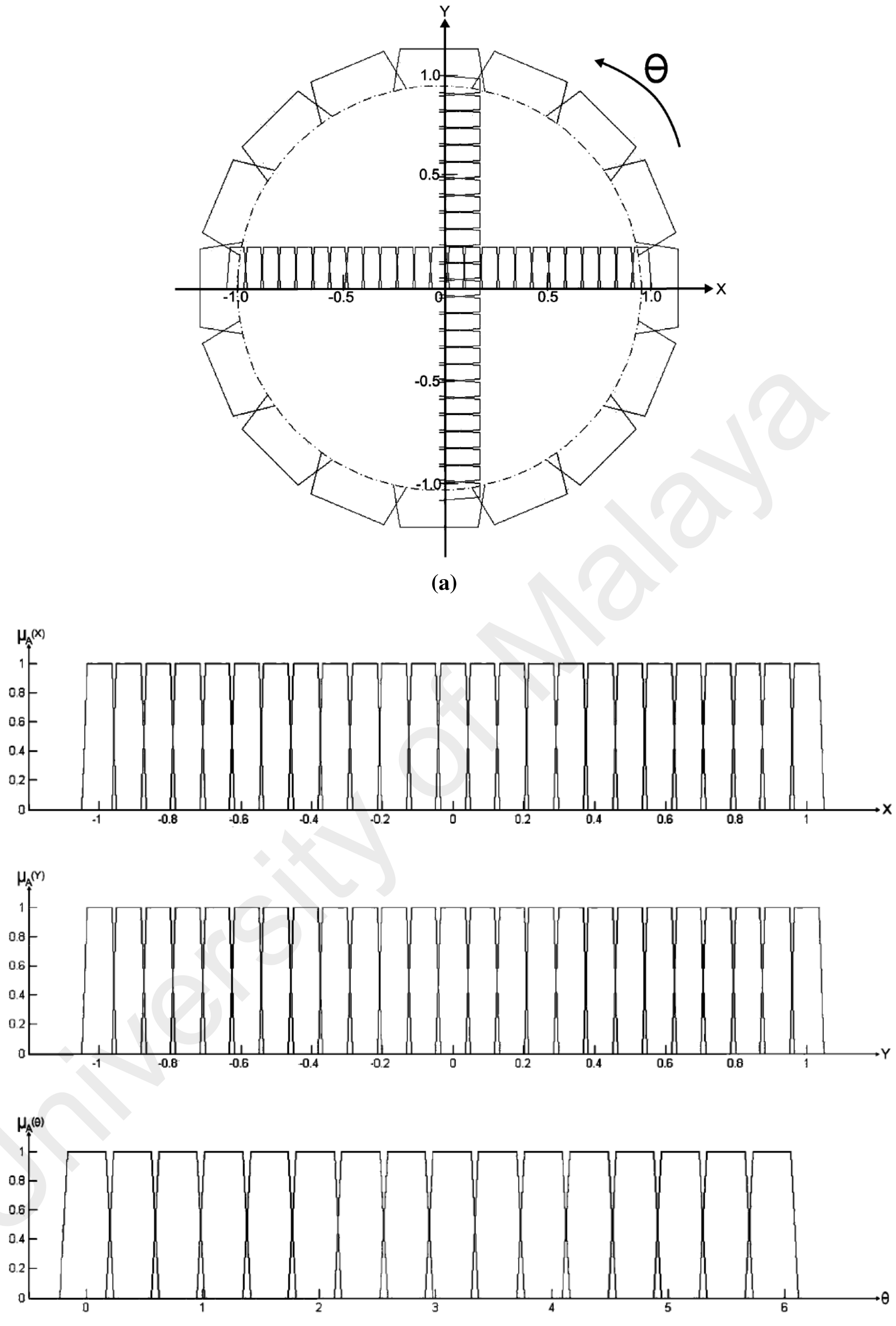
(iv) Hand distance - model the hand movement from start to end frame.

(v) Ground distance - model the distance of the human body from the ground.

It is crucial for the tracking algorithm to be free from problems such as tracking precision issue resulted due to the position changes of each body part (head, torso and leg) evolving over time. Also, the tracking algorithm should take into account the cumulative errors generated because of the uncertainties arising due to different height, size and step size of each human. These problems can directly affect the performance of the higher level task. Therefore, the tracking output is fuzzified using fuzzy qualitative quantity space, as discussed in the following subsection.

5.2.2.1 Fuzzy qualitative quantity space

The fuzzy qualitative quantity space can be defined as a set of overlapped fuzzy numbers whose individual distance among them is defined by a pre-defined metric (Chan & Liu, 2009). Four tuple fuzzy numbers [a, b, α, β] are employed to describe each state in the fuzzy qualitative unit circle (Figure 5.3a) that is a finite and convex discretization of the real number line. In this work, the main motivation behind employing the fuzzy qualitative unit circle is to model the accumulated errors due to the position changes of



(a)

(b)

Figure 5.3: (a) Conventional unit circle: The Cartesian translation and the orientation is replaced by the fuzzy quantity space. (b) Element of the fuzzy quantity space for every variable (translation (X, Y), and orientation θ) in the fuzzy qualitative unit circle is a finite and convex discretization of the real number line (Chan & Liu, 2009).

each body part (head, torso and leg) evolving over time. Besides that, this approach helps in dealing with the tracking errors and precision problem because of the uncertainties arising due to different height, size and step size of each human. In the hybrid model, fuzzy quantity space helps in the normalization of the tracking output generated as a result of the part-based covariance tracking

In the proposed method, the rigid motion of each body part is represented using the fuzzy qualitative translation states. A fuzzy qualitative unit circle as presented in Figure 5.3 is constructed using Eq. 5.2, by following the approach in Chan and Liu (2009):

$$\lim_{s \to s_0 = 10} C_t(s) = QS(qp_l) \tag{5.2}$$

where, the translation component in the conventional unit circle is replaced by the fuzzy qualitative quantity space; and s denotes the number of states representing the $x - y$ translation employed in the quantity space to represent the fuzzy qualitative unit circle. Empirically, the translation is selected as $s = 10$. The fuzzy qualitative quantity space Q consists of the translation component Q^d represented as:

$$Q^d = QS_d(l_j), \quad for j = 1, 2, ..., n \tag{5.3}$$

where, $QS_d(l_j)$ denotes the state of a distance l_j, and n represents the number of elements in the translation component. The final output generated is the fuzzified tracking result, normalized using the fuzzy qualitative quantity space with values between 0 and 1.

5.2.3 Hybrid Model

The output from the human body part-based covariance tracking, normalized using fuzzy qualitative quantity space, is integrated with fuzzy BK subproduct with proper hybridization process to perform human action recognition, as is presented in Figure 5.1.

Given an input action video, let $A = \{f_i | i = 1, \cdots, I\}$ denote the set of features extracted from image frames i of the video describing the human action. Let set $C = \{a_k | k = 1, \cdots, K\}$ be the set of human action. A has no direct relation with C, since there is no information about which action is being performed and by whom. However, if there exists an intermediate set B, which is in relation with both A and C, the indirect relationship between A and C can be derived using fuzzy BK subproduct, and utilize this information to detect an action as early as possible. Therefore, let set $B = \{m_j | j = 1, \cdots, J\}$ constitute the human body part-based model, obtained as a result of covariance tracking. Using this intermediate set, the relationship between image features f in set A and the action a in set C can be therefore obtained as:

$$R \triangleleft S = \{(f, a) | (f, a) \in A \times C \text{ and } fR \subseteq Sa\} \tag{5.4}$$

where $fR \subseteq Sa$ is the main element in retrieving the relationship between f and a, and is obtained from the covariance tracking. The composition of relation between $f_i \in A$ and $a_k \in C$ can be defined using the fuzzy subsethood measure as follows:

$$BK : \ R \triangleleft_{BK} S(f, a) = \frac{1}{J} \sum_{m \in B} (R(f, m) \rightarrow S(m, a)) \tag{5.5}$$

where, $R(f, m)$ represents the membership function of the relation R between f and m, and $S(b, c)$ represents the membership function of the relation S between m and a. Therefore, Eq. 5.5 represents the hybrid model mathematically. The hybrid model performs the integration of the models obtained from human motion tracking into the intermediate set B of BK subproduct. As a result, set B includes five distinct models $m_1 - m_5$ generated from the covariance tracking, i.e. head distance, body distance, leg distance, hand distance, and ground distance.

For each image frame, the membership function values generated from Eq. 5.5 are modeled for early human action detection. For example, for an action video with n number of frames, invoking fuzzy BK subproduct inference engine for each frame will yield a membership function value for each frame as an output. The early detector models the frame-by-frame membership function values generated from fuzzy BK subproduct and triggers an action when it exceeds a pre-defined threshold monotonically. Even if a single action is being continued, the membership grades are constructed using fuzzy BK subproduct (Eq. 5.5) frame-by-frame, and the early detector detects the action in a similar manner. When the membership function value attains the desired threshold value at a certain frame, the detector stops, and triggers the action at that particular frame number. Section 5.2.4 explains the overall process in detail.

5.2.4 Early Anticipation of Human Action

Early anticipation of human action involves processing in real-time. The detector reads from a stream of input video and keeps a sequence of observations in the memory. It continuously monitor the occurrence of the target action. If the target action is detected, the frame number at which the detector triggers is returned.

However, in order to detect an action as early as possible, partial action are used as positive training examples, instead of a complete action sequence. Let (X^1, y^1),$\cdots$,(X^n, y^n) be the set of a series of actions performed by a human and the associated ground truth annotations for the action of interest such that y^i = [s^i, e^i], where s^i denotes the start of the action and e^i denotes the end of the action in the time series of action X^i. Let t_0 denote the beginning of the action video, and the length of the partial and complete action that the detector needs to detect be bounded by l_{min} and l_{max}. Let $Y(t_0, t)$ denote the set of length-bounded time intervals from time t_0 to time t. Also, for a time series of action X

of length l, let $Y(l)$ denote the set of all possible locations of an action in a video. For an interval $y = [s, e] \in Y(l)$, let X_y denote the subsegment of X from frame s to e inclusive. Then, the output of detector that is the segment having the highest membership value (degree of belongingness to an action) is represented as:

$$D(X^i_{[t_0,t]}) = y^i_t = \underset{y \in Y(t_0,t)}{\arg\max}\, \mu(X^i_y) \tag{5.6}$$

where, $D(X^i_{[1,t]})$ denotes the output of detector on the subsequence of X^i from the initial frame to the t^{th} frame only, instead of entire X^i. If $D(X^i_{[t_0,t]}) = \{\emptyset\}$, no action is detected. $\mu(X^i_y)$ represents the membership function of the segment X^i_y belonging to the time series of action X^i. Similarly, the detector's output at $t + 1$ can be computed as:

$$D(X^i_{[t_0,t+1]}) = y^i_{t+1} = \underset{y \in Y(t_0,t+1), y(2)=t+1}{\arg\max} \mu(X^i_y) \tag{5.7}$$

where y^i_{t+1} is the segment that attains the maximum membership function at $t + 1$. The overall computational cost involved for the detection is $O(l)$, where $l_{min} \leq l \leq l_{max}$.

For early human action detection, it is desirable for the membership function $\mu(X^i_y)$ to be monotonic and non-decreasing. Therefore, Eq. 5.6 must hold with the desired property:

$$\mu(X^i_{y^i_t}) \geqslant \mu(X^i_y) \forall i, \forall t = 1, ..., l^i, \forall y \in Y(t) \tag{5.8}$$

The constraint in Eq. 5.8 is enforced for all $t = 1, 2, \cdots, l^i$, instead of $t = l^i$ as the partial actions are being trained instead of a complete action. The learning formulation for early human action detection is obtained as in Eq. 5.6-5.8, where the membership function $\mu(X^i_y)$ is learned using the proposed hybrid technique.

In this work, the target action of multiple classes is detected. Therefore, the detectors are trained and used separately for each of the target action classes. The challenge is to study the indirect relationship between the human subject and the action being performed in the video, modeling the frame-by-frame arrival of data, and subsequently perform action classification on the basis of the membership function values generated from the hybrid model. Therefore, Eq. 5.5 can be re-written as:

$$\begin{gathered}\mu(X_y^i) = R \triangleleft_{BK} S(f,a) = \frac{1}{J}\sum_{m\in B}(R(f,m) \rightarrow S(m,a)) \\ \forall i, \forall t = 1,\dots,l^i, \forall y \in Y(t)\end{gathered} \tag{5.9}$$

where Eq. 5.9 yields the desired membership function required for early human action detection. When the membership function value monotonically exceeds a pre-defined threshold, the detector triggers the action.

5.3 Impact of Implication Operators

An important property of Eq. 5.9 to be taken into consideration is which implication operator '→ ' to employ to infer the relation '$R(f,m) \rightarrow S(m,a)$'. Let r and s defines the membership functions for relations R and S respectively. There exists a number of fuzzy implication operators in the literature (C. K. Lim & Chan, 2015). For example:

(i) Standard Sharp ($S\#$): It is represented as $r \rightarrow_{\mathrm{S\#}} s$. The standard sharp operator outputs 1 iff $r \neq 1$ or $s = 1$, and outputs 0 otherwise.

(ii) Standard Strict (S): It is represented as $r \rightarrow_{\mathrm{S}} s$. The standard strict operator is defined as 1 iff $r \leq 1$, and 0 otherwise.

(iii) Gaines 43 ($G43$): It is represented as $r \rightarrow_{\mathrm{G43}} s$. This fuzzy implication operator is defined as: $\min(1, \frac{r}{s})$

(iv) Kleene-Dienes operator (KD): It is a popularly used fuzzy implication operator and is represented as $r \rightarrow_{\mathrm{KD}} s$. KD operator is defined as: $\max(s, 1 - r)$.

(v) Reichenbach (R): Reichenbach operator is represented as $r \rightarrow_{\mathrm{R}} s$. It is mathematically defined as: $1 - r + rs = \min(1, 1 - r + s)$.

(vi) Łukasiewicz operator (L): Łukasiewicz operator is another well-known implication operator represented as $r \rightarrow_{\text{Ł}} s$, and defined as: $\min(1, 1 - r + s)$.

(vii) Yager operator (Y): It is represented as $r \rightarrow_{\mathrm{Y}} s$, and defined as: s^r.

