Moment Invariant based Human Motion Detection, Recognition and Classification at Dynamic Thresholding Techniques

By

Hashim Yasin

2005-NUST-MS PhD-CSE-28

MS-5

Thesis Supervisor:
Lt Col Dr Shoab Ahmad Khan

Department of Computer Engineering
College of Electrical and Mechanical Engineering
National University of Sciences and Technology, Pakistan.

2008

بسم الله الرحمن الرحيم

Moment Invariant based Human Motion Detection, Recognition and Classification at Dynamic Thresholding Techniques

By

Hashim Yasin

2005-NUST-MS PhD-CSE-28

MS-5

Submitted to the Department of Computer Engineering
in fulfillment of the requirements for the degree of

Master of Science
in
Computer Software Engineering

Thesis Supervisor:
Lt Col Dr Shoab Ahmad Khan

College of Electrical and Mechanical Engineering
National University of Sciences and Technology, Pakistan.

2008

Acknowledgement

All thanks to Allah Almighty, the most Merciful, the most Beneficent, who has made man the supreme of all creatures through endowing him with wisdom and knowledge. He, the Omnipotent, has declared man His vicegerent and bestowed on him many of His virtues and excellences; and edified his spiritual, contemplative, cognitive and perceptive capabilities. He has also conferred on him the faculties of logic and reasoning along with the sensational, emotional and sentimental aspects with regard to his dealings with the other human beings and his belongings. He has also bestowed on man the aesthetic and artistic approach towards life and his surroundings. Allah Almighty, the Omnipresent and the Omniscient has called attention of man at various places in the Holy Quran to exploring the universe and contemplating its making and pondering over all its systems and their functions. He has also invited man to delve into the universe and conquer it through his endeavor and research. He has made those nations deserve the progress and improvement who adopt the reflective and contemplative approach and vision about the universe and all therein. After this the researcher sending blessing and benediction is very thankful to Hazrat Muhammad (Sallala Ho Elahey Waallay Hi Wasallam), the messenger of peace and humanity. He (S.A.W.) is the benefactor of humanity who enhanced the status of knowledge and laid emphasis to its acquisition from cradle to grave exerting all efforts, and raised its attainment to the status of worship. He (S.A.W.) has even granted knowledge and scholarship superiority over worship.

The researcher is very much thankful to his supervisor *Lt Col. Dr. Shoab Ahmad Khan* for his full help in the research work and preparing this dissertation. He has been a light for the researcher through the unexplored regions of research. He first of all assimilated the ideas in the researcher’s mind and then showed him the right direction from among the enmeshed and unseen ones. He always demonstrated the ways to unscramble and straighten out the jumbled and puzzled courses of thoughts and made the researcher feel easy to grope through the difficult ways of the field of research and have an insight into the unfathomable and profound sea of knowledge and a vast and infinite region of search. In the course of research his indulgence and benevolence always encouraged the researcher through the difficult paths. His heartiest concern and encouraging style of supervision have always consorted with the researcher. Many heartfelt thanks to *Lt Col Dr Shoab Ahmad Khan*.

The researcher is also very thankful to the thesis assessment committee comprising *Brig Dr Muhammad Younus Javed* (Head of Department), *Dr Shaleeza Sohail* and *Dr Aasia Khanum*. The committee had been a thought- provoking source and a complete academic and technical guidance in the proper lay-out and arrangement of this thesis. They provided the know-how to give the conspectus of the research a conspicuous look and make it consonant with the true aims of research so that it might reflect all aspects of the research work. Their over all guidance and time to time suggestions and hints saved a lot of time and smoothened the way to a successful preparation of this thesis.

DEDICATED

To

MY

PARENTS

Under whose feet my heaven lies.

Table of Contents

Acknowledgement.. iii

Abstract.. vi

Table of Contents.. viii

List of Figures.. xii

List of Tables... xvi

Acronyms.. xviii

1. Introduction 1

1.1 Motivation of the Research …………………….................................... **1**

1.2 Purpose of the Research ………………………..…………………….. **2**

1.3 Problem Statement... **3**

1.4 Overview of the Research.. **4**

1.5 Goals and Objectives …………………….. **6**

1.6 Thesis Organization …………………………………….……………….. **6**

1.7 Chapter Summary ……………………... **7**

2. Literature Review 8

2.1 Low Level Vision … Motion Detection …………………..………… **8**

2.1.1 Segmentation ………………………………………….……... 9

2.1.1.1 Background Subtraction Approaches ………....................... 9

2.1.1.2 Temporal Differencing Approach …..……....................……... 11

2.1.1.3 Optical Flow …..…………………….................….............. 11

2.1.2 Thresholding Techniques...…………………… 11

2.1.3 Human Motion Detection.………………….................………... 12

2.2 High Level Vision … Motion Representation and Recognition. **14**

2.2.1 View-based Motion Recognition.…………...................……...… 15

2.2.2 View-Invariant Motion Recognition.….................…………….. 16

2.3 Human Pose Representation and Recognition ....… **17**

2.4 Chapter Summary ……………………….........................……… **19**

3. Methodology 20

3.1 Introduction …………….. **20**

3.2 Problem Definition.…….. **21**

3.3 Methodology of Unit-I….. **22**

3.3.1 Frames Extraction... 23

3.3.2 Background Subtraction... 24

3.3.3 Segmentation & Thresholding... 25

3.3.3.1 Otsu’s Threshold Method.. 26

3.3.3.2 Isodata Threshold Method... 27

3.3.3.3 Image Histogram Shape based Threshold Method.............. 27

3.3.3.4 Maximum Entropy based Threshold Method....................... 30

3.3.4 Motion History Image (MHI) Formation..................................... 30

3.3.5 Hu Moment Invariants... 33

3.3.5.1 Raw Moments.. 34

3.3.5.2 Central Moments.. 35

3.3.5.3 Scale Invariant Moments.. 37

3.3.5.4 Rotation Invariant Moments... 37

3.3.6 Training & Testing Datasets Formation.. 39
3.3.7 Recognition & Classification... 39
3.3.7.1 Mahalanobis Distance (MD) Classifier.............................. 40
3.3.7.2 Quadratic Bayes Gaussian (QBG) Classifier....................... 40
3.3.7.3 Linear Bayes Gaussian (LBG) Classifier............................ 41
3.3.7.4 K-Nearest Neighbor (KNN) Classifier............................... 42
3.3.7.5 Fuzzy K-Nearest Neighbor (FKNN) Classifier.................. 42
3.4 Methodology of Unit-II…….. **43**
3.4.1 Module-I and Module-II……………….. 43
3.4.2 Module-III……………………………....................................... 45
3.5 Methodology of Unit-III... **48**
3.6 Chapter Summary.. **53**

4. Results & Discussion 54

4.1 Assumptions and Preconditions... **54**
4.2 Experimental Results of Unit-I... **56**
4.2.1 Module-I of Unit-I *without* BGS... 56
4.2.1.1 Performance Evaluations of Thresholding Methods *without* BGS ……………………………………………. 58
4.2.1.2 Performance Evaluations of Classifiers *without* BGS…........ 58
4.2.2 Module-II of Unit-I *with* BGS... 60
4.2.2.1 Performance Evaluations of Thresholding Methods *with* BGS ……………………………………................. 61
4.2.2.2 Performance Evaluations of Classifiers with BGS …........... 62
4.2.3 Experimental Assessment of Module-I & Module-II of Unit-I...... 63
4.2.3.1 Assessment of Background Subtraction Approach.... 63
4.2.3.2 Assessment of Threshold Methods............................ 64
4.2.3.3 Assessment of Classifiers... 64
4.3 Experimental Results of Unit-II.. **67**

4.3.1 Performance Assessment of Unit-II …………………………........... 68

4.3.1.1 Performance Assessment in Terms of Success Rate.... 68

4.3.1.2 Performance Assessment in Terms of Time & Speed. 70

4.3.1.3 Performance Assessment in Terms of Storage Capacity…. 71

4.4 Experimental Results of Unit-III... .... **72**

4.4.1 Performance Evaluation of Unit-III …………………………........... 73

4.4.1.1 Performance Evaluation in Terms of Success Rate............ 73

4.4.1.2 Performance Evaluation in Terms of Time & Speed….….... 74

4.5 Chapter Summary.. **75**

5. Analysis & Conclusion 76

5.1 Analysis – Unit-I.. **76**

5.2 Analysis – Unit-II... **78**

5.3 Analysis – Unit-III.. **79**

5.4 Chapter Summary.. **79**

6. Summary & Future Work 81

6.1 Summary.. **81**

6.2 Future Work... **82**

Bibliography 84

A. Frames & MHIs 94

B. Hu Moment Invariants 100

C. PDV and PMV 103

C.1 Pixels' Difference Values (PDV).. **103**

C.2 Pixels' Motion Values (PMV)... **103**

List of Figures

Figure 1.1: Selected frames from video “Sitting down” motion [1]-[3] 2

Figure 1.2: A basic “Framework Model” for Human Motion Analysis System [4]…………………………………………………… 3