(viii) Early Zadeh operator (EZ): It is represented as $r \rightarrow_{\mathrm{EZ}} s$. Early Zadeh operator is defined as: $(r \wedge s) \vee (1 - r)$.

However, these fuzzy implication operators are capable of handling 2D data only. A third dimension 'time' plays a crucial role in human action recognition to determine how human movement changes over time. Therefore, there is a need to define a new fuzzy implication operator that can handle space-time data.

Hence, in this work, a new space-time fuzzy implication operator is proposed, which can be efficiently employed in HMA domain. To this end, the popular fuzzy implication operators i.e. Łukasiewicz ($p \rightarrow_{\text{Ł}} q$) and Kleene-Dienes ($p \rightarrow_{\mathrm{KD}} q$) operators are modified to accommodate 'time' as an additional dimension, as follows:

$$p \rightarrow_{\text{newŁ}} q = min(1, 1 - p_t + q_t), \forall i, \forall t = 1, ..., l^i \tag{5.10}$$

$$p \rightarrow_{\text{newKD}} q = max(q_t, 1 - p_t), \forall i, \forall t = 1, ..., l^i \tag{5.11}$$

where $t = 1, \cdots, l^i$ taking partial action frame-by-frame, for the length of an action bounded by l_{min} and l_{max}. With these set of implication operators, each inference yields an interval in the range [0, 1]. The upper bound of an inference is given by Eq. 5.10, and the lower bound is given by Eq. 5.11. The implication operators must follow the constraint in Eq. 5.8 for reliable detection.

5.4 Study on Inference Structures

There exists a number of inference structures developed using operators such as ∧, ∨ and t-norm (Meng, 1997) that are employed in various applications. For example, the inference structures $K7$ and $K9$ delivered good performance for the medical expert system in C. K. Lim and Chan (2011). $K7$ and $K9$ are represented as:

$$K_7 : \ R \triangleleft_{K7} S(a,c) = \min\Big(\frac{1}{J}\sum_{b\in B}(R(a,b) \rightarrow S(b,c)), \mathrm{OrBot}(\mathrm{AndBot}(R(a,b), S(b,c)))\Big) \tag{5.12}$$

$$K_9 : \ R \triangleleft_{K9} S(a,c) = \min\Big(\frac{1}{J}\sum_{b\in B}(R(a,b) \rightarrow S(b,c)), \mathrm{OrBot}(\mathrm{AndTop}(R(a,b), S(b,c)))\Big) \tag{5.13}$$

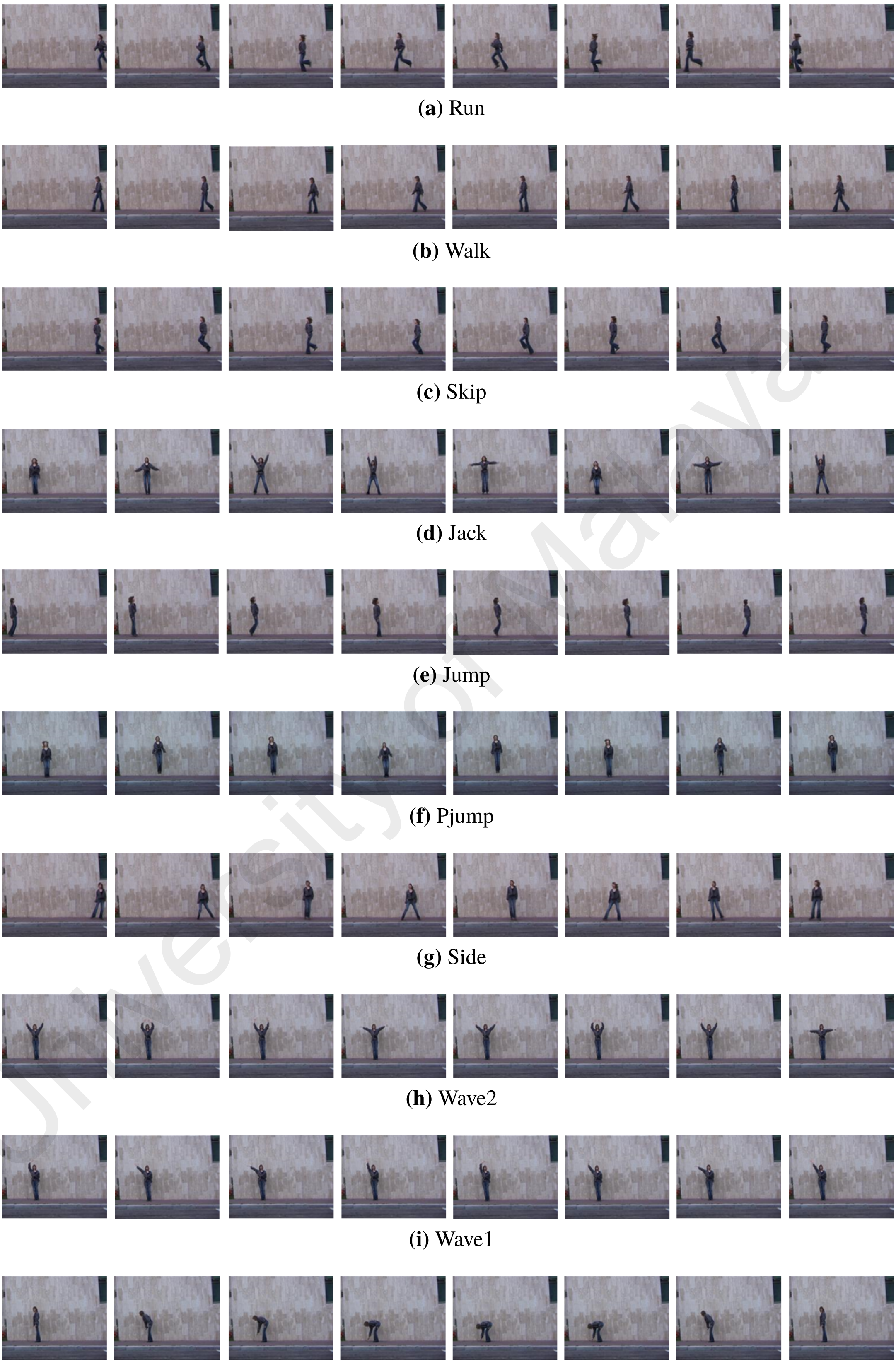
where $AndTop(p,q) = \min(p,q)$, $AndBot(p,q) = \max(0, p+q-1)$ and $OrBot(p,q) = \min(1, p+q)$ are the logical connectives. Furthermore, the inference structures instantiated from the original BK subproduct (Eq. 5.14) along with the combination of $K7$ and $K9$ were applied for scene classification in (Vats et al., 2012, 2015).

$$BK : \ R \triangleleft_{BK} S(a,c) = \frac{1}{J}\sum_{b\in B}(R(a,b) \rightarrow S(b,c)) \tag{5.14}$$

However, in order to find the suitable inference structure for human action recognition, the detector performance is tested using the classical inference structures: $K7$, $K9$ and original BK subproduct. The comparison results are shown in section 5.5.1.

5.5 Validation

In order to test the effectiveness of the proposed method, the experiments are performed on the Weizmann human actions dataset (Gorelick et al., 2007). As discussed in section 3.1.2,



(**a**) Run

(**b**) Walk

(**c**) Skip

(**d**) Jack

(**e**) Jump

(**f**) Pjump

(**g**) Side

(**h**) Wave2

(**i**) Wave1

(**j**) Bend

Figure 5.4: Example images from the Weizmann human actions dataset for ten action classes (Gorelick et al., 2007).

ten natural actions: ‘run’, ‘walk’, ‘skip’, ‘jack’ (jumping-jack), ‘jump’ (jump-forward-on-two-legs), ‘pjump’ (jump-in-place-on-two-legs), ‘side’ (gallop sideways), ‘wave2’ (wave-two-hands), ‘wave1’ (waveone-hand), and ‘bend’ are performed by nine different people. Figure 5.4 presents some example of image frames from the Weizmann human action dataset, representing ten action classes used in the experiments.

The following pseudo-code represents the implementation of hybrid technique for early human action detection in a step-by-step manner:

```
Step 1: Input an action video.
Step 2: Perform human detection.
Step 3: Segment the human body into three parts: head, torso+arm and leg.
Step 4: Perform feature extraction by constructing covariance matrix of features.
Step 5: Feature image is constructed. Save in Set A.
Step 6: Perform part-based covariance tracking.
Step 7: Human body part-based models are obtained. Save in Set B.
Step 8: Perform hybridization on the tracking output generated and Set B.
Step 9: Normalize the results obtained using min-max normalization. Save as membership function R.
Step 10: Obtain the converse relation S between actions and models by normalizing the tracking results using min-max normalization.
Step 11: Call BK subproduct inference engine frame-by-frame utilizing R and S.
Step 12: Set the threshold value as the cutting point for the detector.
Step 13: When membership degree exceeds the threshold value, stop.
Step 14: Output the frame number.
Step 15: Done.
```

Similar to section 3.1.2, the preprocessing of images, feature extraction and covari-

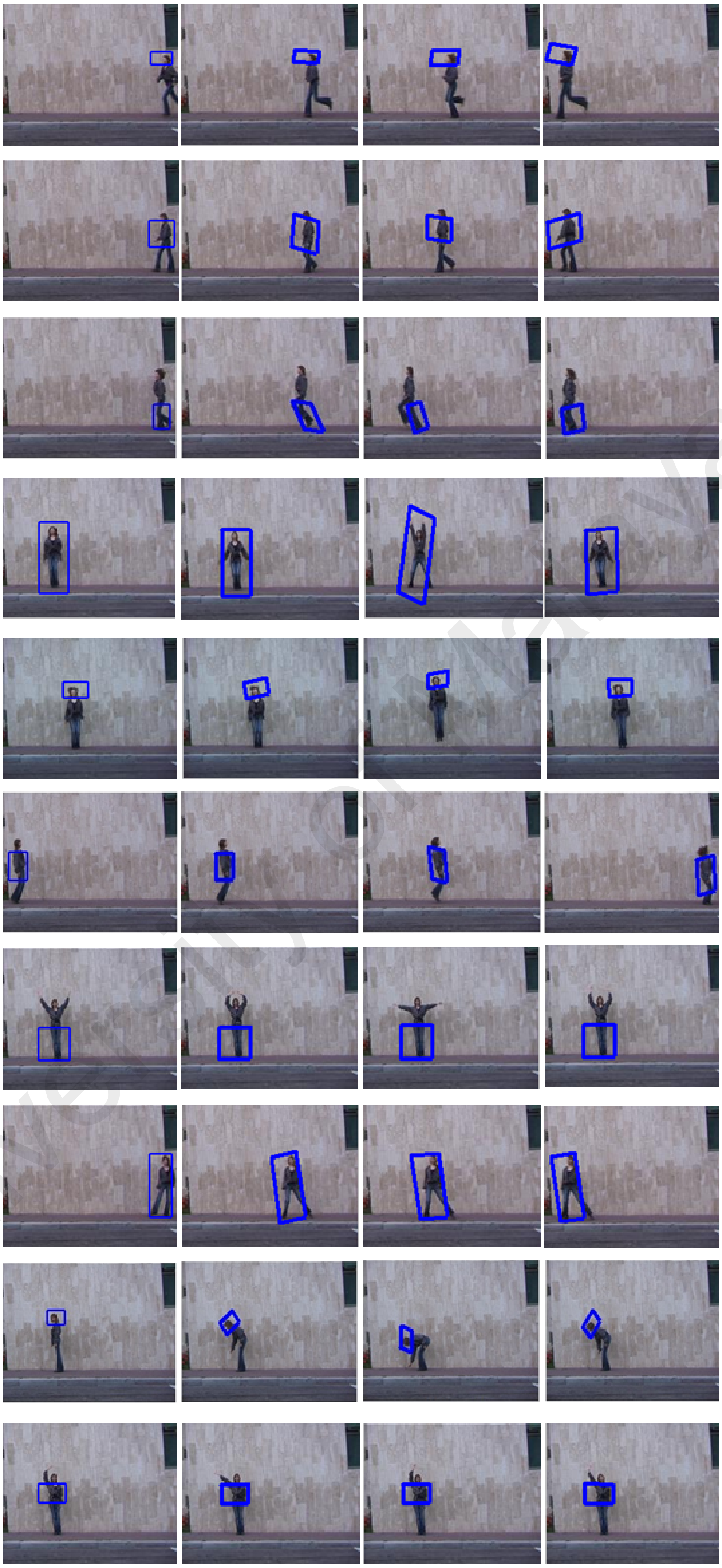
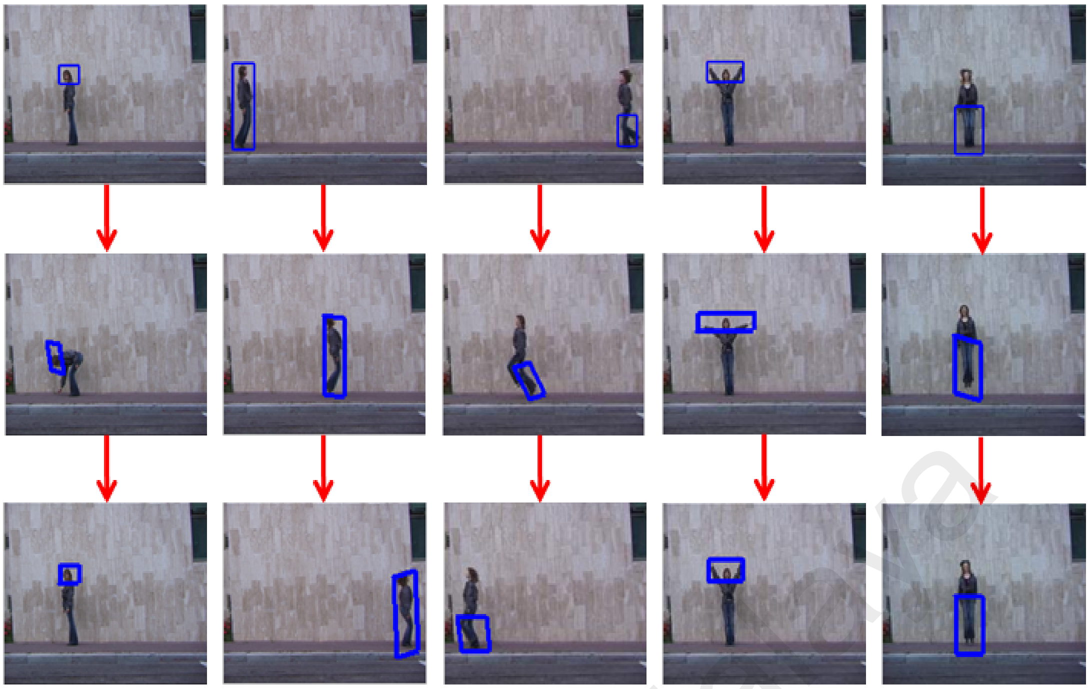


Figure 5.5: Sample human motion tracking results: From top to bottom row represents the part-based covariance tracking results for run, walk, skip, jack, pjump, jump, wave2, side, bend and wave1 action, represented using blue colored bounding box.



(a) m_1 for 'bend' **(b)** m_2 for 'jump' **(c)** m_3 for 'skip' **(d)** m_4 for 'wave2' **(e)** m_5 for 'pjump'

Figure 5.6: Part-based human body model generated from human motion tracking: m_1-m_5 for five example action sequences.

ance tracking are performed using the method in Porikli et al. (2006). And further the method is modified to generate the part-based human body model with separate tracks for full body, head, torso (arm included) and legs. Sample human motion tracking results for part-based covariance tracking for the ten action classes are shown in Figure 5.5.

Utilizing the results obtained from Figure 5.5, five models are constructed: m_1 - model the head movement from start to end frame, m_2 - model the position changes of the human body from the first frame, m_3 - model the distance between both the legs, m_4 - model the hand movement from start to end frame, and m_5 - model the distance of the human body from the ground. Figure 5.6 represents the model which forms the set B for BK relational product.

The membership function $R(f, m)$ is generated by normalizing the results obtained from the model-based covariance tracking using the fuzzy qualitative quantity states ($s = 10$). As can be seen in Table 5.1, $R(f, m)$ represents the one-to-many relationship

Table 5.1: Example of membership function $R(f, m)$, for models m_1-m_5.

Frame no.	m_1	m_2	m_3	m_4	m_5
1	1.00	0.00	0.00	0.00	0.00
10	1.00	0.10	0.00	0.10	0.20
20	0.70	0.60	0.00	0.30	0.70
30	0.20	0.70	0.00	0.90	0.30
40	0.00	0.30	0.00	0.90	0.00
50	0.40	0.40	0.00	0.50	0.30
60	1.00	0.10	0.00	0.10	1.00
70	1.00	0.10	0.00	0.10	0.00
80	1.00	0.10	0.00	0.20	0.10
90	0.50	0.10	0.80	0.00	0.00
100	0.40	0.30	0.80	0.00	0.00
110	0.40	0.40	0.80	0.00	0.80
120	0.50	0.60	0.60	0.00	0.80
130	0.20	0.80	0.30	0.00	0.50
140	1.00	0.90	0.10	0.00	0.60
150	0.50	1.00	0.20	0.00	0.30

Table 5.2: Example of membership function $S(m, a)$ for ten action classes.

Model	Bend	Jump	Jack	Skip	Pjump	Run	Side	Walk	Wave1	Wave2
m_1	0.60	0.80	0.01	0.80	0.11	0.80	0.70	0.80	0.00	0.00
m_2	0.01	0.80	0.12	0.90	0.20	0.90	0.85	0.90	0.00	0.00
m_3	0.01	0.10	0.82	0.25	0.01	0.85	0.75	0.88	0.00	0.00
m_4	0.70	0.01	0.90	0.15	0.18	0.60	0.65	0.60	0.50	0.90
m_5	0.01	0.25	0.20	0.20	0.30	0.20	0.01	0.01	0.00	0.00

between the images (set A) and the models (set B), describing the degree of belongingness between an image and several models. The membership function $S(m, a)$ represents the relationship between the model (set B) and the action (set C) being performed. Table 5.2 highlights the membership function values generated for the one-to-many relationship between model and action, with each model having a degree of belongingness to the action classes. Generating R and S for each image frame, fuzzy BK subproduct inference engine is invoked. Utilizing Eq. 5.9, 5.10 and 5.11, human action classification is performed. Since the partial human action is modeled instead of the complete action, the detector is

Table 5.3: Results for early human action detection using hybrid technique.

Action	Total no. of frames	t=0.70	t=0.75	t=0.80	t=0.85	t=0.90	Frames seen (%)
Bend	84	36	39	-	-	-	42.85
Jump	67	31	32	-	-	-	46.26
Jack	89	25	26	27	28	29	28.08
Skip	57	11	12	12	13	13	19.29
Pjump	62	12	13	15	17	32	19.35
Run	42	4	-	-	-	-	9.52
Side	53	4	5	5	-	-	7.54
Walk	84	8	9	-	-	-	9.52
Wave1	82	20	52	-	-	-	24.39
Wave2	81	13	13	14	14	-	16.04

capable of detecting an action early, before its completion.

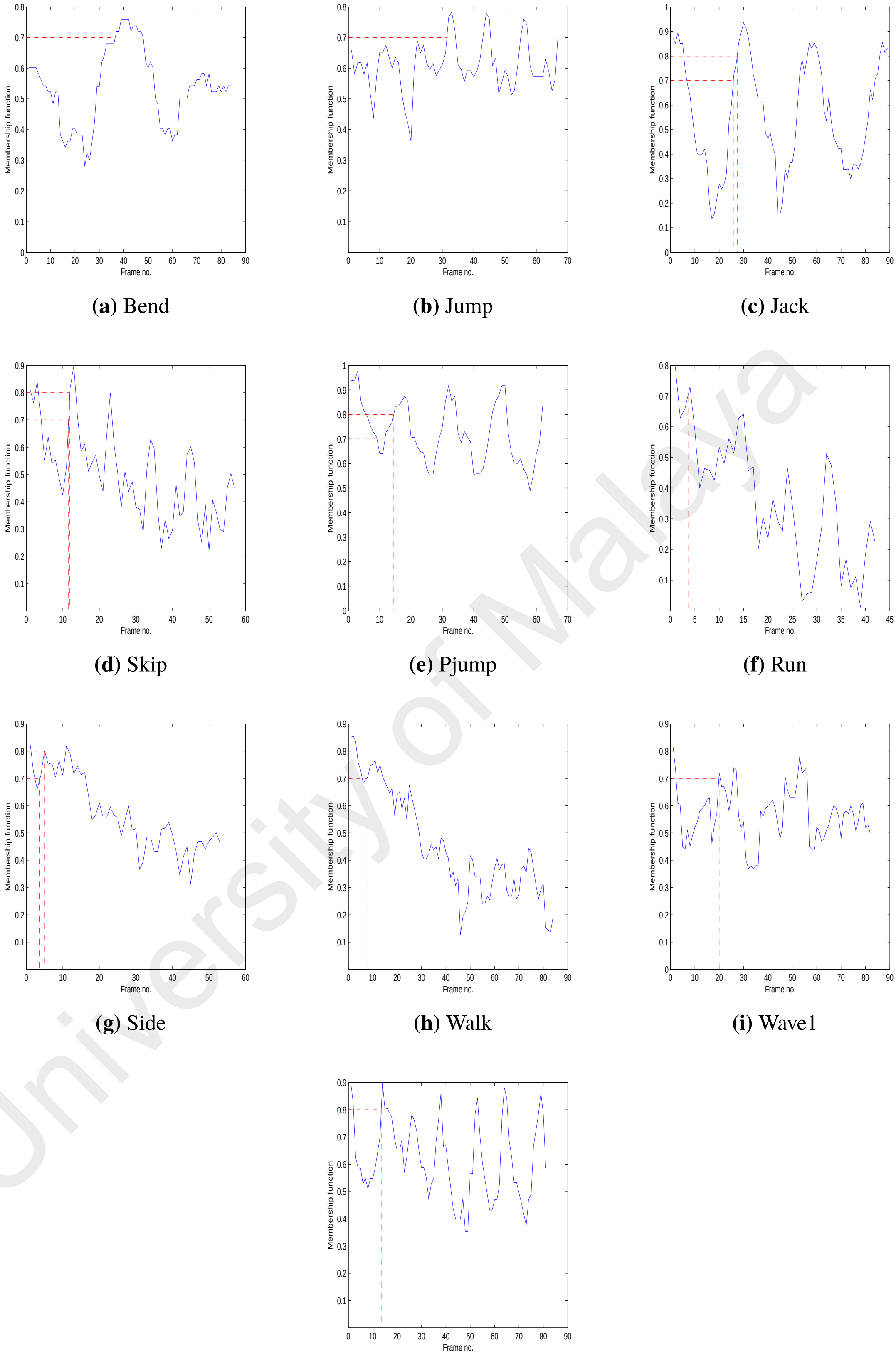
The proposed detector infers an action when the membership function value exceeds the pre-defined threshold monotonically. It is observed that automatic thresholding doesn't provide optimal solution for early detection. It is required to set a fixed threshold value for all the actions in order to detect an action early. In the experiments, results are tested using different threshold values i.e. 0.70, 0.75, 0.80, 0.85 and 0.90. Table 5.3 presents the early human action detection results obtained, where 't' refers to the threshold value and the last column represents the percentage of frames observed before the detector triggers the action when the threshold is set to 0.70.

Following observations are made on testing the detector performance for different threshold values:

(i) When the threshold is set to 0.70, the detector is able to detect all the actions performed upon seeing ~23% of the frames on an average.

(ii) On increasing the threshold to 0.75, the detector misses the detection for only 'run' action, and able to make early detection for all other actions upon seeing ~37% of the frames on an average.

(iii) With the threshold value set to 0.80, the detector successfully detects all actions



(a) Bend

(b) Jump

(c) Jack

(d) Skip

(e) Pjump

(f) Run

(g) Side

(h) Walk

(i) Wave1

(j) Wave2

Figure 5.7: Graphical results for early human action detection. The detector triggers the action upon seeing ~23% of the frames on an average when the membership function attains a certain threshold (e.g. 0.70 and 0.80 here, represented using red dotted lines) monotonically.

except ‘bend’, ‘jump’, ‘run’, ‘walk’ and ‘wave1’, upon seeing ~60% of the frames on an average.

(iv) Even with the threshold value 0.90, the detector is able to detect ‘jack’, ‘skip’ and ‘pjump’ action upon seeing ~33%, ~23%, and ~52% of the frames respectively.

(v) A threshold value lower than 0.70 is not recommended as then the detector will turn out to be unreliable.

Figure 5.7 highlights the experimental results qualitatively for the ten action classes, where the proposed early detector infers an action upon observing ~23% of the frames (on an average) when the membership function attains a certain threshold (e.g. 0.70 or 0.80) monotonically.

5.5.1 Comparison with the state-of-the-art

The conventional CV solutions for early human action detection includes Ryoo (2011); G. Yu et al. (2012); Ryoo et al. (2014); K. Li and Fu (2012); Hoai and De la Torre (2012, 2014). In terms of timeliness and accuracy of detection, MMED proposed in Hoai and De la Torre (2012, 2014) outperforms the other algorithms. The experiments were performed on the Auslan dataset (Australian Sign Language), the extended Cohn-Kanade dataset (CK+) and the Weizmann human actions dataset. On an average, MMED requires seeing ~37% of the sentence for Australian sign language recognition. To detect facial expression (CK+), MMED detects when it completes ~47% of the expression. For human action recognition using Weizmann human actions dataset, MMED requires seeing ~40% of the action (with a score of 0.7). In this work, the experiments were performed using the same human action dataset, and it is found that the detector significantly outperforms MMED where the detector requires seeing ~23% of the image frames on an average in an action video (with membership function score of 0.7). Nonetheless, the computational cost involved in MMED is high as it requires extensive labeling on each of the training samples. Due to the inherent advantages from fuzzy BK subproduct inference mechanism, the computational cost involved in the proposed early detector is lower as compared to MMED, i.e. $O(l)$, where l is length of the action. Moreover, MMED lacks in terms of handling the vague feature data and uncertainty involved in the training stage. The proposed method is based on fuzzy BK subproduct and therefore inherits the capabilities of fuzzy theory in handling the uncertainties involved therein using the fuzzy membership function values generated by invoking fuzzy BK subproduct inference engine.

However, there exists several methods that employ fuzzy logic for human action recognition. For example, FIS was successfully applied in Le Yaouanc and Poli (2012); Yao et al. (2014) for effectively distinguishing the human motion patterns using the flexible membership functions and the fuzzy rules with endurance to the vague feature data. In Gkalelis et al. (2008), FVQ incorporated with FCM was used to model the human movements with flexibility to support complex continuous actions. Despite of the inherent advantages of fuzzy logic in performing human action recognition, these approaches require seeing the complete action video to detect an action. Hence, these approaches lack in ability to detect an action early and cannot be quantitatively compared with the proposed methodology.