Figure 2.1: Fujiyoshi and Lipton’s Star-Skeleton Model [66]……......... 18

Figure 3.1: Five different types of human motion …………………….... 21

Figure 3.2: Classifiers’ performances with “Trail & Error Threshold” method …………………………………...………..….. 22

Figure 3.3: Frames extracted from “Bending down” motion video clip.. 24

Figure 3.4: “Bending down” motion: (a) Known background image (b) Last static image frame (c) Image frame after Background Subtraction (d) Thresholded image of “b” (e) Thresholded image of “c”………………..………………………….. 25

Figure 3.5: “Gun Shot” motion: (a) Last static image frame (b) Image frame after Background Subtraction (c) Image frame after “Otsu’s Threshold” method ………………………………... 26

Figure 3.6: “Kicking front” motion: (a) Last static image frame (b) Image frame after Background Subtraction (c) Image

frame after “Isodata Threshold” method ……………….…. 27
Figure 3.7: “Bending down” motion: (a) Last static image frame (b) Image frame after Background Subtraction (c) Histogram of “a” (d) Histogram of “b”.. 28
Figure 3.8: “Punching forward” motion: (a) Last static image frame (b) Image frame after Background Subtraction (c) Image frame after “Image Histogram Shape based Threshold” method …………………………………………………... 29
Figure 3.9: Histogram of last static image frame of “Punching forward” motion…………………………………………… 29
Figure 3.10: “Jumping up” motion: (a) Last static image frame (b) Image frame after Background Subtraction (c) Image frame after “Maximum Entropy based Threshold” method. 30
Figure 3.11: “Bending down” motion (a) Motion Energy Image (MEI) (b) Motion History Image (MHI)....................................... 32
Figure 3.12: Motion History Images created using “Isodata Threshold” method and *without* Background Subtraction approach (a) Bending down (b) Gun Shot (c) Jumping up (d) Kicking front (e) Punching forward.… …… ………………………. 32
Figure 3.13: Motion History Image created using “Isodata Threshold” method and *with* Background Subtraction (a) Bending down (b) Gun Shot (c) Jumping up (d) Kicking front (e) Punching forward …………..…………………………... 33
Figure 3.14: “Kicking front” motion: (a) Image frame no. 19 (b) Image frame after Background Subtraction. (c) Image frame after “Otsu’s Threshold” method….…………….……….. 44
Figure 3.15: Hu Moment Invariants of all created thresholded image frames of motion types: (a) Bending down (b) Gun Shot

(c) Jumping up (d) Kicking front (e) Punching forward….. 52

Figure 4.1: Different types of “Bending down” motion video clips……. 55

Figure 4.2: Classifiers’ performances under different Thresholding techniques *without* Background Subtraction approach…….. 59

Figure 4.3: Classifiers’ performances under different Thresholding techniques *with* Background Subtraction (BGS) approach… 63

Figure 4.4: Classifiers’ performances under “Otsu’s Threshold” method *with* BGS & *without* BGS approaches................................. 65

Figure 4.5: Classifiers’ performances under “Isodata Threshold” method *with* BGS & *without* BGS approaches …………….. 65

Figure 4.6: Classifiers’ performances under “Image Histogram Shape based Threshold” method *with* BGS & *without* BGS approaches... 66

Figure 4.7: Classifiers’ performances under “Maximum Entropy based Threshold” method *with* BGS & *without* BGS approaches.... 66

Figure 4.8: Classifiers’ performances in terms of success rate................ 69

Figure 4.9: Classifiers’ performances in terms of time & speed ……….. 71

Figure 4.10: Recognition & Classification techniques’ performances in terms of memory capacity …... 72

Figure 4.11: Classifiers’ performances in terms of success rate ……..… 74

Figure 4.12: Classifiers’ performances in terms of time & speed ……..... 75

Figure A.1: Motion History Images: (a) Bending down (b) Gun Shot (c) Jumping up (d) Kicking front (e) Punching forward............ 94

Figure A.2: Frames extracted from “Bending down” motion video clip... 95

Figure A.3: Motion History Image (MHI) formation from extracted frames of “Bending down” motion video clip..... 95

Figure A.4: Frames extracted from “Gun Shot” motion video clip.......... 96

Figure A.5: Motion History Image (MHI) formation from extracted frames of "Gun Shot" motion video clip............................. 96
Figure A.6: Frames extracted from "Jumping up" motion video clip….. 97
Figure A.7: Motion History Image (MHI) formation from extracted frames of "Jumping up" motion video clip..... 97
Figure A.8: Frames extracted from "Kicking front" motion video clip… 98
Figure A.9: Motion History Image (MHI) formation from extracted frames of "Kicking front" motion video clip………………. 98
Figure A.10: Frames extracted from "Punching forward" motion video clip…………………………………………………….... 99
Figure A.11: Motion History Image (MHI) formation from extracted frames of "Punching forward" motion video clip……….. 99
Figure B.1: Hu Moment Invariants of all created thresholded image frames of motion types: (a) Bending down (b) Gun Shot (c) Jumping up (d) Kicking front (e) Punching forward….. 102
Figure C.1: "Bending down" motion (a) Extracted frames (b) Pixels' Difference Values (PDV) (c) Pixels' Motion Values (PMV) 104
Figure C.2: "Gun Shot" motion (a) Extracted frames (b) Pixels' Difference Values (PDV) (c) Pixels' Motion Values (PMV) 105
Figure C.3: "Jumping up" motion (a) Extracted frames (b) Pixels' Difference Values (PDV) (c) Pixels' Motion Values (PMV) 106
Figure C.4: "Kicking front" motion (a) Extracted frames (b) Pixels' Difference Values (PDV) (c) Pixels' Motion Values (PMV) 107
Figure C.5: "Punching forward" motion (a) Extracted frames (b) Pixels' Difference Values (PDV) (c) Pixels' Motion Values (PMV) ……………………………………………………… 108

List of Tables

Table 3.1: Some examples of Hu Moment Invariants........................... 38

Table 3.2: Training & Testing Datasets' sizes....................................... 39

Table 3.3: Training & Testing Datasets for Module-I........................... 45

Table 3.4: Training & Testing Datasets for Module-II.......................... 45

Table 4.1: Successful detections under "Otsu's Threshold" method *without* Background Subtraction approach…......................... 56

Table 4.2: Successful detections under "Isodata Threshold" method *without* Background Subtraction approach........................... 57

Table 4.3: Successful detections under "Image Histogram Shape based Threshold" method *without* Background Subtraction approach... 57

Table 4.4: Successful detections under "Maximum Entropy based Threshold" method *without* Background Subtraction approach.. 57

Table 4.5: Classifiers and Thresholding methods' evaluations *without* Background Subtraction approach....................................... 59

Table 4.6: Successful detections under "Otsu's Threshold" method

with Background Subtraction approach………………..……… 60
Table 4.7: Successful detections under “Isodata Threshold” method *with* Background Subtraction approach................................ 60
Table 4.8: Successful detections under “Image Histogram Shape based Threshold” method *with* Background Subtraction approach... 61
Table 4.9: Successful detections under “Maximum Entropy based Threshold” method *with* Background Subtraction approach... 61
Table 4.10: Classifiers and Thresholding methods’ evaluations *with* Background Subtraction approach....................................... 62
Table 4.11: Classifiers’ success rate when Training Dataset = 105x7 elements………………………….………………………..……. 67
Table 4.12: Classifiers’ success rate when Training Dataset = 200x 7 elements... 67
Table 4.13: TTRC’s success rate... 68
Table 4.14: Recognition & Classification Techniques’success rate…... 69
Table 4.15: Recognition & Classification Techniques’ executed time... 70
Table 4.16: Recognition & Classification Techniques’ memory capacity……………………………………………….………… 71
Table 4.17: Classifiers’ success rate under Unit-III............................... 72
Table 4.18: Classifiers’ executed time under Unit-III............................ 73

Acronyms

BGS	Background Subtraction
MHI	Motion History Image
MEI	Motion Energy Image
TTRC	Trained Table based Recognition & Classification
MIRC	Moment Invariant based Recognition and Classification
MD	Mahalanobis Distance
QBG	Quadratic Bayes Gaussian
LBG	Linear Bayes Gaussian
KNN	K- Nearest Neighbor
FKNN	Fuzzy K- Nearest Neighbor
GMM	Gaussian Mixture Models
Pfinder	Person Finder’
HSV	Hue, Saturation, and Value
VSAM	Visual Surveillance and Monitoring
EM	Expectation Maximization
LDA	Linear Discriminate Analysis
SVM	Support Vector Machines
FoV	Field of View

PCA	Principal Components Analysis
tMHI	timed Motion History Image
HMM	Hidden Markov Models
CHMM	Coupled Hidden Markov Models
NN	Neural Network
MLD	Moving Light Display
MEV	Motion Energy Volume
MHV	Motion History Volume
fps	Frames per Second
avi	Audio Video Interleave
TT	Trained Table
MaxVC	Maximum value of Column
MinVC	Minimum value of Column
PDV	Pixels' Difference Values
PMV	Pixels' Motion Values