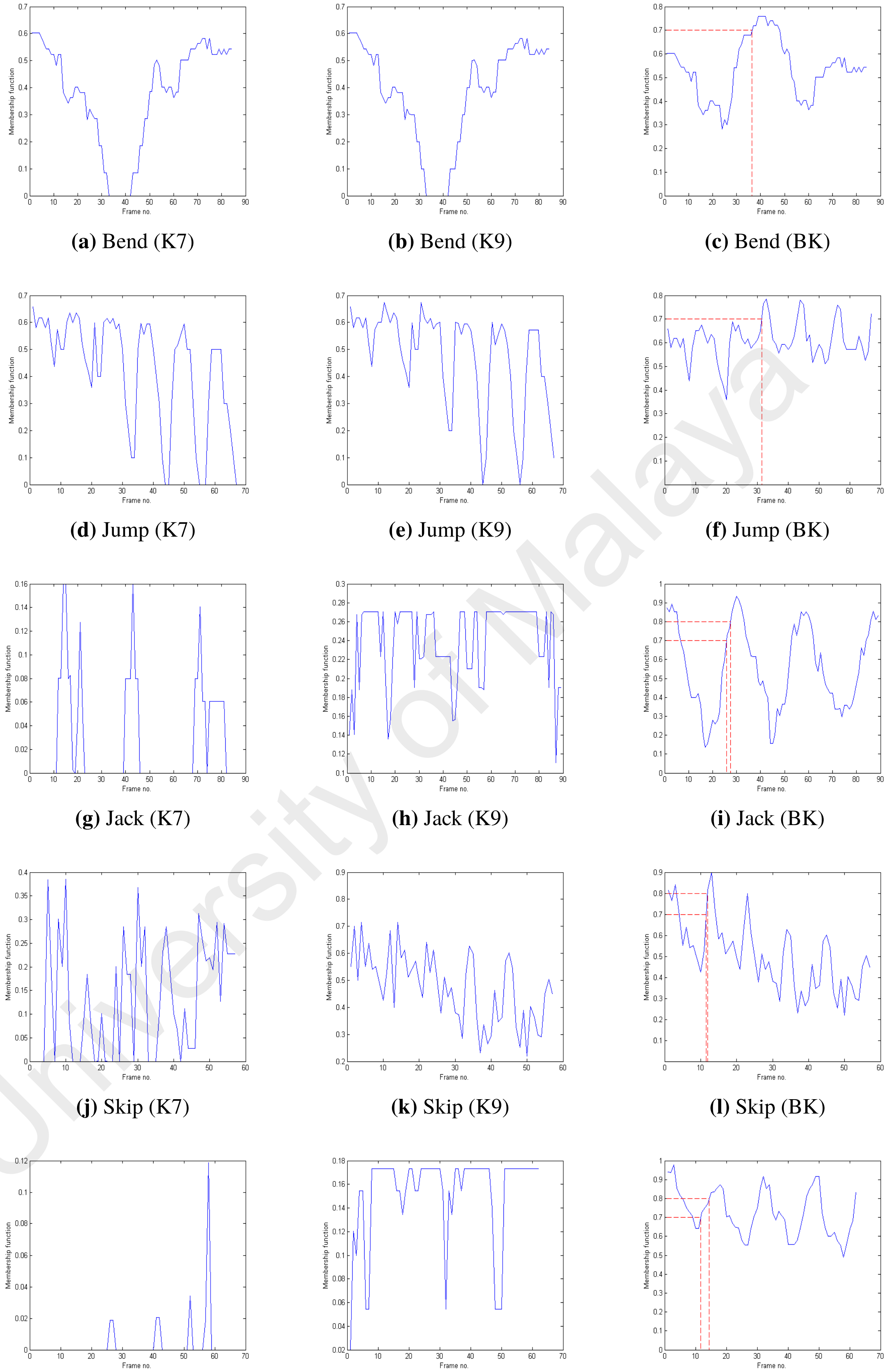
Recently, there has been a tremendous growth of research exploration of fusing elements of intelligence using efficient hybrid techniques. For example, Acampora et al. (2012); Hosseini and Eftekhari-Moghadam (2013) effectively integrated fuzzy logic with machine learning techniques for human action recognition where optimum membership function and flexible fuzzy rules were used to infer the human behavior. However, the conventional hybrid methods for human action recognition are not capable of inferring an action early. This work reveals the inherent strength of hybridization of computational

Table 5.4: Membership function values for inference structures.

Inference structure	Frame no.	Bend	Jump	Jack	Skip	Pjump	Run	Side	Walk	Wave1	Wave2
K7	1	0.60	0.66	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00
	10	0.52	0.50	0.00	0.39	0.00	0.42	0.00	0.00	0.50	0.30
	20	0.40	0.36	0.05	0.10	0.00	0.23	0.00	0.00	0.00	0.50
	30	0.19	0.50	0.00	0.37	0.00	0.16	0.03	0.03	0.00	0.30
	40	0.00	0.50	0.08	0.10	0.02	0.06	0.33	0.00	0.30	0.30
K9	1	0.60	0.66	0.14	0.55	0.02	0.61	0.29	0.13	0.00	0.00
	10	0.52	0.60	0.27	0.43	0.17	0.53	0.62	0.75	0.50	0.40
	20	0.40	0.36	0.27	0.50	0.17	0.23	0.55	0.64	0.00	0.60
	30	0.20	0.60	0.22	0.38	0.17	0.16	0.52	0.43	0.40	0.40
	40	0.00	0.57	0.22	0.29	0.17	0.18	0.49	0.43	0.50	0.40
Original BK	1	0.60	0.66	0.87	0.81	0.94	0.79	0.83	0.85	0.82	0.90
	10	0.52	0.65	0.47	0.43	0.64	0.53	0.71	0.75	0.52	0.55
	20	0.40	0.36	0.28	0.50	0.70	0.23	0.61	0.64	0.72	0.65
	30	0.54	0.61	0.94	0.38	0.75	0.16	0.52	0.43	0.54	0.59
	40	0.76	0.57	0.46	0.29	0.56	0.18	0.49	0.43	0.60	0.67

methods (CV solutions and fuzzy BK subproduct) for early human action detection in a way that the strength of fuzzy set theory can alleviate the limitation of CV solutions. To the best of the authors' knowledge, this is the first work in the community that employs hybrid technique for solving the problem of early human action detection and stands out against other conventional methods with good detection rate where the detector requires seeing only ~23% of the frames on an average to detect an action.

In order to justify the choice of employing fuzzy BK subproduct for HMA, the performance of detector is evaluated using the classical inference structures: $K7$, $K9$ and original BK (i.e. fuzzy BK subproduct). It is found that overall the original BK performed the best for all the action classes as shown in Table 5.4. Whereas, $K9$ delivered comparable results for some action sequences (e.g. bend, jump, skip, run and walk) and $K7$ performed fairly poorer for all the action classes. Figure 5.8 and 5.9 evaluates the results qualitatively for the ten example action classes. It can be observed from the graphical representation that the membership function values generated using $K7$ and $K9$ inference structures are much lower as compared to original BK. Therefore, it is deduced that original BK is the most suitable inference structure to be used to perform HMA.



(a) Bend (K7) **(b)** Bend (K9) **(c)** Bend (BK)

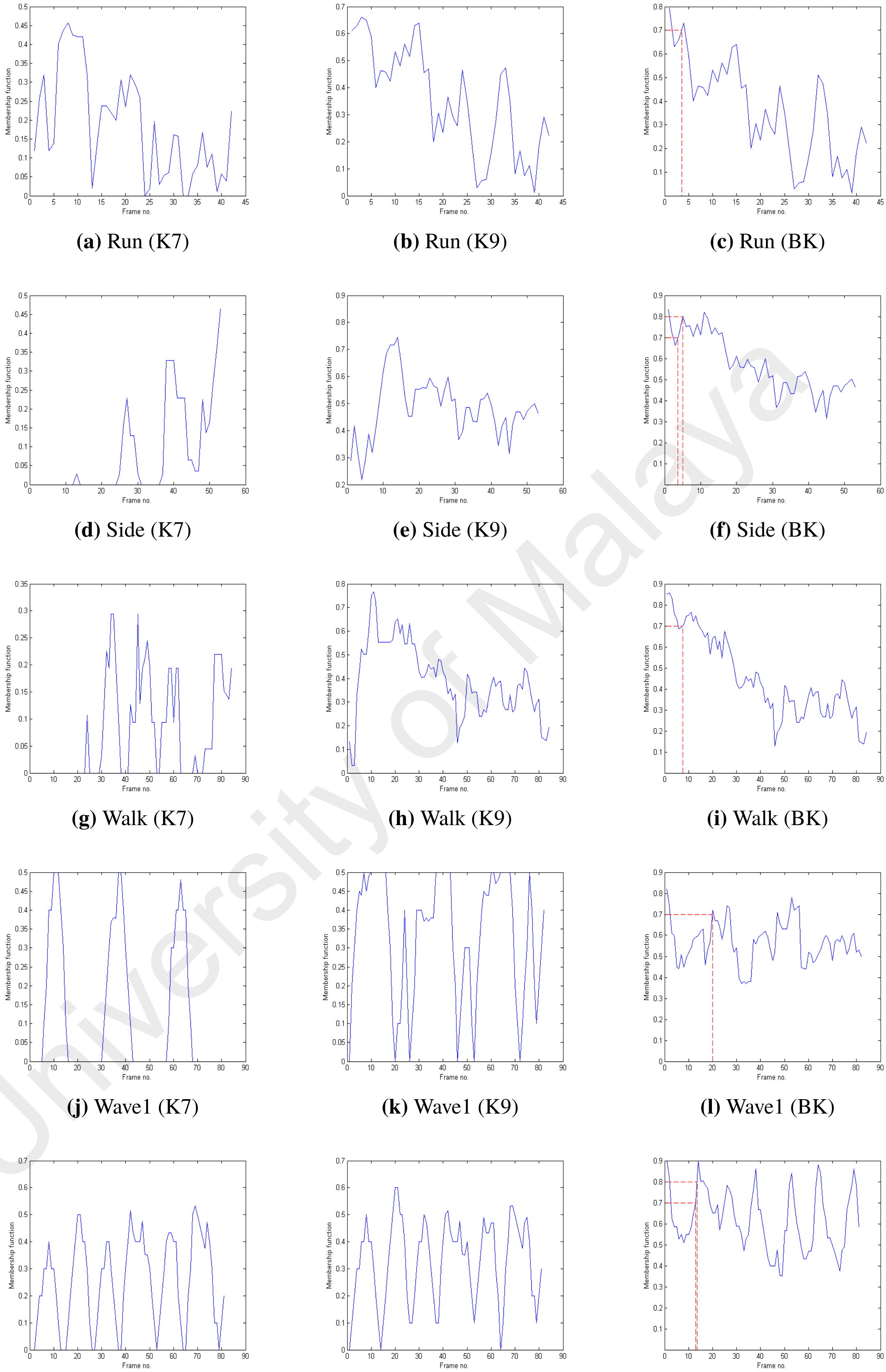
(d) Jump (K7) **(e)** Jump (K9) **(f)** Jump (BK)

(g) Jack (K7) **(h)** Jack (K9) **(i)** Jack (BK)

(j) Skip (K7) **(k)** Skip (K9) **(l)** Skip (BK)

(m) Pjump (K7) **(n)** Pjump (K9) **(o)** Pjump (BK)

Figure 5.8: Graphical results representing the early detector performance using $K7$, $K9$ and original BK inference structure (BK) for example actions: bend, jump, jack, skip and pjump.



(a) Run (K7) **(b)** Run (K9) **(c)** Run (BK)

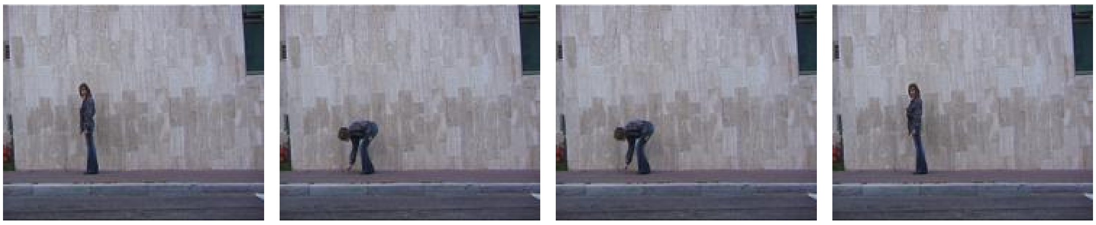
(d) Side (K7) **(e)** Side (K9) **(f)** Side (BK)

(g) Walk (K7) **(h)** Walk (K9) **(i)** Walk (BK)

(j) Wave1 (K7) **(k)** Wave1 (K9) **(l)** Wave1 (BK)

(m) Wave2 (K7) **(n)** Wave2 (K9) **(o)** Wave2 (BK)

Figure 5.9: Graphical results representing the early detector performance using $K7$, $K9$ and original BK inference structure (BK) for example actions: run, side, walk, wave1 and wave2.



(a) 0.00 **(b)** 0.42 **(c)** 0.46 **(d)** 1.00

Figure 5.10: NTtoD for *bend*. (a) Onset frame, (b) NTtoD with threshold 0.70 (the proposed early detector fires), (c) NTtoD with threshold 0.80, (d) Peak frame.

To evaluate the timeliness of detection, NTtoD is used. Assume for a given action sequence, where the action occurs from start frame s to end frame e, the detector triggers the action at time t. For successful detection, $s \leq t \leq e$, NTtoD is defined as $\frac{t-s+1}{e-s+1}$ i.e. the fraction of action occurred. When $t < s$, NTtoD = 0 i.e. false detection, and when $t > e$, NTtoD = ∞ i.e. false rejection. For the well-known classifiers (e.g. SVM, KNN), the classification is performed observing the complete action sequence and therefore NTtoD is 1. NTtoD for the detector in this work for the ten example actions is as follows: bend=0.42, jump=0.46, jack=0.28, skip=0.19, pjump=0.19, run=0.09, side=0.07, walk=0.09, wave1=0.24 and wave2=0.16. Figure 5.10 highlights the NTtoD results obtained using the detector for bend action.

5.6 Summary

This work takes the initiative to fuse the benefits of CV and fuzzy set theory to develop a hybrid technique to perform early human action detection. Human action classification problem is modified into frame-by-frame level classification where the partial human actions are modeled to enable early detection. The membership function values generated for each human action are utilized to infer an action. Detection is triggered when the membership function attains a pre-defined threshold monotonically.

To the best of my knowledge, this is the first work in the community that employs hybrid technique for solving the problem of early human action detection and stands out against other conventional methods with good detection rate. The experimental results demonstrate the capability of the proposed detector to carry out reliable early human action detection. On average, the detector is able to infer an action upon viewing ~23% of the frames.

CHAPTER 6: DISCUSSION AND CONCLUSION

This thesis presented a fuzzy approach for early human action detection and demonstrated its benefits in understanding the human behavior from human actions performed. The development of fuzzy BK subproduct based early human action detector was driven by the importance of detecting temporal human actions that belong to a certain pre-defined class, or the motions that are performed repeatedly. The framework was designed to model the complex human action and detect them as early as possible, after the action has begun but before it is completed. The experimental results demonstrate the capability of the proposed detector to carry out reliable early human action detection.

6.1 Summarized Contributions

6.1.1 Fuzzy BK subproduct as a classifier

This thesis addressed the most fundamental problem of selecting a classifier to employ for the classification task. As a solution, fuzzy BK subproduct was used as a classifier. In order to demonstrate the capability of fuzzy BK subproduct in handling both 3D video data and 2D image data, its performance was empirically tested for HMA and scene classification. Experimental results on standard public datasets demonstrated the effectiveness of fuzzy BK subproduct in performing HMA and scene classification. This was the first attempt of using fuzzy BK subproduct as a classifier.

6.1.2 Fuzzy approach for early human action detection

This thesis proposed a novel framework to detect human action early based on fuzzy BK subproduct inference mechanism by utilizing the fuzzy capabilities in handling the uncertainties that exist in the real-world for reliable decision making. Frame-by-frame action classification was performed for early detection where the fuzzy membership function generated from fuzzy BK subproduct provided the basis to detect an action before it is completed when a certain threshold is attained in a suitable way. In order to test the effectiveness of the proposed framework, a set of experiments was performed for few action sequences where the aim of the detector was to recognize an action upon seeing minimum number of frames possible.

Based on the best of my knowledge and a recent survey paper by C. H. Lim et al. (2015), this was the first work in the fuzzy community dealing with early human action detection. The closest research to this work was MMED proposed in Hoai and De la Torre (2012, 2014). For human action recognition using Weizmann human actions dataset, MMED required seeing ~40% of the action (with a score of 0.7). In this work, the experiments were performed using the same human action dataset, and it is found that the detector significantly outperforms MMED where the detector required seeing ~32% of the image frames on an average in an action video (with membership function score of 0.7).

6.1.3 Hybrid technique for early human action detection

The proposed framework was analyzed from a broader perspective where it can be represented as a hybrid model of CV and fuzzy set theory based on fuzzy BK subproduct. The proposed solution was the synergistic integration of CV solutions and fuzzy set theory where the relationship between a human subject and the action being performed was studied using fuzzy BK subproduct, efficiently integrated with CV techniques including feature extraction and motion tracking to perform human action recognition effectively. Hybrid techniques addressed issues such as uncertainty, vagueness or imprecision to a considerable extent by exploiting the strengths of fuzzy set theory to alleviate the limitations of CV solutions.

Human action classification problem was modified into frame-by-frame level classification where the partial human actions were modeled to enable early detection. The membership function values generated for each human action were utilized to infer an action. Detection was triggered when the membership function attained a pre-defined threshold monotonically. To the best of my knowledge, this is the first work in the community that employs hybrid technique for solving the problem of early human action detection and stands out against other conventional methods with good detection rate. The experimental results demonstrate the capability of the proposed detector to carry out reliable early human action detection. On average, the detector is able to infer an action upon viewing ~23% of the frames.

6.1.4 Fuzzy space-time implication operator

A study was performed on the impact of various fuzzy implication operators and the inference structures in retrieving the relationship between the human subject and the action. The existing fuzzy implication operators were capable of handling 2D data only. However, a third dimension 'time' plays a crucial role in human action recognition to model human movement changes over time. Therefore, a new space-time fuzzy implication operator was introduced, by modifying the existing implication operators to accommodate time as an added dimension.

Although the current framework is relevant to and effectual in performing early human detection, it has a few limitations. This chapter highlights the limitations of the current framework, and the future directions to improve and further extend it.

6.2 Limitations and Future Directions

6.2.1 Dataset biased

The current framework is designed for limited action classes in a testing dataset, and therefore is dataset biased (or in other words, dataset dependent). This is because the focus of this thesis is to validate and evaluate the performance of the proposed early detector, for which the experiments were performed on a well-known human action dataset. However, the lack of experiments on several complex datasets do not restrict the scope of this thesis and its applicability in various real-world applications. The proposed framework was tested for Weizmann human action dataset, and the early detector performs convincingly well. Hence, one of the future works can be the extension of the current framework to incorporate a large variety of complex datasets.

6.2.2 Detecting spatio-temporal events

In this thesis, temporal human actions were modeled for early detection. Localizing an event in time satisfies the goal of this thesis i.e. human action detection. However, it may not satisfy the applications where an event can occur at the same temporal locations but different spatial locations. Therefore, a possible future direction can be extending the current framework for detecting spatio-temporal events, localizing in both space and time.

6.2.3 Inter-segment dependency in action time series

Inter-segment dependency refers to the relationship amongst the segments of action time series. For example, “hand waving” is often followed by greeting (saying “good bye”), or a “hand shake” is followed by greeting (saying “hello”), or in a football match “kicking” a ball is often followed by “running”. The current framework ignores this inter-segment dependency. Therefore, it can be an interesting future study direction to extend the current framework to take into account inter-segment dependency in a series of actions.

6.2.4 Optimization

In the experiments, fuzzy BK subproduct inference mechanism worked well in detecting human action seeing minimum number of possible frames. However, the optimization of fuzzy BK subproduct for better initialization strategies can be investigated as a potential future work, and will be worth exploring.

6.2.5 Fuzzy datasets

In research, public datasets play a very important role in order to show the effectiveness of a proposed algorithm. Even so, from the findings in Table 6.1, there were not many works from the fuzzy community that had explored these public datasets. Only a handful of works in fuzzy HMA (as referred in Table 6.1) had employed those datasets and compared their works with other algorithms. In order to justify and improve the competency of the fuzzy approaches in HMA, it is believed that a way forward is to start employing these datasets for baseline comparisons.

On the other hand, the datasets listed in Table 6.1 undeniably have met the objectives for baseline evaluation. However, Boutell et al. (2004); Parikh and Grauman (2011); C. H. Lim and Chan (2012) raised an argument that many situations in the real life are ambiguous, especially the human behavior due to different perceptions of people. The current datasets, at this stage might be too ideal to reflect the real world scenarios, i.e. the current datasets are mutually exclusive, allowing a data to belong to one class (action) only at a time. Therefore, another potential area which can be explored as future work is having an appropriate psycho-physical dataset with fuzzy ground truths, or in simple words: fuzzy datasets. To the best of my knowledge, there do not exist any fuzzy datasets modeling the human activities and their behavior till date.

6.2.6 Fuzzy deep learning

Deep learning has created a research wave in the CV community with its outstanding performance in the recent years. Several real-time applications of deep learning include image recognition (Krizhevsky, Sutskever, & Hinton, 2012; Farabet, Couprie, Najman, & LeCun, 2013; Tompson, Jain, LeCun, & Bregler, 2014; Szegedy et al., 2014), speech recognition (Mikolov, Deoras, Povey, Burget, & Černockỳ, 2011; Hinton et al., 2012; Sainath, Mohamed, Kingsbury, & Ramabhadran, 2013) etc. A worth exploring problem

Table 6.1: The current best results of applying the fuzzy approaches and other stochastic methods on the well-known datasets in HMA. (RA indicates the recognition accuracy and TP is the tracking precision.)

Name	Dataset Established Year	Dataset Reference	Fuzzy paper that uses this dataset	Best accuracy in fuzzy approach(s) (%)	Best accuracy in other method(s) (%)
KTH	2004	Schuldt, Laptev, and Caputo (2004)	Chan and Liu (2009); Chan et al. (2008); Iosifidis et al. (2013)	RA = 93.52 Iosifidis et al. (2013)	RA = 96.76 Sapienza, Cuzzolin, and Torr (2014)
CAVIAR	2004	Fisher (2004)	-	-	TP = 91.90 Nie et al. (2014)
WEIZMANN *Actions*	2005	Blank, Gorelick, Shechtman, Irani, and Basri (2005)	Chan and Liu (2009); Chan et al. (2008); Gkalelis et al. (2008); Yao et al. (2014); Mozafari et al. (2012)	RA = 100.00 Chan and Liu (2009)	RA = 100.00 C.-C. Chen and Aggarwal (2009)
IXMAS	2006	Weinland et al. (2006)	C. H. Lim and Chan (2013); Iosifidis, Tefas, Nikolaidis, and Pitas (2012)	RA = 83.47 Iosifidis, Tefas, Nikolaidis, and Pitas (2012)	RA = 95.54 D. Wu and Shao (2014)

Table 6.1 (continued): The current best results of applying the fuzzy approaches and other stochastic methods on the well-known datasets in HMA. (RA indicates the recognition accuracy and TP is the tracking precision.)

Name	Dataset Established Year	Dataset Reference	Fuzzy paper that uses this dataset	Best accuracy in fuzzy approach(s) (%)	Best accuracy in other method(s) (%)
CASIA *Action*	2007	Y. Wang, Huang, and Tan (2007)	-	-	RA = 99.90 Lu, Boukharouba, Boonært, Fleury, and Lecœuche (2014)
ETISEO	2007	Nghiem, Bremond, Thonnat, and Valentin (2007)	-	-	TP = 100.00 Simha, Chau, and Bremond (2014)
UIUC - Complex action	2007	Ikizler and Duygulu (2007)	Chan et al. (2010)	RA > 80.00 Chan et al. (2010)	-
UIUC	2008	Tran and Sorokin (2008)	-	-	RA = 93.30 Tu, Xia, and Wang (2014)
CMU MoCap	2008	De la Torre, Hodgins, Montano, Valcarcel, and Macey (2009)	Gkalelis et al. (2008)	RA = 98.90 Gkalelis et al. (2008)	RA = 98.30 John and Trucco (2014)

Table 6.1 (continued): The current best results of applying the fuzzy approaches and other stochastic methods on the well-known datasets in HMA. (RA indicates the recognition accuracy and TP is the tracking precision.)

Name	Dataset Established Year	Dataset Reference	Fuzzy paper that uses this dataset	Best accuracy in fuzzy approach(s) (%)	Best accuracy in other method(s) (%)
ViHASi	2008	Ragheb, Velastin, Remagnino, and Ellis (2008)	-	-	RA = 72.00 L. Zhang et al. (2013)
HOLLYWOOD	2008	Laptev, Marszalek, Schmid, and Rozenfeld (2008)	-	-	RA = 61.50 Du, Zhai, Guo, Tang, and Lung (2014)
HOLLYWOOD-2	2009	Marszalek, Laptev, and Schmid (2009)	-	-	RA = 64.30 H. Wang and Schmid (2013)
UCF-Sports	2008	Rodriguez, Ahmed, and Shah (2008b)	Iosifidis et al. (2013)	RA = 85.77 Iosifidis et al. (2013)	RA = 89.70 S. Wu, Oreifej, and Shah (2011)
UCF-11 *Youtube*	2009	J. Liu, Luo, and Shah (2009)	-	-	RA = 89.79 Sapienza et al. (2014)
i3DPost	2009	Gkalelis, Kim, Hilton, Nikolaidis, and Pitas (2009)	Iosifidis, Tefas, Nikolaidis, and Pitas (2012); Iosifidis et al. (2013); Iosifidis, Tefas, and Pitas (2012b)	RA = 100.00 Iosifidis et al. (2013)	RA = 98.44 Holte, Chakraborty, Gonzalez, and Moeslund (2012)

Table 6.1 (continued): The current best results of applying the fuzzy approaches and other stochastic methods on the well-known datasets in HMA. (RA indicates the recognition accuracy and TP is the tracking precision.)

Name	Dataset Established Year	Dataset Reference	Fuzzy paper that uses this dataset	Best accuracy in fuzzy approach(s) (%)	Best accuracy in other method(s) (%)
UT-Interaction	2009	Ryoo and Aggarwal (2009)	-	-	RA = 91.67 Fu, Jia, and Kong (2014)
UT-Tower	2009	C.-C. Chen and Aggarwal (2009)	-	-	
MSR *Action*	2009	Yuan, Liu, and Wu (2009)	-	-	
MSR *3D Action*	2010	W. Li, Zhang, and Liu (2010)	-	-	RA = 97.80 Yang and Tian (2014)
BEHAVE	2010	Blunsden and Fisher (2010)	-	-	RA = 65.50 Z. Cheng, Qin, Huang, Yan, and Tian (2014)
MuHAVi	2010	Singh, Velastin, and Ragheb (2010)	-	-	RA = 100.00 Chaaraoui and Flórez-Revuelta (2014)
Olympic Sports	2010	Niebles, Chen, and Fei-Fei (2010)	-	-	RA = 91.10 H. Wang and Schmid (2013)

Table 6.1 (continued): The current best results of applying the fuzzy approaches and other stochastic methods on the well-known datasets in HMA. (RA indicates the recognition accuracy and TP is the tracking precision.)

Name	Dataset Established Year	Dataset Reference	Fuzzy paper that uses this dataset	Best accuracy in fuzzy approach(s) (%)	Best accuracy in other method(s) (%)
TV Human Interaction	2010	Patron-Perez, Marszalek, Zisserman, and Reid (2010)	-	-	RA = 46.00 Marín-Jiménez, Muñoz-Salinas, Yeguas-Bolivar, and de la Blanca (2014)
HMDB51	2011	Kuehne, Jhuang, Garrote, Poggio, and Serre (2011)	-	-	RA = 57.20 H. Wang and Schmid (2013)
VideoWeb	2011	Denina et al. (2011)	-	-	RA = 72.00 Zha, Zhang, Wang, Luan, and Chua (2013)
UCF-101	2012	Soomro, Zamir, and Shah (2012)	-	-	RA = 83.50 Cai, Wang, Peng, and Qiao (2014)
UCF-50	2013	Reddy and Shah (2013)	-	-	RA = 91.20 H. Wang and Schmid (2013)

can be designing “fuzzy deep learning” model which can be applied in several applications. In view of the encouraging results obtained in this work, hybridization of deep learning and fuzzy set theory for human action recognition can be a potential future work.

6.3 Conclusion

This thesis presented a fuzzy BK subproduct based approach for detecting human actions early and utilized the benefits of both CV and fuzzy set theory. The conventional human action classification problem was modified into frame-by-frame level classification where the partial human actions were modeled to enable early detection. The membership function values generated for each human action from fuzzy BK subproduct inference engine were utilized to infer an action. The detection is triggered when the membership function attains a pre-defined threshold monotonically. The experimental results demonstrated the capability of the proposed detector to carry out reliable early human action detection. On an average, the detector was able to infer an action upon viewing ~23% of the frames for test data under the experimental settings. It is worth mentioning that the proposed framework not only benefits the HMA applications, but also can be applied to several other research domains.

REFERENCES

Acampora, G., Foggia, P., Saggese, A., & Vento, M. (2012). Combining neural networks and fuzzy systems for human behavior understanding. In *Ieee ninth international conference on advanced video and signal-based surveillance* (pp. 88–93).

Aggarwal, J. K., & Cai, Q. (1997). Human motion analysis: A review. In *Ieee nonrigid and articulated motion workshop* (pp. 90–102).