1

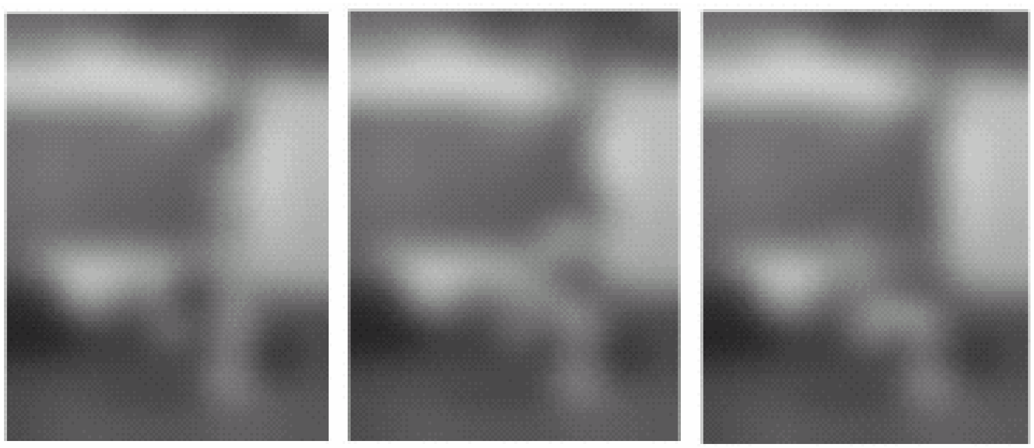
Introduction

Computer Vision and Digital Image Processing techniques play a vital role in the area of human behavior study, human motion reconstruction and human motion analysis especially in detection, recognition and classification of human motions under video surveillance system. Human motion analysis and understanding of human actions under video surveillance system is the demand of such an insecure situation of the world. Human Motion Analysis Systems find application & implementation in numerous sectors of our society including security surveillance application (for example, characterization of motion for identification of different types of actions and recognition/detection of suspicious or impostor behavior in video surveillance) as well as non-security related applications (for example, detection of postural disturbances due to mobility disorders or aging).

1.1 Motivation of the Research

Over last few years, the increasing need of security has promoted the importance of human behavior study and human motion & action analysis on a real time basis. A keen interest has appeared in the areas of detection, recognition and classification of human motions and understanding of actions. There are other domains as well that motivate the research in this area. They are content-based image storage & image retrieval, robotic motions, motion tracking, video surveillance systems, video conferencing, virtual reality and human–computer interactions.

The Figure 1.1 illustrates another aspect of the motivation of this research work. This Figure has been taken from [1]-[3] which represent that there is some movement. These images are blurred images; even then the motion is recognized as someone is sitting. Despite that even there is a total lack of recognizable features of these static images, but the movement is easily recognized [1]-[3]. In other words, motions or actions can be recognized and detected even when the most of information of static images is not available. This feature is very attractive and striking when Human Motion Analysis System receives incomplete and blurred video data due to some environmental factor.



Frame 5 Frame 25 Frame 40

Figure 1.1: Selected frames from video "Sitting down" motion [1]-[3]

1.2 Purpose of the Research

The purpose of this research work, based on computer vision techniques and human motion detection, recognition and classification system, is identification and understanding of the human motion and action in a video frame sequence. This research work may be a step ahead in this crucial field of human motion understanding.

There are a lot of approaches that have investigated the above mentioned problem. However, all of the approaches have almost a similar basic framework model in order to describe the problem. This basic framework model (see Figure 1.2) contains processes of Segmentation and Thresholding, Feature Extraction, Object Classification, Action Representation and Recognition. This research will focus on low level vision (Detection and Classification) and high level vision (Action Recognition and Behavior Understanding).

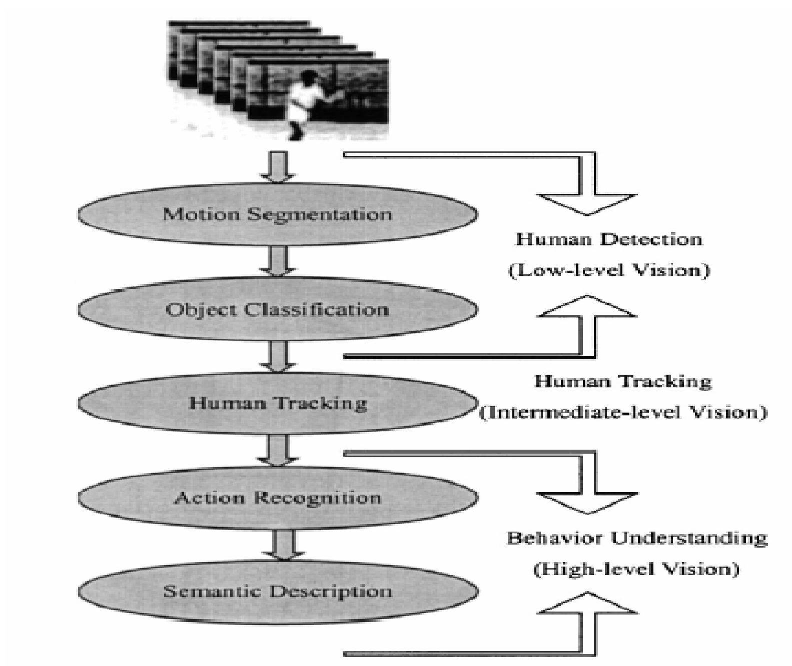


Figure 1.2: A basic “Framework Model” for Human Motion Analysis System [4]

1.3 Problem Statement

Classification and recognition of motions and actions on real time basis has been a very challenging and promising problem in the area of computer vision. All types of human motion analysis systems demand strength, accuracy, time and speed [4, 5]. Strength is required, because all systems need stability, constancy, continuity and automatic system. Accuracy is needed, because high security institutions demand steadiness and error-free human motion analysis systems. Time & Speed is necessary because it is the demand of real-time motion analysis systems [6]. So this research will converge on the above mentioned issues. In short, there is a crucial need of such a human motion analysis system that performs well, has high success rate and consumes less time. Another major requirement, on which this research work will focus, is to evaluate the performances of different approaches, techniques and classification methodologies that contribute in the process of recognition and classification.

In this research work, five different types of mistrustful motions and actions have been recorded in 'avi' file format. These actions are: Bending down, Gun Shot, Jumping up, Kicking front and Punching forward, performed by different subjects at different times. Each subject performs each action five to seven times forming training dataset and testing dataset for each type of action.

1.4 Overview of the Research

This research work has been divided into three units: Unit-I, Unit-II and Unit-III. Unit-I investigates the behavior of different classifiers under Otsu threshold method, Isodata threshold method, Image Histogram shape based threshold method and Maximum Entropy based threshold method. These thresholding techniques are used to find out threshold levels of the images created without Background Subtraction (BGS) approach while in the second category threshold levels are computed from images with Background Subtraction (BGS) approach. Motion History Image (MHI) is formed from these thresholded images using temporal template methodology [1]–[3]. Seven φ values of Hu Moment Invariants are used for feature descriptions of MHIs. From these Hu Moment Invariants calculated from MHIs, we develop training datasets as well as testing datasets. Using these training dataset and testing dataset, different classifiers' performances are measured at above mentioned dynamic thresholding techniques with and without Background Subtraction approaches. At the end, performances of classifiers are compared with Background Subtraction approach and without Background Subtraction approaches at different dynamic thresholding techniques in terms of accuracy and success rate.

Unit-II deals with temporal template methodology for motion recognition process, and Hu Moment Invariants are used for feature description. Firstly, two types of training datasets based on Hu Moment Invariants have been developed. One training dataset is of 105x7 elements and the other consists of 200x7 elements. Secondly, a new simple approach for motion recognition and classification called Trained Table based Recognition & Classification (TTRC) has been proposed. In TTRC approach, a simple data table has been trained on the placement of training dataset. The training process consists of the behavioral study of seven φ values of Hu Moment Invariants. Performance of TTRC is evaluated with other classification techniques using these training datasets in the context of accuracy, success rate, time & speed and memory capacity.

In Unit-III, a new recognition and classification methodology, Moment Invariant based Recognition and Classification (MIRC) has been proposed [6]. In this Unit, avi data is given as input and from that input, frames are extracted at the rate of 10 frames per second. After frame extraction, known background is subtracted from each created image frame. Otsu's threshold method is used to convert each image frame into binary image frame. Hu Moment Invariants are used in order to describe features of image frames. In motion analysis process, first motion existence is detected, and then classification and recognition processes are performed. The motion existence is identified by the difference of values of Hu Moment Invariants of different image frames. MIRC is based on values of Hu Moment Invariant and the Euclidean Distances between these values of same image frames as well as other image frames. Performance of MIRC is compared with other classification methodologies using temporal template motion detection approach in terms of success rate, time and speed.

The classification techniques which are used in this research thesis are: Mahalanobis Distance (MD) classifier, Quadratic Bayes Gaussian (QBG) classifier, Linear Bayes Gaussian (LBG) classifier, K-Nearest Neighbor (KNN) classifier with K = 1,3, 5 and Fuzzy K-Nearest Neighbor (FKNN) classifier with K = 1,3,5.