Aggarwal, J. K., Cai, Q., Liao, W., & Sabata, B. (1994). Articulated and elastic non-rigid motion: A review. In *Ieee workshop on motion of non-rigid and articulated objects* (pp. 2–14).

Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, *43*(3), 16.

Al-Jarrah, O., & Halawani, A. (2001). Recognition of gestures in arabic sign language using neuro-fuzzy systems. *Artificial Intelligence*, *133*(1), 117–138.

Anderson, D. T., Keller, J. M., Anderson, M., & Wescott, D. J. (2011). Linguistic description of adult skeletal age-at-death estimations from fuzzy integral acquired fuzzy sets. In *Ieee international conference on fuzzy systems* (pp. 2274–2281).

Anderson, D. T., Keller, J. M., Skubic, M., Chen, X., & He, Z. (2006). Recognizing falls from silhouettes. In *International conference of the ieee engineering in medicine and biology society* (pp. 6388–6391).

Anderson, D. T., Luke, R., Skubic, M., Keller, J. M., Rantz, M., & Aud, M. (2008). Evaluation of a video based fall recognition system for elders using voxel space. *Gerontechnology*, *7*(2), 68.

Anderson, D. T., Luke, R. H., Keller, J. M., & Skubic, M. (2008). Extension of a soft-computing framework for activity analysis from linguistic summarizations of video. In *Ieee international conference on fuzzy systems* (pp. 1404–1410).

Anderson, D. T., Luke, R. H., Keller, J. M., Skubic, M., Rantz, M. J., & Aud, M. A. (2009a). Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Computer Vision and Image Understanding*, *113*(1), 80–89.

Anderson, D. T., Luke, R. H., Keller, J. M., Skubic, M., Rantz, M. J., & Aud, M. A. (2009b). Modeling human activity from voxel person using fuzzy logic. *IEEE Transactions on Fuzzy Systems*, *17*(1), 39–49.

Anderson, D. T., Luke III, R. H., Stone, E. E., & Keller, J. M. (2009). Fuzzy voxel object. In *World congress of the international fuzzy systems association / conference of the european society for fuzzy logic and technology* (pp. 282–287).

Balcilar, M., & Sonmez, A. C. (2013). Region based fuzzy background subtraction using choquet integral. In *Adaptive and natural computing algorithms* (pp. 287–296). Springer.

Bandler, W., & Kohout, L. (1980a). Fuzzy power sets and fuzzy implication operators. *Fuzzy Sets and Systems*, *4*(1), 13–30.

Bandler, W., & Kohout, L. J. (1980b). Semantics of implication operators and fuzzy relational products. *International Journal of Man-Machine Studies*, *12*(1), 89–116.

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics*, *23*(2), 145–152.

Barrenechea, E., Bustince, H., Fernandez, J., Paternain, D., & Sanz, J. A. (2013). Using the choquet integral in the fuzzy reasoning method of fuzzy rule-based classification systems. *Axioms*, *2*(2), 208–223.

Bělíček, T., Kidéry, J., Kukal, J., Matěj, R., & Rusina, R. (2013). Morphological analysis of 3d spect images via nilpotent t-norms in diagnosis of alzheimer's disease. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, *24*(2), 313–321.

Bezdek, J. C. (1992). Computing with uncertainty. *IEEE Communications Magazine*, *30*(9), 24–36.

Bhattacharyya, S., & Maulik, U. (2013). Target tracking using fuzzy hostility induced segmentation of optical flow field. In *Soft computing for image and multimedia data processing* (pp. 97–107). Springer.

Bhattacharyya, S., Maulik, U., & Dutta, P. (2009). High-speed target tracking by fuzzy hostility-induced segmentation of optical flow field. *Applied Soft Computing*, *9*(1), 126–134.

Binh, N. D., & Ejima, T. (2005). Hand gesture recognition using fuzzy neural network. In *Conference on graphics, vision and image proces* (pp. 1–6).

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73.

Blank, M., Gorelick, L., Shechtman, E., Irani, M., & Basri, R. (2005). Actions as space-time shapes. In *Ieee international conference on computer vision* (Vol. 2, pp. 1395–1402).

Blunsden, S., & Fisher, R. (2010). The behave video dataset: ground truthed video for multi-person behavior classification. *Annals of the British Machine Vision Association*, *2010*(4), 1–12.

Bobick, A. F. (1997). Movement, activity and action: the role of knowledge in the perception of motion. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1358), 1257–1265.

Bobick, A. F., & Davis, J. W. (2001). The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(3), 257–267.

Bosch, A., Zisserman, A., & Muñoz, X. (2006). Scene classification via plsa. In *European conference on computer vision* (pp. 517–530). Springer.

Boutell, M. R., Luo, J., Shen, X., & Brown, C. M. (2004). Learning multi-label scene classification. *Pattern recognition*, *37*(9), 1757–1771.

Bouwmans, T., & El Baf, F. (2009). Modeling of dynamic backgrounds by type-2 fuzzy gaussians mixture models. *MASAUM Journal of of Basic and Applied Sciences*, *1*(2), 265–276.

Bui, L.-D., & Kim, Y.-G. (2006). An obstacle-avoidance technique for autonomous underwater vehicles based on bk-products of fuzzy relation. *Fuzzy Sets and Systems*, *157*(4), 560–577.

Bustince, H., Burillo, P., & Soria, F. (2003). Automorphisms, negations and implication operators. *Fuzzy Sets and Systems*, *134*(2), 209–229.

Cai, Z., Wang, L., Peng, X., & Qiao, Y. (2014). Multi-view super vector for action recognition. In *Ieee international conference on computer vision and pattern recognition* (pp. 596–603).

Calvo-Gallego, E., Brox, P., & Sánchez-Solano, S. (2013). A fuzzy system for background modeling in video sequences. In *Fuzzy logic and applications* (pp. 184–192). Springer.

Candamo, J., Shreve, M., Goldgof, D. B., Sapper, D. B., & Kasturi, R. (2010). Understanding transit scenes: A survey on human behavior-recognition algorithms. *IEEE Transactions on Intelligent Transportation Systems*, *11*(1), 206–224.

Cao, Y., Barrett, D., Barbu, A., Narayanaswamy, S., Yu, H., Michaux, A., ... Wang, S. (2013). Recognize human activities from partially observed videos. In *Ieee conference on computer vision and pattern recognition* (pp. 2658–2665).

Cédras, C., & Shah, M. (1995). Motion-based recognition a survey. *Image and vision computing*, *13*(2), 129–155.

Chaaraoui, A. A., & Flórez-Revuelta, F. (2014). Optimizing human action recognition based on a cooperative coevolutionary algorithm. *Engineering Applications of Artificial Intelligence*, *31*, 116–125.

Chaira, T. (2012). Intuitionistic fuzzy color clustering of human cell images on different color models. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, *23*(2, 3), 43–51.

Chan, C. S., & Liu, H. (2009). Fuzzy qualitative human motion analysis. *IEEE Transactions on Fuzzy Systems*, *17*(4), 851–862.

Chan, C. S., Liu, H., Brown, D. J., & Kubota, N. (2008). A fuzzy qualitative approach to human motion recognition. In *Ieee international conference on fuzzy systems* (pp. 1242–1249).

Chan, C. S., Liu, H., & Lai, W. K. (2010). Fuzzy qualitative complex actions recognition. In *Ieee international conference on fuzzy systems* (pp. 1–8).

Chaquet, J. M., Carmona, E. J., & Fernández-Caballero, A. (2013). A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding*, *117*(6), 633–659.

Chen, C.-C., & Aggarwal, J. K. (2009). Recognizing human action from a far field of view. In *Workshop on motion and video computing* (pp. 1–7).

Chen, G., Xie, Q., & Shieh, L. S. (1998). Fuzzy kalman filtering. *Information Sciences*, *109*(1), 197–209.

Chen, L., Wei, H., & Ferryman, J. (2013). A survey of human motion analysis using depth imagery. *Pattern Recognition Letters*, *34*(15), 1995–2006.

Chen, X., He, Z., Anderson, D. T., Keller, J. M., & Skubic, M. (2006). Adaptive silouette extraction and human tracking in complex and dynamic environments. In *Ieee international conference on image processing* (pp. 561–564).

Chen, X., He, Z., Keller, J. M., Anderson, D. T., & Skubic, M. (2006). Adaptive silhouette extraction in dynamic environments using fuzzy logic. In *Ieee international conference on fuzzy systems* (pp. 236–243).

Cheng, T. W., Goldgof, D., & Hall, L. (1995). Fast clustering with application to fuzzy rule generation. In *Ieee international conference on fuzzy systems* (Vol. 4, pp. 2289–2295).

Cheng, Z., Qin, L., Huang, Q., Yan, S., & Tian, Q. (2014). Recognizing human group action by layered model with multiple cues. *Neurocomputing*, *136*, 124–135.

Chowdhury, A., & Tripathy, S. S. (2014). Detection of human presence in a surveillance video using fuzzy approach. In *International conference on signal processing and integrated networks* (pp. 216–219).

Cordón, O., Herrera, F., & Villar, P. (2001). Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *Transactions on Fuzzy Systems*, *9*(4), 667–674.

Cristani, M., Raghavendra, R., Del Bue, A., & Murino, V. (2013). Human behavior analysis in video surveillance: A social signal processing perspective. *Neurocomputing*, *100*, 86–97.

Dawn, D. D., & Shaikh, S. H. (2015). A comprehensive survey of human action recognition with spatio-temporal interest point (stip) detector. *The Visual Computer*, 1–18.

De Baets, B., & Kerre, E. (1993). Fuzzy relational compositions. *Fuzzy Sets and Systems*, *60*(1), 109–120.

DeKruger, D., Hodge, J., Bezdek, J. C., Keller, J. M., & Gader, P. (2001). Detecting mobile land targets in ladar imagery with fuzzy algorithms. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, *10*(3, 4), 197–213.

De la Torre, F., Hodgins, J., Montano, J., Valcarcel, S., & Macey, J. (2009). Guide to the carnegie mellon university multimodal activity (cmu-mmac) database. *Robotics Institute, Carnegie Mellon University*.

Denina, G., Bhanu, B., Nguyen, H. T., Ding, C., Kamal, A., Ravishankar, C., . . . Varda, B. (2011). Videoweb dataset for multi-camera activities and non-verbal communication. In *Distributed video sensor networks* (pp. 335–347). Springer.

Du, J.-X., Zhai, C.-M., Guo, Y.-L., Tang, Y.-Y., & Lung, P. C. C. (2014). Recognizing complex events in real movies by combining audio and video features. *Neurocomputing*, *137*, 89–95.

El Baf, F., Bouwmans, T., & Vachon, B. (2008a). A fuzzy approach for background subtraction. In *Ieee international conference on image processing* (pp. 2648–2651).

El Baf, F., Bouwmans, T., & Vachon, B. (2008b). Fuzzy integral for moving object detection. In *Ieee international conference on fuzzy systems* (pp. 1729–1736).

El Baf, F., Bouwmans, T., & Vachon, B. (2008c). Type-2 fuzzy mixture of gaussians model: application to background modeling. In *Advances in visual computing* (pp. 772–781). Springer.

El Baf, F., Bouwmans, T., & Vachon, B. (2009). Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos. In *Ieee computer society conference on computer vision and pattern recognition workshop* (pp. 60–65).

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine*

Intelligence, *35*(8), 1915–1929.

Fei-Fei, L., & Perona, P. (2005). A bayesian hierarchical model for learning natural scene categories. In *Ieee international conference on computer vision and pattern recognition* (Vol. 2, pp. 524–531).

Fisher, R. B. (2004). The pets04 surveillance ground-truth data sets. In *Ieee international workshop on performance evaluation of tracking and surveillance* (pp. 1–5).

Fu, Y., Jia, Y., & Kong, Y. (2014). Interactive phrases: Semantic descriptions for human interaction recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1.

García, J., Molina, J. M., Besada, J. A., Portillo, J. I., & Casar, J. R. (2002). Robust object tracking with fuzzy shape estimation. In *International conference on information fusion* (Vol. 1, pp. 64–71).

Garcia, J., Patricio, M. A., Berlanga, A., & Molina, J. M. (2011). Fuzzy region assignment for visual tracking. *Soft Computing*, *15*(9), 1845–1864.

Gavrila, D. M. (1999). The visual analysis of human movement: A survey. *Computer vision and image understanding*, *73*(1), 82–98.

Gkalelis, N., Kim, H., Hilton, A., Nikolaidis, N., & Pitas, I. (2009). The i3dpost multi-view and 3d human action/interaction database. In *Conference for visual media production* (pp. 159–168).

Gkalelis, N., Tefas, A., & Pitas, I. (2008). Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *18*(11), 1511–1521.

Gorelick, L., Blank, M., Shechtman, E., Irani, M., & Basri, R. (2007). Actions as space-time shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(12), 2247–2253.

Gowsikhaa, D., Abirami, S., & Baskaran, R. (2014). Automated human behavior analysis from surveillance videos: a survey. *Artificial Intelligence Review*, *42*(4), 747–765.

Groenemans, R., Van Ranst, E., & Kerre, E. (1997). Fuzzy relational calculus in land evaluation. *Geoderma*, *77*(2), 283–298.

Guo, G., & Lai, A. (2014). A survey on still image based human action recognition. *Pattern Recognition*, *47*(10), 3343–3361.

Guo, Y., Xu, G., & Tsuji, S. (1994). Tracking human body motion based on a stick figure model. *Journal of Visual Communication and Image Representation*, *5*(1), 1–9.

Haering, N., Venetianer, P. L., & Lipton, A. (2008). The evolution of video surveillance: an overview. *Machine Vision and Applications*, *19*(5-6), 279–290.

Hatakeyama, Y., Mitsuta, A., & Hirota, K. (2008). Detection algorithm for color dynamic images by multiple surveillance cameras under low luminance conditions based on fuzzy corresponding map. *Applied Soft Computing*, *8*(4), 1344 - 1353.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., . . . Sainath, T. N. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*(6), 82–97.

Hoai, M., & De la Torre, F. (2012). Max-margin early event detectors. In *Ieee conference on computer vision and pattern recognition* (pp. 2863–2870).

Hoai, M., & De la Torre, F. (2014). Max-margin early event detectors. *International Journal of Computer Vision*, *107*(2), 191–202.