The developed Human Motion Analysis System is designed for Windows based environment and the coding of the system is written in computing language "MATLAB 7.0". The system input is in the form of .avi file.

1.5 Goals and Objectives

The main goals and objectives under which this research work has been conducted are:

- The primary objective of the research is to explore comparative study and analysis of computer vision and digital image processing techniques that are suitable for human motion detection, recognition and classification in video surveillance applications.
- The secondary objective is to design and develop the algorithms for the recognition and classification of human motion. Two new classification approaches, named, Trained Table based Recognition & Classification (TTRC) and Moment Invariant based Recognition and Classification (MIRC), have been proposed. A system will be developed that will test the performance of the above-mentioned classifiers.
- The designed system may be helpful to investigate the detailed analysis of the obtained results in terms of accuracy, time, memory capacity and success rate.

1.6 Thesis Organization

There are six chapters in this report. The chapter details and organization of the dissertation are as follows:

- Chapter 2 introduces the previous work done in this area, relevant methodologies and gives an in-depth review of the human motion detection, recognition and tracking approaches.
- Chapter 3 deals with the formal definition of the problem, solutions to the task specific problems, detailed exposition of methodologies and classification criterion to classify the testing dataset of the Unit-I, Unit-II and Unit-III. Technique and methodology of the new proposed algorithms TTRC and MIRC have also been described in detail in this chapter.
- Chapter 4 illustrates the experimental results of the Unit-I, Unit-II and Unit-III in tabular as well as in graphical format.
- Chapter 5 explains an in-depth analysis of the results obtained during the experiments in terms of accuracy, time, memory capacity and success rate.
- Chapter 6 summarizes the research and highlights the conclusion of the research and discusses some possible extensions.

1.7 Chapter Summary

This Chapter covers the broader aspects and features of the research work. It has presented purpose and general motivation behind the selection of this research area. It has explained the problem statement and then possible solution. It also gives a bird's eye view of the research methodology. It has highlighted objectives and goals of the research to be completed and achieved. In the end, organization of this report is presented as a guide for those who wish to study some particular topic and sections of the report.

2

Literature Review

This research work focuses on low level vision which deals with segmentation, thresholding, detection and classification, and high level vision that is related to action recognition and behavior understanding according to the framework of human motion analysis as discussed in section 1.2 and Figure 1.2. This chapter gives a bird's eye view of various approaches that address to low level vision as well as high level vision of human motion analysis. This chapter also recapitulates the previous work relevant to human motion analysis.

2.1 Low Level Vision…Motion Detection

Prior to the processes of human motion recognition and classification, the system needs to detect a human first and then initialize a model of the human. Human Detection involves the process of segmentation and thresholding in which human body is extracted from the background and optionally, detecting various body-parts and body-joints, and the process of objects classification. After detection of the existence of human beings then next step is the exposure of Human motion and ultimately understanding & analysis of motion. Human motion analysis & behavior understanding are the elements of high level vision.

2.1.1 Segmentation

Segmentation is the process that subdivides an image into its regions. Segmentation accuracy determines how efficiently the analysis of the system can be made. In other words, it explores the eventual success and failure of the system [7, 8]. Several approaches helpful in the process of motion segmentation can be described as under:

2.1.1.1 Background Subtraction Approaches

Background Subtraction is the most common approach for identifying and discriminating a moving object from the rest of the scene. It is the process that deals with identification of objects in the image that is relevant to a human motion. It's a very difficult task to identify the objects related to foreground motion from the background objects. In terms of intensity and color, some part of the background can copy some components of the moving human body [8].

Background subtraction approach is very successful when there is a static background and fixed stationary camera. However, it is very sensitive to the changing in background, for example, lightening effect [4].

Background detection and subtraction approaches can be classified into two types: non-adaptive methods and adaptive methods.

- *Non-adaptive Methods:* Pixel-by-Pixel approach and Mean Value Search algorithms are the examples of non-adaptive methods. Adaptive and continuously changing background (moving objects, flying birds, swaying trees, changing environment etc.) make it difficult for the human motion analysis system to converge on the moving human body and finally extract it from the background. In this way, imaging process is inherently noisy, creating another challenge [9].
- *Adaptive Methods:* In case of Adaptive methods, examples are: Kalman Filtering, averaging images over time, alpha-blending [10, 11], Gaussian Mixture Models

(GMM) etc. Kalman filtering method is based on Guassian Distribution [4]. Averaging images and alpha blending are straightforward and speedy methods, but they are not so efficient for scenes with many moving objects particularly if they move slowly [10, 11]. Haritaoglu et al [10]-[12] determine the minimum per-frame change and maximum per-frame change in intensity level at each pixel to find out a pixel of the background or foreground. In other words, background scene is modeled by representing each pixel by three values; minimum intensity, maximum intensity, and the maximum intensity difference between consecutive frames during the process of training period. These values are anticipated over several frames and are updated for background regions periodically. Stauffer and Grimson in their research paper "Learning Patterns of Activity Using Real-Time Tracking" [13] model pixel intensity from a mixture of adaptive Gaussians. A mixture of Gaussians is used to solve the problem of adaptive background variations and to capture multimodality (e.g. moving and sparkling screens, variation on water surfaces etc.).

Pfinder ('Person Finder') system [14] is another approach that deals with mean and covariance matrices at each pixel and update background recursively as the scene changes. The Pfinder system detects and tracks the head and hands of the human body using flesh-color as a prior.

Another hurdle in background subtraction process is the shadow of the objects. Shadows may cause serious problems such as color distortion, merging of objects, images histogram, shape warp and deformations, misidentifications [15]. There are a lot of approaches to find the solution of shadows. Some approaches prefer and works on the HSV (Hue, Saturation, and Value) color space analysis. HSV is one of several color systems that refer approximately to tint, shade, and tone [7]-[8]. These approaches select HSV because a shadow on a background does not change its hue significantly [16]. Other approaches employ saturation information in view of that shadows often lower the saturation of the pixels. In [17], a pixel of image is categorized into one of the four categories, background, foreground, shadow, highlight, depending on the distortion of the brightness and chrominance of the difference.

2.1.1.2 Temporal Differencing Approach

Temporal Differencing [18, 19] is the technique for detecting moving areas, which makes use of pixel-by-pixel difference between two or more than two successive frames. This method works well only if there is a very small motion and it detects just the skeleton of the moving objects if they display a little intensity variation [9]. Lipton et al. [20] make use of temporal differencing for the detection of moving objects. A development of the basic two-frame differencing modeling is the use of three consecutive frames [18]. The Visual Surveillance and Monitoring (VSAM) at Carnegie Mellon University have combined both background and temporal differencing to detect moving objects in a relatively static background. This hybrid algorithm is speedy and very efficient for detecting moving objects from the image sequences [19].

2.1.1.3 Optical Flow

Optical flow can be defined as: it is the distribution of velocities of movement of brightness patterns in an image [21]. Rowley and Rehg contribute to add kinematics motion constraint to every pixel in an image. They combine segmentation approach with Expectation Maximization (EM) computation [22]. Ju et al [23] illustrate a 2D methodology where main body parts are approximated as planar patches connected to each other at joints. Bregler and Malik [24] elaborate 3D kinematics model of the body and make use of optical flow approach to track the human body.

2.1.2 Thresholding Techniques

Thresholding is the process to find out the threshold level at which foreground pixels are successfully distinguished from background pixels. Thresholding is an effective way to separate foreground pixels from background pixels. Sezgin [25] has categorized thresholding techniques into following groups.

1. *Histogram shape based methods* in which threshold level is found out at the mixture of two Gaussian distribution associated with the background pixels and foreground pixels, for example, peak and valley thresholding method [26].
2. *Cluster based methods*, where grey level pixels are clustered into two classes (background and foreground) or alternatively distributed into a mixture of two Gaussians, for example, iterative thresholding (Ridler [28] and Otsu [27]).
3. *Entropy based methods* find out the difference in entropy between foreground and background regions (Kapur [29]).
4. *Object attribute based methods* which measure the similarity between the grey level and the binarized images (Hertz [30]).
5. *Spatial methods* make use of higher order probability distribution and correlation between pixels. (Abutaleb [31])
6. *Local methods* which compute threshold level at each pixel based on properties of the local images like local contrast method. (White [32])

2.1.3 Human Motion Detection

Human motion or action detection is the process of identifying human body/part motion and the human motion as well. Recently the approaches relevant to human motion detection can be divided into several groups.

One of schools of thought is based on explicit structural model [33]-[35]. They achieve motion recognition and understanding by recognizing sequence of static patterns. Other schools of thought try to achieve the process of recognition of the motion directly from the sequence of images [36]-[37]. Polana and Nelson recognize the motion using the spatiotemporal template matching process [37]. Davis used view-based approach, Motion History Image (MHI) and Motion Energy Image (MEI), in order to recognize the motion [36].

Collins et al [19] describes a Linear Discriminant Analysis (LDA) [38] based classifier and a view dependent neural network based classifier to differentiate between object's classes; human, human clutter and vehicle. In addition to these shape based classification methods, motion based methods have also been illustrated.

C. Cedras and M. Shah [39] present two steps in the process of motion-based recognition. The first step involves the extraction of motion while in the second step some testing dataset is matched with a conserved model of known actions. At the end, they perform the comparison for the purpose of classification. Mohan et al in [40] use Haar Wavelets and Support Vector Machines (SVM) in order to detect body parts of the human beings, for example, head, arm and legs and locations from the back view, front view and side view as well as existence of human motion. They illustrate the approach in two steps: the first step engrosses the testing small windows throughout the image for the existence of each body part while the second step combines the results to find out a human or human motion in the image.

D.M. Gavrila [41] has presented a report on visual analysis of human movement and the recent development in this field as well. The report includes only the discussion of human movement regarding whole body or hand motion, but there is no any description about the movement of human faces. Lipton, in [42], performs classification between human body and a rigid solid thing by using the heuristic that the residual flow will be higher for a human body than for a rigid body, due to its articulation.

Haritaoglu et al [10]-[12] describe a system for the purpose of labeling a human body silhouette by applying a chain of heuristics on recursive convex hull computations on the body silhouette. Rosales and Sclaroff in their paper "Inferring Body Pose without Tracking Body Parts" [43] define a neural-networks based methodology to map a 2D silhouette to a set of 2D joint locations. They use a binary image for a 2D silhouette, i.e., inside the silhouette the intensity is 1, outside the silhouette the intensity is 0. The silhouette is then represented by Hu moments which are computed from moments of inertia of a silhouette image. Hu moments are invariant to translation, orientation, rotation and scaling. A neural network is qualified for mapping Hu moments from the silhouettes of each bunch to the known 2D positions.

Q.Cai and J.K. Agarwal [44] proposed a framework of tracking the human motion in an outdoor scene using Multivariate Gaussian Models. More than one camera has been used which are fixed and mounted in various locations. Multiple cameras provide more information for the detection of human body parts and joints. D.M. Gavrila and L Davis in their paper “3-D Model-Based Tracking of Humans in Action” [45] have developed a system to determine the 3D positions of various joints, using multiple calibrated cameras. Principal Components Analysis (PCA) is applied to the silhouette, which determines the head-torso axis in the image, from each view [9]. Grauman et al [46] who have presented a probabilistic technique for measuring the 3D coordinates of different joints of the body of a walking human from images captured from multiple calibrated cameras, using Bayesian approach.

2.2 High Level Vision … Motion Representation and Recognition

Human action recognition and behavior understanding are the high level visions. A human action can be described in terms of a starting pose P_s, an ending pose P_e, and a sequence of continuous modifications that take the body from pose P_s at time t=0 to pose P_e at time t=T [9]. Human action representation is basically tied to the representation of the pose P. The issues related to that process are the speed of the motion/action and the frame rate of the video. Unpredictability in human actions becomes another crucial factor in human motion analysis systems: different persons exhibit variation in performing actions or the same action executed multiple times by the same person.

G. Bradski and J. Davis in [47] describe a gesture recognition method for object motion analysis. This work is the extension of the previous work done by Davis and Bobick [36]. This gesture recognition method is useful for determining the recent pose of the object as well as for measuring the different motions performed by the object in a video clip. After segmenting and thresholding processes, the regions of motion have been allied to those parts of the object that are moving. Accordingly the Motion History Image (MHI) is designed for the purpose of encoding the actual time, which they named as the timed Motion History Image (tMHI).

Hidden Markov Models (HMMs) [48] are specifically used for analysis of time-varying data with spatiotemporal variability. Coupled HMMs (CHMM) [49] combines HMMs and allows alteration of one HMM to be conditioned on the states of other HMMs that are coupled to it. Neural Network (NN) [50] is another approach to analysis time-varying data. Hogg [51] designed a computer program named as "WALKER" which attempts to recover 3D structure of walking individuals. Rohr [52] uses the eigenvector line fitting to make the outline model of human image and then converts 2D projections into 3D human model.

Freeman and Roth [53] use orientation histogram methodology for the gesture recognition. A single histogram of image edge orientations of a user's hand was used to represent and recognize various static gestures of user's hand, with dynamic gestures formed by concatenating histograms of individual poses. Yacoob and Black [54] describe action as view-based temporal trajectories of the optical flow parameters of the thigh, arm, calf and foot. They mold and recover temporal variations in the actions by an 'affine' transformation of time, $t' = \alpha t + L$. Most of Bobick's [55] work concentrated on the trajectories of the Moving Light Display (MLD) of human joints. Their works on MLD help them in activity recognition and understanding.

2.2.1 View-based Motion Recognition

View-based action recognition is very sensitive to the angle or view of the concerned object. Change in view may show the quite different results. For example, the walking man as seen from the left, right, front and back views represents four completely different motions. Chomat and Crowley's approach [56] is sensitive to temporal and viewpoint variations, that make use of Gabor filters to confine the spatiotemporal dynamics of actions. With unknown image sequence, their system calculates the Gabor filter responses and then employs the Bayes rule to find out the action probabilities.

Bobick and Davis [1]-[3] propose the system for human motion recognition that is purely view-based approach. They perform background differencing for each image and create two cumulative static images from the images in the sequence: the Motion Energy Image (MEI) which is a binary cumulative motion image and the Motion History Image (MHI) which represents "how" the image is moving as opposed to "where". Hu moments Invariants are calculated from the MEIs and MHIs to represent the image's feature description. Davis and Bradski [1, 57] extend Motion Energy Image (MEI) as Motion Energy Volume (MEV) and Motion History Image (MHI) as Motion History Volume (MHV). The Binary Motion Energy Volume (MEV) can be defined as:

$$E_\tau(x,t) = \bigcup_{i=0}^{\tau-1} \Omega^D(x,t-i)$$

Similarly, Motion History Volume where each pixel's intensity shows temporal history of the motion at the 3D location is defined as:

$$H_\tau(x,t) = \begin{cases} \tau & \text{if } \Omega^D(x,t) = 1 \\ \max[0, H_\tau(x,t-1)-1] & \text{otherwise} \end{cases}$$

Motion History Volume (MHV) and Motion Energy Volume (MEV) have been used for gait recognition [58] as well as activity modeling [59].

Sharma et al [60] illustrate the performance of view based approach using artificial neural network. The Hu Moments are used as the feature descriptor. Using Multi Layer Perceptron (MLP) based on back propagation, the performance of the system has been evaluated in real time.

2.2.2 View-Invariant Motion Recognition

The view-sensitive approaches have largely ignored the viewpoint-invariance issue, especially when the input is monocular video and the camera parameters are unknown. Even then, some prior work has addressed the viewpoint-invariance issue directly. Seitz and Dyer in [61] illustrate a methodology that detects cyclic motion and it is affine invariant, based on the assumption that feature association between successive frames is known. Their objective of cyclic motion detection is to identify and detect the repeating period of an action by evaluation between two images for resemblance.

Canton-Ferrer et al [62] perform view-independent 3D gesture recognition in low resolution sequence from multiple cameras, which is real time and robust to environmental condition. This is based on Motion History Volume (MHV) and Motion Energy Volume (MEV). They explore a complementary approach in which firstly data fusion is performed and then 3D motion description features (3D moment invariants [63]) are computed for the purpose of motion classification.

Campbell et al in [64] describe the process of extracting view invariant features for 3D gesture recognition. They work on the 3D coordinates of the two hands directly. A survey regarding visual analysis of human movement and the recent development in this area has also been illustrated in [41]. Only the body or hand motion has been highlighted, but not included the survey regarding the work on human faces.

2.3 Human Pose Representation and Recognition

A human body 'pose' means the configuration of the various body parts in a body-centric coordinate system. Quantitatively, pose is defined as the positions and orientations of the various body parts [9]. For human action recognition, the manner under which body poses, determines the extent of applicability of an approach. There are several approaches for view-dependent representations of a single body pose; these approaches are cumbersome. A view-invariant body-pose representation provides the desired cost-cutting measure and sophistication.

In [65] Barron and Kakadiaris estimate anthropometry and pose from a single un-calibrated image by using the scaled orthographic camera assumption and user supplied image joint locations. Bregler and Malik's [24] have an initialization step to find out where the joint angles of the body are calculated by using the scaled orthographic projection approximation and report a search based procedure to recover the joint angles as well as body lengths.

Fujiyoshi and Lipton [66] proposed the ‘star-skeleton’ approach. From the ‘star-skeleton’, two types of motion indication are determined: body posture and cyclic motion of skeleton segment. These indications are used to find out human activity such as walking or running. The ‘star-skeleton’ figure is obtained by connecting the centroid to the local maxima of the boundary distance from the centroid (see Figure 2.1). The bottom maxima are taken to be the feet and the top maximum is taken to be the head by assuming that the human is in the upright posture.