Holte, M. B., Chakraborty, B., Gonzalez, J., & Moeslund, T. B. (2012). A local 3-d motion descriptor for multi-view human action recognition from 4-d spatio-temporal interest points. *IEEE Journal of Selected Topics in Signal Processing*, *6*(5), 553–565.

Holte, M. B., Tran, C., Trivedi, M. M., & Moeslund, T. B. (2011). Human action recognition using multiple views: a comparative perspective on recent developments. In *Proceedings of the joint acm workshop on human gesture and behavior understanding* (pp. 47–52).

Hosseini, M.-S., & Eftekhari-Moghadam, A.-M. (2013). Fuzzy rule-based reasoning approach for event detection and annotation of broadcast soccer video. *Applied Soft Computing*, *13*(2), 846–866.

Hu, W., Tan, T., Wang, L., & Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *34*(3), 334–352.

Hu, W., Xie, D., Tan, T., & Maybank, S. (2004). Learning activity patterns using fuzzy self-organizing neural network. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *34*(3), 1618–1626.

Huntsberger, T. L., Rangarajan, C., & Jayaramamurthy, S. N. (1986). Representation of uncertainty in computer vision using fuzzy sets. *IEEE Transactions on Computers*, *100*(2), 145–156.

Ikizler, N., & Duygulu, P. (2007). Human action recognition using distribution of oriented rectangular patches. In *Human motion–understanding, modeling, capture and animation* (pp. 271–284). Springer.

Iosifidis, A., Tefas, A., Nikolaidis, N., & Pitas, I. (2012). Multi-view human movement recognition based on fuzzy distances and linear discriminant analysis. *Computer Vision and Image Understanding*, *116*(3), 347–360.

Iosifidis, A., Tefas, A., & Pitas, I. (2011). Person specific activity recognition using fuzzy learning and discriminant analysis. In *European signal processing conference* (pp. 1974–1978).

Iosifidis, A., Tefas, A., & Pitas, I. (2012a). Activity-based person identification using fuzzy representation and discriminant learning. *IEEE Transactions on Information Forensics and Security*, *7*(2), 530–542.

Iosifidis, A., Tefas, A., & Pitas, I. (2012b). Multi-view action recognition based on action volumes, fuzzy distances and cluster discriminant analysis. *Signal Processing*.

Iosifidis, A., Tefas, A., & Pitas, I. (2013). Minimum class variance extreme learning machine for human action recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *23*(11), 1968–1979.

Isard, M., & Blake, A. (1998). Condensation—conditional density propagation for visual tracking. *International journal of computer vision*, *29*(1), 5–28.

Iwai, Y., Ogaki, K., & Yachida, M. (1999). Posture estimation using structure and motion models. In *Ieee international conference on computer vision* (Vol. 1, pp. 214–219).

Ji, X., & Liu, H. (2010). Advances in view-invariant human motion analysis: a review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *40*(1), 13–24.

John, V., & Trucco, E. (2014). Charting-based subspace learning for video-based human action classification. *Machine Vision and Applications*, *25*(1), 119–132.

Ju, S. X., Black, M. J., & Yacoob, Y. (1996). Cardboard people: A parameterized model of articulated image motion. In *International conference on automatic face and gesture recognition* (pp. 38–44).

Juang, C.-F., & Chang, C.-M. (2007). Human body posture classification by a neural fuzzy network and home care system application. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *37*(6), 984–994.

Kacprzyk, J., & Yager, R. R. (2001). Linguistic summaries of data using fuzzy logic. *International Journal of General System*, *30*(2), 133–154.

Kakadiaris, I. A., & Metaxas, D. (1996). Model-based estimation of 3d human motion with occlusion based on active multi-viewpoint selection. In *Ieee conference on computer vision and pattern recognition* (pp. 81–87).

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, *82*(1), 35–45.

Kamel, H., & Badawy, W. (2005). Fuzzy logic based particle filter for tracking a maneuverable target. In *Symposium on circuits and systems* (pp. 1537–1540).

Kim, I. S., Choi, H. S., Yi, K. M., Choi, J. Y., & Kong, S. G. (2010). Intelligent visual surveillance – a survey. *International Journal of Control, Automation and Systems*, *8*(5), 926–939.

Kim, Y.-J., Won, C.-H., Pak, J.-M., & Lim, M.-T. (2007). Fuzzy adaptive particle filter for localization of a mobile robot. In *Knowledge-based intelligent information and engineering systems* (pp. 41–48).

Kirtley, C., & Smith, R. (2001). Application of multimedia to the study of human movement. *Multimedia tools and Applications*, *14*(3), 259–268.

Kitani, K. M., Ziebart, B. D., Bagnell, J. A., & Hebert, M. (2012). Activity forecasting. In *European conference on computer vision* (pp. 201–214). Springer.

Ko, T. (2008). A survey on behavior analysis in video surveillance for homeland security applications. In *Ieee applied imagery pattern recognition workshop* (pp. 1–8).

Kobayashi, K., Cheok, K. C., Watanabe, K., & Munekata, F. (1998). Accurate differential global positioning system via fuzzy logic kalman filter sensor fusion technique. *IEEE Transactions on Industrial Electronics*, *45*(3), 510–518.

Kohout, L. J., & Bandler, W. (1985). Relational-product architectures for information processing. *Information Sciences*, *37*(1), 25–37.

Kohout, L. J., & Bandler, W. (1992). How the checklist paradigm elucidates the semantics of fuzzy inference. In *Ieee international conference on fuzzy systems* (pp. 571–578).

Kong, Y., Kit, D., & Fu, Y. (2014). A discriminative model with multiple temporal scales for action prediction. In *European conference on computer vision* (pp. 596–611). Springer.

Kratz, L., & Nishino, K. (2009). Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *Ieee conference on computer vision and pattern recognition* (pp. 1446–1453).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., & Serre, T. (2011). Hmdb: a large video database for human motion recognition. In *Ieee international conference on computer visio* (pp. 2556–2563).

Kumar, S., & Hebert, M. (2003). Discriminative random fields: A discriminative framework for contextual interaction in classification. In *Ieee international conference on computer vision* (pp. 1150–1157).

Laptev, I., Marszalek, M., Schmid, C., & Rozenfeld, B. (2008). Learning realistic human actions from movies. In *Ieee conference on computer vision and pattern recognition* (pp. 1–8).

Lara, O. D., & Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys*, *15*(3), 1192–1209.

Leung, M. K., & Yang, Y.-H. (1995). First sight: A human body outline labeling system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(4), 359–377.

Lewandowski, M., Makris, D., & Nebel, J.-C. (2010). View and style-independent action manifolds for human activity recognition. In *European conference on computer vision* (pp. 547–560). Springer.

Le Yaouanc, J.-M., & Poli, J.-P. (2012). A fuzzy spatio-temporal-based approach for activity recognition. In *Advances in conceptual modeling* (pp. 314–323). Springer.

Li, K., & Fu, Y. (2012). Arma-hmm: A new approach for early recognition of human activity. In *Ieee international conference on pattern recognition* (pp. 1779–1782).

Li, T., Chang, H., Wang, M., Ni, B., Hong, R., & Yan, S. (2015). Crowded scene analysis: A survey. *Circuits and Systems for Video Technology, IEEE Transactions on*, *25*(3), 367–386.

Li, W., Zhang, Z., & Liu, Z. (2010). Action recognition based on a bag of 3d points. In *Ieee computer society conference on computer vision and pattern recognition workshops* (pp. 9–14).

Li, X. (2003). Gesture recognition based on fuzzy c-means clustering algorithm. *Department Of Computer Science The University Of Tennessee Knoxville*.

Li, Z., Liu, W., & Zhang, Y. (2012). Adaptive fuzzy apporach to background modeling using pso and klms. In *World congress on intelligent control and automation* (pp. 4601–4607).

Lim, C. H., & Chan, C. S. (2012). A fuzzy qualitative approach for scene classification. In *Ieee international conference on fuzzy systems* (pp. 1–8).

Lim, C. H., & Chan, C. S. (2013). Fuzzy action recognition for multiple views within single camera. In *Ieee international conference on fuzzy systems* (pp. 1–8).

Lim, C. H., Risnumawan, A., & Chan, C. S. (2014). A scene image is nonmutually exclusive—a fuzzy qualitative scene understanding. *IEEE Transactions on Fuzzy*

Systems, *22*(6), 1541–1556.

Lim, C. H., Vats, E., & Chan, C. S. (2015). Fuzzy human motion analysis: A review. *Pattern Recognition*, *48*(5), 1773–1796. (Lim and Vats contributed equally.)

Lim, C. K., & Chan, C. S. (2011). Logical connectives and operativeness of bk subtriangle product in fuzzy inferencing. *International Journal of Fuzzy Systems*, *13*(4), 237–245.

Lim, C. K., & Chan, C. S. (2015). A weighted inference engine based on interval-valued fuzzy relational theory. *Expert Systems with Applications*, *42*(7), 3410–3419.

Lin, C., Chung, I., & Sheu, L. (2000). A neural fuzzy system for image motion estimation. *Fuzzy sets and systems*, *114*(2), 281–304.

Liu, H., Brown, D. J., & Coghill, G. M. (2008a). A fuzzy qualitative framework for connecting robot qualitative and quantitative representations. *IEEE Transactions on Fuzzy Systems*, *16*(3), 808–822.

Liu, H., Brown, D. J., & Coghill, G. M. (2008b). Fuzzy qualitative robot kinematics. *IEEE Transactions on Fuzzy Systems*, *16*(6), 1522–1530.

Liu, H., & Coghill, G. M. (2005). Fuzzy qualitative trigonometry. In *Ieee conference on systems, man and cybernetics* (pp. 1291–1296).

Liu, H., Coghill, G. M., & Barnes, D. P. (2009). Fuzzy qualitative trigonometry. *International Journal of Approximate Reasoning*, *51*(1), 71–88.

Liu, J., Luo, J., & Shah, M. (2009). Recognizing realistic actions from videos “in the wild”. In *Ieee conference on computer vision and pattern recognition* (pp. 1996–2003).

Lu, Y., Boukharouba, K., Boonært, J., Fleury, A., & Lecœuche, S. (2014). Application of an incremental svm algorithm for on-line human recognition from video surveillance using texture and color features. *Neurocomputing*, *126*, 132–140.

Lyons, M. J., Budynek, J., & Akamatsu, S. (1999). Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(12), 1357–1362.

Maddalena, L., & Petrosino, A. (2010). A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection. *Neural Computing and Applications*, *19*(2), 179–186.

Mahapatra, A., Mishra, T. K., Sa, P. K., & Majhi, B. (2013). Background subtraction and human detection in outdoor videos using fuzzy logic. In *Ieee international conference on fuzzy systems* (pp. 1–7).

Marín-Jiménez, M., Muñoz-Salinas, R., Yeguas-Bolivar, E., & de la Blanca, N. P. (2014). Human interaction categorization by using audio-visual cues. *Machine Vision and Applications*, *25*(1), 71–84.

Marszalek, M., Laptev, I., & Schmid, C. (2009). Actions in context. In *Ieee conference on computer vision and pattern recognition* (pp. 2929–2936).

Meng, Y. K. (1997). Interval-based reasoning in medical diagnosis. In *Ieee iis* (p. 32).

Mikolov, T., Deoras, A., Povey, D., Burget, L., & Černockỳ, J. (2011). Strategies for training large scale neural network language models. In *Ieee workshop on automatic speech recognition and understanding* (pp. 196–201).

Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *37*(3), 311–324.

Mitra, S., & Hayashi, Y. (2000). Neuro-fuzzy rule generation: survey in soft computing framework. *IEEE Transactions on Neural Networks*, *11*(3), 748–768.

Moeslund, T. B., & Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, *81*(3), 231–268.

Moeslund, T. B., Hilton, A., & Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, *104*(2), 90–126.

Mozafari, K., Charkari, N. M., Boroujeni, H. S., & Behrouzifar, M. (2012). A novel fuzzy hmm approach for human action recognition in video. In *Knowledge technology* (pp. 184–193). Springer.

Nghiem, A. T., Bremond, F., Thonnat, M., & Valentin, V. (2007). Etiseo, performance evaluation for video surveillance systems. In *Ieee conference on advanced video and signal based surveillance* (pp. 476–481).

Nie, W., Liu, A., Su, Y., Luan, H., Yang, Z., Cao, L., & Ji, R. (2014). Single/cross-camera multiple-person tracking by graph matching. *Neurocomputing*, *139*, 220–232.

Niebles, J. C., Chen, C.-W., & Fei-Fei, L. (2010). Modeling temporal structure of decomposable motion segments for activity classification. In *European conference on computer vision* (pp. 392–405). Springer.

Ning, H., Tan, T., Wang, L., & Hu, W. (2004). Kinematics-based tracking of human walking in monocular video sequences. *Image and Vision Computing*, *22*(5), 429–441.

Niyogi, S. A., & Adelson, E. H. (1994). Analyzing and recognizing walking figures in xyt. In *Ieee conference on computer vision and pattern recognition* (pp. 469–474).

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, *42*(3), 145–175.

Parikh, D., & Grauman, K. (2011). Relative attributes. In *Ieee international conference on computer vision* (pp. 503–510).

Patron-Perez, A., Marszalek, M., Zisserman, A., & Reid, I. D. (2010). High five: Recognising human interactions in tv shows. In *British machine vision conference* (pp. 1–11).

Peng, H., Wang, J., Pérez-Jiménez, M. J., & Shi, P. (2013). A novel image thresholding method based on membrane computing and fuzzy entropy. *Journal of Intelligent and Fuzzy Systems*, *24*(2), 229–237.

Pentland, A. (2000). Looking at people: Sensing for ubiquitous and wearable computing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 107–119.

Popoola, O. P., & Wang, K. (2012). Video-based abnormal human behavior recognition - a review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C Applications and Reviews*, *42*(6), 865–878.

Poppe, R. (2007). Vision-based human motion analysis: An overview. *Computer vision and image understanding*, *108*(1), 4–18.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and vision computing*, *28*(6), 976–990.

Porikli, F., Tuzel, O., & Meer, P. (2006). Covariance tracking using model update based on lie algebra. In *Ieee conference on computer vision and pattern recognition* (Vol. 1, pp. 728–735).

Ragheb, H., Velastin, S., Remagnino, P., & Ellis, T. (2008). Vihasi: virtual human action silhouette data for the performance evaluation of silhouette-based action recognition methods. In *Ieee international conference on distributed smart cameras* (pp. 1–10).