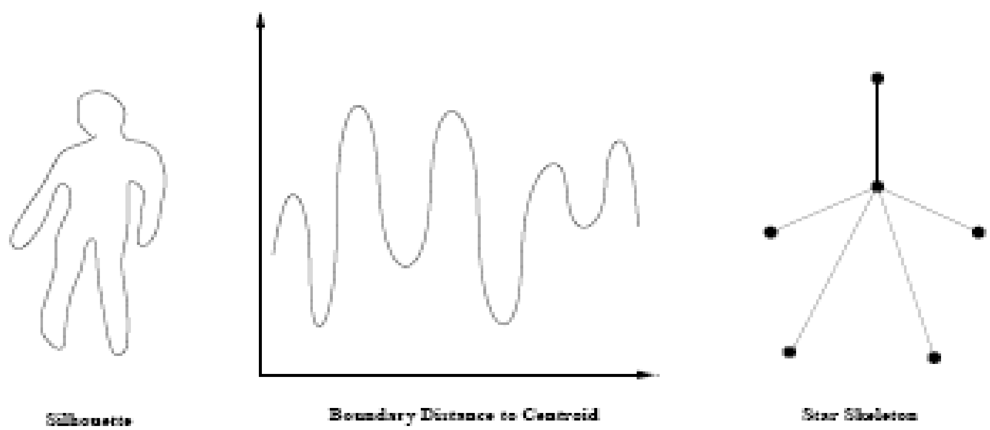


Figure 2.1: Fujiyoshi and Lipton’s Star-Skeleton Model [66]

It is conceivable that individual body parts can be tracked and labeled, e.g., [12, 67] but human pose is much hard to be estimated. Ali and Aggarwal in [68] present a method for automatic segmentation and recognition of continuous human activity. They make use of a skeleton-based representation of pose for action recognition from the side-view. They determine the angles subtended by three major components of the body: with the vertical axis, the upper part of the leg and the lower part of the leg.

2.4 Chapter Summary

This chapter presents recent development in human motion analysis and work done relevant to the human action detection, recognition and classification, during last 20 years. Various image processing techniques and the framework of different approaches that have been developed in this area of computer vision are discussed in concise in this chapter. This is a thorough and comprehensive collection of researches and approximately covers most of major techniques used in the field of human motion analysis.

3

Methodology

This chapter explores main aspects of recognition & classification problems and solutions in human motion analysis system. First of all, problem statement is illustrated and then it is followed by the methodologies of different approaches used in this research work.

3.1 Introduction

This research work is decomposed into three Units; Unit-I, Unit-II and Unit-III. The details of these Units are described one by one as under.

Unit-I deals with temporal template [1]-[3] methodology. Using this methodology, the effects of segmentation & thresholding, on recognition and classification with background subtraction approach and without background subtraction approach are investigated.

In Unit-II, evaluation of training datasets has been done. This evaluation is based on the experiments conducted at different sizes of training datasets. The behavior of the system is checked at these sizes of training datasets. Then a new proposed approach

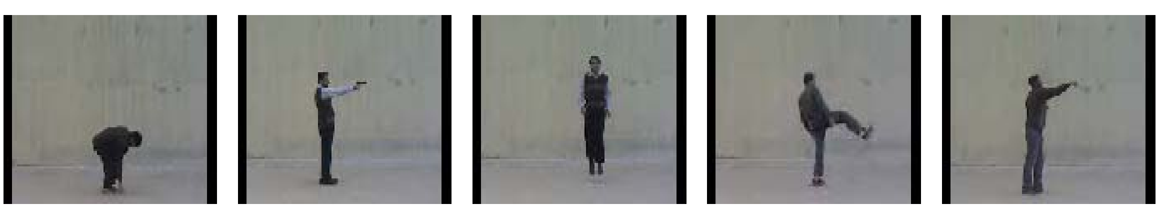
Trained Table based Recognition and Classification (TTRC) as well is discussed in detail in Unit-II.

At the end, in Unit-III, a recognition and classification approach, Moment Invariant based Recognition and Classification (MIRC) has been proposed. The overall methodology of MIRC has been discussed in detail in this Unit-III.

3.2 Problem Definition

In analysis process of a Visual Surveillance System, the input data to the system is usually in the form of a stream of video frames extracted from video from the surveillance camera. These extracted frames contain information of various types of human motions performed within the field of view (FoV) of the surveillance camera [9]. The main goal of the research work is to determine which type of motion has been performed in the FoV and then recognize and classify these types of motions. For that purpose, different methodologies involved in this process are evaluated and some new concepts and ideas in this regard have been proposed.

Five different types of human motions and actions are performed and recorded for this research work. These motions and actions are: Bending down, Gun Shot, Jumping up, Kicking front and Punching forward as shown in Figure 3.1. From these human motions, training datasets as well as testing datasets have been developed. Both training datasets and testing datasets are very different from each other so that ambiguity in results can be avoided.



Bending down   Gun Shot   Jumping up   Kicking front   Punching forward

Figure 3.1: Five different types of human motion

Unit-I

3.3 Methodology of Unit-I

Unit-I deals with segmentation and thresholding effects on human motion analysis. These effects of segmentation and thresholding are checked and investigated with Background Subtraction (BGS) approach and without Background Subtraction (BGS) approach. Selection of appropriate threshold level in recognition and classification is a crucial factor in evaluating performance of classification technique. Importance of a suitable threshold level in performance of classification techniques is quite obvious from the Experiment 1 in which threshold levels are manually selected using Trial & Error Thresholding technique (from 20 to 100 threshold level with the difference of value 5) as shown in Figure 3.2. At suitable and appropriate threshold level, classifiers give best performance. For example, at threshold levels 50 to 60 each classifier performs best as in Figure 3.2.

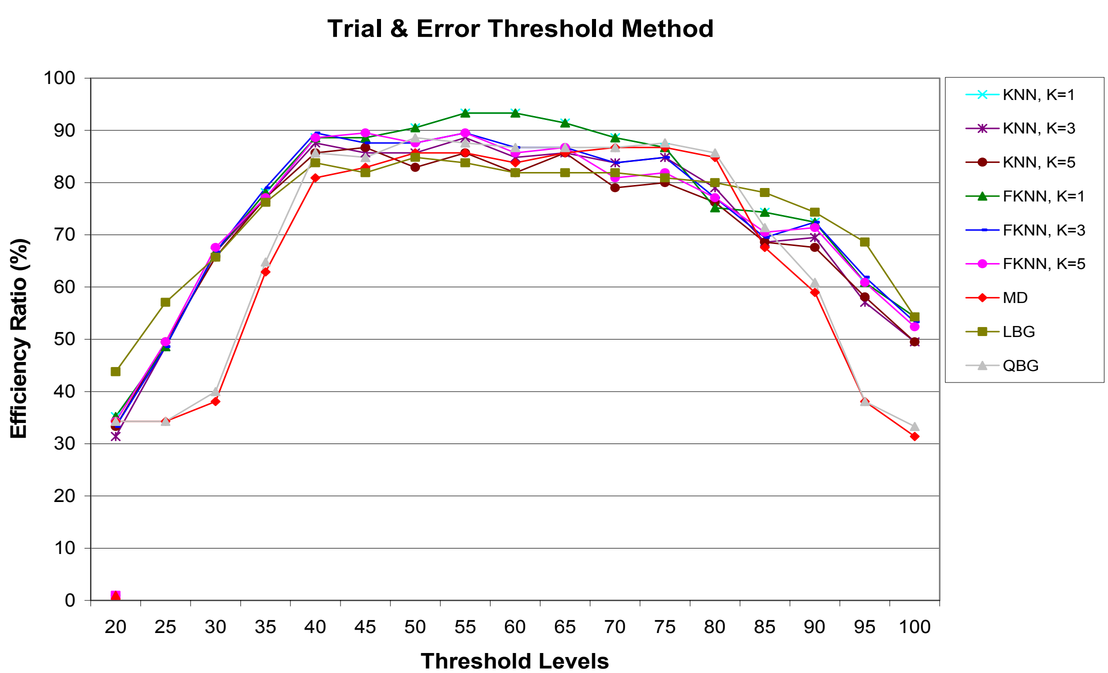


Figure 3.2: Classifiers' performances with "Trail & Error Threshold" method

Unit-I consists of two modules: Module-I and Module-II. In Module-I, behavior of different classification methodologies have been investigated under different dynamic thresholding techniques which are: Otsu threshold method, Isodata threshold method, Image Histogram shape based threshold method and Maximum Entropy based threshold method. These dynamic thresholding techniques are used to find out threshold levels of the static images created without Background Subtraction (BGS) approach, while in Module-II the threshold levels are computed from static images with Background Subtraction (BGS) approach. At the end, efficiency and performance of different classifiers is compared at different dynamic thresholding techniques with Background Subtraction approach and without Background Subtraction approach in terms of accuracy, time and success rate. Unit-I also investigates the performance of different classifiers under Background Subtraction approach as well as without Background Subtraction approach.

Unit-I includes several different steps. These steps are fashioned in such a manner that the output of every step becomes the input for the next coming step. However, the primary input of the system is the raw video signal from the Video Surveillance Camera. Ultimate output of the system is in the form of recognition as well as the classification of human motion existing in raw video clips. The steps involved in Unit-I are: Frame Extraction, Background Subtraction, Segmentation & Thresholding, Hu Moment Invariants, and Motion Recognition & Classification. These steps are discussed in detail as follows.