Rautaray, S. S., & Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, *43*(1), 1–54.

Reddy, K. K., & Shah, M. (2013). Recognizing 50 human action categories of web videos. *Machine Vision and Applications*, *24*(5), 971–981.

Rehg, J. M., & Kanade, T. (1995). Model-based tracking of self-occluding articulated objects. In *International conference on computer vision* (pp. 612–617).

Rhee, F. C.-H., & Krishnapuram, R. (1993). Fuzzy rule generation methods for high-level computer vision. *Fuzzy Sets and Systems*, *60*(3), 245–258.

Rodriguez, M. D., Ahmed, J., & Shah, M. (2008a). Action mach a spatio-temporal maximum average correlation height filter for action recognition. In *Ieee conference on computer vision and pattern recognition* (pp. 1–8).

Rodriguez, M. D., Ahmed, J., & Shah, M. (2008b). Action mach: a spatio-temporal maximum average correlation height filter for action recognition. In *Ieee international conference on computer vision and pattern recognition* (pp. 1–8).

Rohr, K. (1994). Towards model-based recognition of human movements in image sequences. *Image understanding*, *59*(1), 94–115.

Rubin, S. H. (1999). Computing with words. *IEEE Transactions on Systems, Man, and*

Cybernetics, Part B: Cybernetics, *29*(4), 518–524.

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). Labelme: a database and web-based tool for image annotation. *International journal of computer vision*, *77*(1-3), 157–173.

Ryoo, M. S. (2011). Human activity prediction: Early recognition of ongoing activities from streaming videos. In *Ieee international conference on computer vision* (pp. 1036–1043).

Ryoo, M. S., & Aggarwal, J. K. (2009). Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In *International conference on computer vision* (pp. 1593–1600).

Ryoo, M. S., Fuchs, T. J., Xia, L., Aggarwal, J. K., & Matthies, L. (2014). Early recognition of human activities from first-person videos using onset representations. *arXiv preprint arXiv:1406.5309*.

Sainath, T. N., Mohamed, A.-r., Kingsbury, B., & Ramabhadran, B. (2013). Deep convolutional neural networks for lvcsr. In *Ieee international conference on acoustics, speech and signal processing* (pp. 8614–8618).

Sanchez-Valdes, D., Alvarez-Alvarez, A., & Trivino, G. (2015). Walking pattern classification using a granular linguistic analysis. *Applied Soft Computing*, *33*, 100 - 113.

Sapienza, M., Cuzzolin, F., & Torr, P. H. (2014). Learning discriminative space–time action parts from weakly labelled videos. *International Journal of Computer Vision*, 1–18.

Schuldt, C., Laptev, I., & Caputo, B. (2004). Recognizing human actions: a local svm approach. In *International conference on pattern recognition* (pp. 32–36).

See, J., Lee, S., & Hanmandlu, M. (2005). Human motion detection using fuzzy rule-base classification of moving blob regions. In *International conference on robotics, vision, information and signal processing* (pp. 398–402).

Shakeri, M., Deldari, H., Foroughi, H., Saberi, A., & Naseri, A. (2008). A novel fuzzy background subtraction method based on cellular automata for urban traffic applications. In *International conference on signal processing* (pp. 899–902).

Silaghi, M.-C., Plänkers, R., Boulic, R., Fua, P., & Thalmann, D. (1998). Local and global skeleton fitting techniques for optical motion capture. In *Modelling and motion capture techniques for virtual environments* (pp. 26–40). Springer.

Simha, S. M., Chau, D. P., & Bremond, F. (2014). Feature matching using co-inertia analysis for people tracking. In *International conference on computer vision theory and applications.*

Singh, S., Velastin, S. A., & Ragheb, H. (2010). Muhavi: A multicamera human action video dataset for the evaluation of action recognition methods. In *Ieee international conference on advanced video and signal based surveillance* (pp. 48–55).

Soomro, K., Zamir, A. R., & Shah, M. (2012). Ucf101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2014). Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*.

Tompson, J. J., Jain, A., LeCun, Y., & Bregler, C. (2014). Joint training of a convolutional network and a graphical model for human pose estimation. In *Advances in neural information processing systems* (pp. 1799–1807).

Tran, D., & Sorokin, A. (2008). Human activity recognition with metric learning. In *European conference on computer vision* (pp. 548–561). Springer.

Trivino, G., & van der Heide, A. (2008). Linguistic summarization of the human activity using skin conductivity and accelerometers. In *Conference on information processing and management of uncertainty in knowledge-based systems* (pp. 1583–1589).

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, *2*(5), 2.

Tsochantaridis, I., Joachims, T., Hofmann, T., & Altun, Y. (2005). Large margin methods for structured and interdependent output variables. In *Journal of machine learning research* (pp. 1453–1484).

Tu, H.-b., Xia, L.-m., & Wang, Z.-w. (2014). The complex action recognition via the correlated topic model. *The Scientific World Journal*, *2014*.

Turaga, P., Chellappa, R., Subrahmanian, V. S., & Udrea, O. (2008). Machine recognition of human activities: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, *18*(11), 1473–1488.

Ullah, A. S., D'Addona, D., & Arai, N. (2014). Dna based computing for understanding complex shapes. *Biosystems*, *117*, 40–53.

Várkonyi-Kóczy, A. R., & Tusor, B. (2011). Human–computer interaction for smart environment applications using fuzzy hand posture and gesture models. *IEEE Transactions on Instrumentation and Measurement*, *60*(5), 1505–1514.

Vats, E., Lim, C. K., & Chan, C. S. (2012). A bk subproduct approach for scene classification. In *Iieej image electronics and visual computing workshop* (pp. 1–5).

Vats, E., Lim, C. K., & Chan, C. S. (2015). An improved bk sub-triangle product approach for scene classification. *Journal of Intelligent & Fuzzy Systems*, *29*(5), 1923–1931.

Verma, R., & Dev, A. (2009). Vision based hand gesture recognition using finite state machines and fuzzy logic. In *International conference on ultra modern telecommunications & workshops* (pp. 1–6).

Vogel, J., & Schiele, B. (2004). Natural scene retrieval based on a semantic modeling step. In *Image and video retrieval* (pp. 207–215). Springer.

Vogel, J., & Schiele, B. (2007). Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, *72*(2), 133–157.

Wachs, J., Kartoun, U., Stern, H., & Edan, Y. (2002). Real-time hand gesture telerobotic system using fuzzy c-means clustering. In *Ieee biannual world automation congress* (pp. 403–409).

Wachs, J. P., Stern, H., & Edan, Y. (2005). Cluster labeling and parameter estimation for the automated setup of a hand-gesture recognition system. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *35*(6), 932–944.

Wachter, S., & Nagel, H.-H. (1997). Tracking of persons in monocular image sequences. In *Ieee nonrigid and articulated motion workshop* (pp. 2–9).

Wang, H., & Schmid, C. (2013). Action recognition with improved trajectories. In *Ieee international conference on computer vision* (pp. 3551–3558).

Wang, L., Hu, W., & Tan, T. (2003). Recent developments in human motion analysis. *Pattern Recognition*, *36*(3), 585–601.

Wang, L.-X., & Mendel, J. M. (1992). Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*, *22*(6), 1414–1427.

Wang, X., Wang, Y., Xu, X., Ling, W., & Yeung, D. S. (2001). A new approach to fuzzy rule generation: fuzzy extension matrix. *Fuzzy Sets and Systems*, *123*(3), 291–306.

Wang, Y., Huang, K., & Tan, T. (2007). Human activity recognition based on r transform. In *Ieee conference on computer vision and pattern recognition* (pp. 1–8).

Wang, Z., & Zhang, J. (2008). Detecting pedestrian abnormal behavior based on fuzzy associative memory. In *Conference on natural computation* (pp. 143–147).

Weinland, D., Ronfard, R., & Boyer, E. (2006). Free viewpoint action recognition using motion history volumes. *Computer Vision and Image Understanding*, *104*(2), 249–257.

Weinland, D., Ronfard, R., & Boyer, E. (2011). A survey of vision-based methods for action representation, segmentation and recognition. *Computer Vision and Image Understanding*, *115*(2), 224–241.

Wilbik, A., & Keller, J. M. (2013). A fuzzy measure similarity between sets of linguistic summaries. *IEEE Transactions on Fuzzy Systems*, *21*(1), 183–189.

Wilbik, A., Keller, J. M., & Alexander, G. L. (2011). Linguistic summarization of sensor data for eldercare. In *Ieee international conference on systems, man, and cybernetics* (pp. 2595–2599).

Wongkhuenkaew, R., Auephanwiriyakul, S., & Theera-Umpon, N. (2013). Multi-prototype fuzzy clustering with fuzzy k-nearest neighbor for off-line human action recognition. In *Ieee international conference on fuzzy systems* (pp. 1–7).

Wren, C. R., Azarbayejani, A., Darrell, T., & Pentland, A. P. (1997). Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 780–785.

Wu, D., & Mendel, J. M. (2007). Uncertainty measures for interval type-2 fuzzy sets. *Information Sciences*, *177*(23), 5378–5393.

Wu, D., & Shao, L. (2014). Multi-max-margin support vector machine for multi-source human action recognition. *Neurocomputing*, *127*, 98–103.

Wu, H., & Mendel, J. M. (2002). Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems. *IEEE Transactions on Fuzzy Systems*, *10*(5), 622–639.

Wu, H., Sun, F., & Liu, H. (2008). Fuzzy particle filtering for uncertain systems. *IEEE Transactions on Fuzzy Systems*, *16*(5), 1114–1129.

Wu, S., Moore, B. E., & Shah, M. (2010). Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In *Ieee conference on computer vision and pattern recognition* (pp. 2054–2060).

Wu, S., Oreifej, O., & Shah, M. (2011). Action recognition in videos acquired by a moving camera using motion decomposition of lagrangian particle trajectories. In *Ieee international conference on computer vision* (pp. 1419–1426).

Wu, Y., & Huang, T. S. (1999). Vision-based gesture recognition: A review. In *Gesture-based communication in human-computer interaction* (pp. 103–115). Springer.

Xie, D., Hu, W., Tan, T., & Peng, J. (2004). A multi-object tracking system for surveillance video analysis. In *International conference on pattern recognition* (pp. 767–770).

Yager, R. R. (2002). Uncertainty representation using fuzzy measures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *32*(1), 13–20.

Yang, X., & Tian, Y. (2014). Effective 3d action recognition using eigenjoints. *Journal of Visual Communication and Image Representation*, *25*(1), 2–11.

Yao, B., Hagras, H., Al Ghazzawi, D., & Alhaddad, M. J. (2012). An interval type-2 fuzzy logic system for human silhouette extraction in dynamic environments. In *Autonomous and intelligent systems* (pp. 126–134). Springer.

Yao, B., Hagras, H., Alhaddad, M. J., & Alghazzawi, D. (2014). A fuzzy logic-based system for the automation of human behavior recognition using machine vision in intelligent environments. *Soft Computing*, 1–8.

Yeguas-Bolivar, E., Muñoz-Salinas, R., Medina-Carnicer, R., & Carmona-Poyato, A. (2014). Comparing evolutionary algorithms and particle filters for markerless human motion capture. *Applied Soft Computing*, *17*, 153 - 166.

Yew, K., & Kohout, L. (1996). Interval-valued fuzzy relational inference structures. In *International conference on intelligent information management systems* (pp. 173–175).

Yoon, C., Cheon, M., & Park, M. (2013). Object tracking from image sequences using adaptive models in fuzzy particle filter. *Information Sciences*, *253*, 74–99.

Yu, G., Yuan, J., & Liu, Z. (2012). Predicting human activities using spatio-temporal structure of interest points. In *Acm international conference on multimedia* (pp. 1049–1052).

Yu, M., Naqvi, S. M., Rhuma, A., & Chambers, J. (2011). Fall detection in a smart room by using a fuzzy one class support vector machine and imperfect training data. In *Ieee international conference on acoustics, speech and signal processing* (pp. 1833–1836).

Yuan, J., Liu, Z., & Wu, Y. (2009). Discriminative subvolume search for efficient action detection. In *Ieee conference on computer vision and pattern recognition* (pp. 2442–2449).

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man and Cybernetics*(1), 28–44.

Zadeh, L. A. (1996). Fuzzy logic= computing with words. *IEEE Transactions on Fuzzy Systems*, *4*(2), 103–111.

Zaki, M., & Abulwafa, M. (2002). The use of invariant features for object recognition from a single image. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, *12*(2), 79–95.

Zeng, J., & Liu, Z.-Q. (2006). Type-2 fuzzy sets for handling uncertainty in pattern recognition. In *Ieee international conference on fuzzy systems* (pp. 1247–1252).

Zha, Z.-J., Zhang, H., Wang, M., Luan, H., & Chua, T.-S. (2013). Detecting group activities with multi-camera context. *IEEE Transactions on Circuits and Systems for Video Technology*, *23*(5), 856–869.

Zhang, H., & Xu, D. (2006). Fusing color and texture features for background model. In *International conference on fuzzy systems and knowledge discovery* (pp. 887–893).

Zhang, L., Tao, D., Liu, X., Sun, L., Song, M., & Chen, C. (2013). Grassmann multimodal implicit feature selection. *Multimedia Systems*, 1–16.

Zhao, Z., Bouwmans, T., Zhang, X., & Fang, Y. (2012). A fuzzy background modeling approach for motion detection in dynamic backgrounds. In *Multimedia and signal processing* (pp. 177–185). Springer.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Lim, C. H., Vats, E., & Chan, C. S. (2015). Fuzzy human motion analysis: A review. *Pattern Recognition*, *48*(5), 1773–1796. (Lim and Vats contributed equally.)

Vats, E., & Chan, C. S. (2015). Early anticipation of human behaviour – a hybrid approach. *Applied Soft Computing*. (In Press.)

Vats, E., Lim, C. K., & Chan, C. S. (2015). Early human actions detection using bk sub-triangle product. *IEEE International Conference on Fuzzy Systems*, 1–8. (Accepted. Best student paper award nomination.)

Vats, E., Lim, C. K., & Chan, C. S. (2015). An improved bk sub-triangle product approach for scene classification. *Journal of Intelligent & Fuzzy Systems*, *29*(5), 1923–1931.