3.3.1 Frames Extraction

In first step of Unit-I, static image frames are created from '.avi (Audio Video Interleave)' files at the rate of 10 frames per second (fps). This process has been performed for each collected video clip. In this way, complexity of the problem has been minimized because video data has been transferred into static image frames. For example, 41 frames extracted from a video clip of Bending down motion have been shown in Figure 3.3.

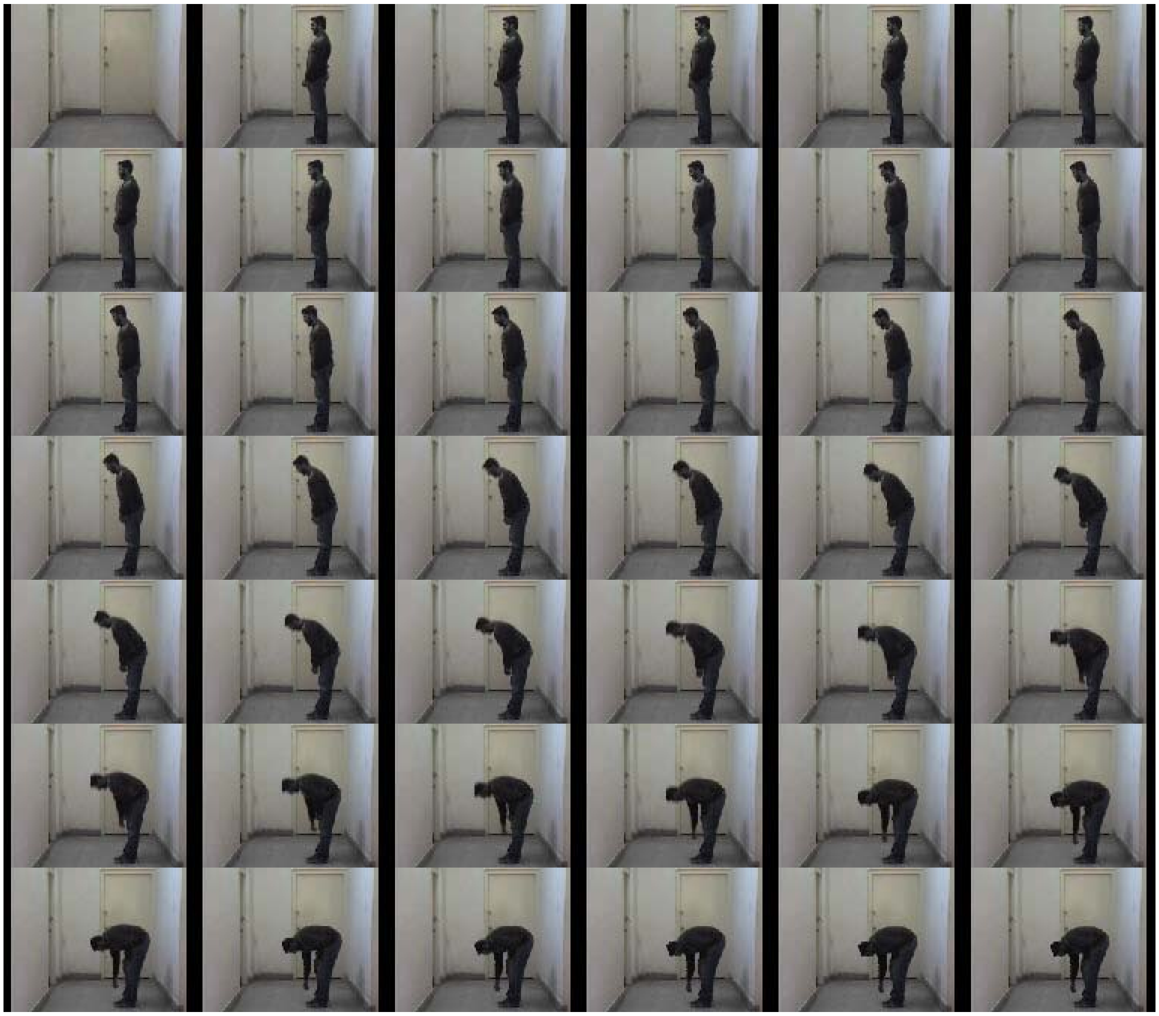


Figure 3.3: Frames extracted from "Bending down" motion video clip

3.3.2 Background Subtraction

In Module-I, threshold levels are calculated from the images without Background Subtraction (BGS) approach. Background Subtraction (BGS) process is used in Module-II. In other words, Threshold levels are calculated from the static images after Background Subtraction (BGS) approach.

Background Subtraction is used to differentiate between foreground and background pixels. This approach is particularly popular for motion segmentation and thresholding process in the condition where static & known background is available. Using pixel-by-pixel background subtraction approach [69, 70], known background image (see Figure 3.4 (a)) is subtracted from each created image frame before thresholding (see Figure 3.4 (c)). Pixels that show some change represent moving subject, and static pixels represent background. For example, thresholded images of last static image of Bending down video clip without Background Subtraction (see Figure 3.4 (d)) and with Background Subtraction (see Figure 3.4 (e)) using Otsu threshold method is shown in Figure 3.4.

(a) (b) (c)

(d) (e)

Figure 3.4: "Bending down" motion: (a) Known background image (b) Last static image frame (c) Image frame after Background Subtraction (d) Thresholded image of "b" (e) Thresholded image of "c"

3.3.3 Segmentation & Thresholding

Segmentation & Thresholding process divides gray-level image into its component regions. It separates foreground points $(f(x, y) > T)$ from background points $(f(x, y) < T)$. A thresholded image $f'(x, y)$ is [7, 8]:

$$f'(x, y) = \begin{cases} 1 & f(x, y) > T \\ 0 & f(x, y) \leq T \end{cases} \tag{3.1}$$

Thresholding approaches used in this Unit-I are:

- Otsu's Threshold Method
- Isodata Threshold Method
- Image Histogram Shape based Threshold Method
- Maximum Entropy based Threshold Method

3.3.3.1 Otsu's Threshold Method

Otsu's threshold method [27] has been used for thresholding, which is a histogram-based threshold approach. It finds threshold level that maximizes between-class variance σ^2_B, which can be defined as [7]:

$$\sigma_B^2 = \omega_0 (\mu_0 - \mu_T)^2 + \omega_1 (\mu_1 - \mu_T)^2 \tag{3.2}$$

Otsu also defines within-class variance σ_w^2 and total-class variance σ_T^2 as [27]:

$$\sigma_w^2 = \omega_o \sigma_o^2 + \omega_b \sigma_b^2 \tag{3.3}$$

$$\sigma_T^2 = \sum_{i=0}^{l-1} (i - \mu_T)^2 p(i) \tag{3.4}$$

$$\mu_T = \omega_o \mu_o + \omega_b \mu_b \tag{3.5}$$

Where ω_o and ω_b are the weights relevant to the object and background classes respectively. μ_o and μ_b are the mean values of the object and background classes. $p(i)$ is the histogram density function. The result of Otsu threshold method with background subtraction is shown in Figure 3.5 where a person is performing a "Gun Shot" motion.

(a) (b) (c)

Figure 3.5: "Gun Shot" motion: (a) Last static image frame (b) Image frame after Background Subtraction (c) Image frame after "Otsu's Threshold" method

3.3.3.2 Isodata Threshold Method

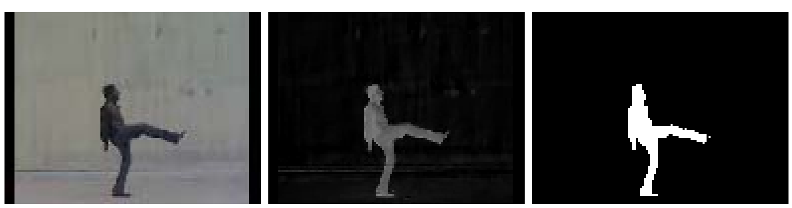
Isodata threshold method is used to compute global threshold level by iteration technique which is developed by Ridler and Calvard [28]. The steps involved in Isodata threshold method are:

- First threshold level is selected by segmenting the image histogram.
- Threshold the image at this specific level.
- Calculate the sample mean of foreground pixels (μ_1) and the sample mean of background pixels (μ_2). Comput the new threshold value by using these two sample means.

$$T = ½\,(\mu_1 + \mu_2) \tag{3.6}$$

- This process is repeated until there is no more change in threshold level and threshold value becomes $T^{(t+1)} = T^t$.

The output of Isodata threshold method with background subtraction approach is shown in Figure 3.6 where a person is performing a “Kicking front” motion.



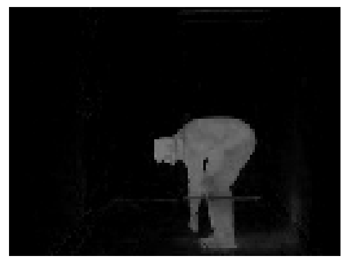
(a) (b) (c)

Figure 3.6: “Kicking front” motion: (a) Last static image frame (b) Image frame after Background Subtraction (c) Image frame after “Isodata Threshold” method

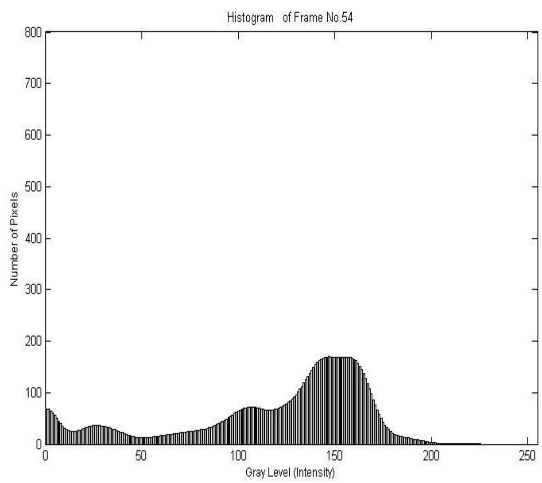
3.3.3.3 Image Histogram Shape based Threshold Method

This category of threshold methods computes threshold level based on the shape properties of the histogram [25, 26]. Using Image Histogram Shape based Threshold method, threshold level is calculated by finding the low valley location in image histogram. The threshold level would be somewhere between the first terminating valley and second initiating zero crossing valley. For example, in Bending down action video clip, image histograms of the last image frames without Background Subtraction (see Figure 3.7 (c)) and with Background Subtraction (see Figure 3.7 (d)) is elaborated in Figure 3.7.

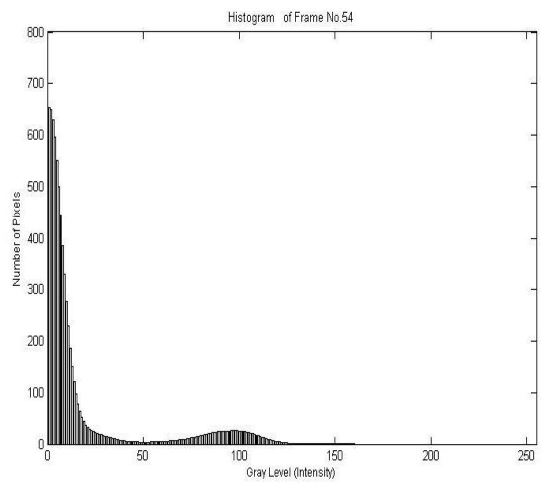
(a)



(b)



(d)

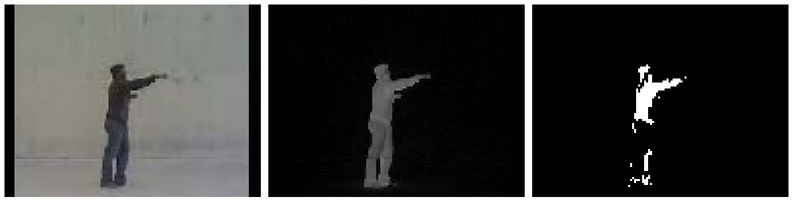


(e)

Figure 3.7: "Bending down" motion: (a) Last static image frame (b) Image frame after Background Subtraction (c) Histogram of "a" (d) Histogram of "b".

In a video clip of "Punching forward" motion, total 17 static image frames are obtained. After performing thresholding of last static image frame (Frame No. 17) using Image Histogram Shape based Threshold method with background subtraction approach, the obtained result is shown in Figure 3.8.

The histogram of this last static image frame through which threshold level is selected and the result obtained in Figure 3.8 is shown in Figure 3.9.



(a) (b) (c)

Figure 3.8: "Punching forward" motion: (a) Last static image frame (b) Image frame after Background Subtraction (c) Image frame after "Image Histogram Shape based Threshold" method

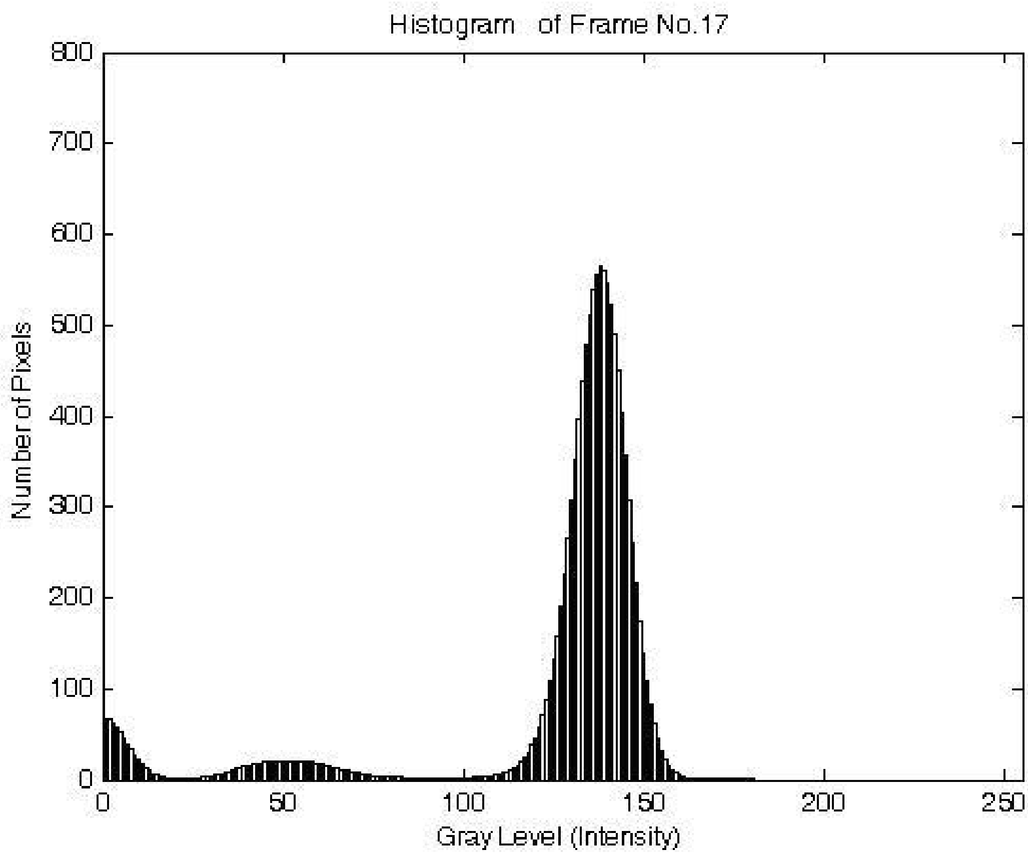


Figure 3.9: Histogram of last static image frame of "Punching forward" motion

3.3.3.4 Maximum Entropy based Threshold Method

In this method, maximum entropy in the grey-level in the image is calculated [71, 72] [29]. Entropy *H*, the average information per source output is:

$$H = -\sum_{i=1}^{L} P(e_i) \log p(e_i) \tag{3.7}$$

Where e_i are the random variables and $P(e_i)$ are the associated probabilities. Kapur et al. [29] define optimal threshold level maximizing the sum of two class entropies.

$$H = \max\left[-\sum_{i=o}^{Topt} p_i \log(p_i) - \sum_{i=Topt+1}^{255} p_i \log(p_i)\right] \tag{3.8}$$

Maximum Entropy based Threshold method with background subtraction is shown in Figure 3.10 where "Jumping up" motion exists in video clip.

(a) (b) (c)

Figure 3.10: "Jumping up" motion: (a) Last static image frame (b) Image frame after Background Subtraction (c) Image Frame after "Maximum Entropy based Threshold" method

3.3.4 Motion History Image (MHI) Formation

After creation of thresholded images, Motion History Image (MHI) has been designed from these static thresholded images. Motion History Image (MHI) is basically a view-specific approach in which only the action is described over time. Davis [1]-[3] defined Motion History Image (MHI) as: MHI is formed in order to represent how (as opposed to where) the image is moving. In other words, MHI represents "how" the motion exists. The strength of this technique is the use of a compact, expressive, concurrent and real-time representation of the sequence of motions in a single static image called Motion History Image (MHI).

In an MHI “*Hτ*”, pixel intensity is a function of the temporal history of motion at that point [2]. MHI “*Hτ*” can be represented as:

$$H\tau(x,y,t)=\begin{cases}\tau & \text{if } D(x,y,t)=1\\ \max(0,H\tau(x,y,t-1)-1) & \text{otherwise}\end{cases} \tag{3.9}$$

Brightness of pixel in Motion History Image is proportional to “how” recently change in intensity level has occurred as shown in Figure 3.11. Higher intensity level of pixels shows the recent part of occurrence of motion.

Davis also defined Motion Energy Image (MEI) [1]-[3] as cumulative motion image. MEI points out where the motion occurred (as opposed to how the motion exists). In case of MEI, time dimension has been fully cut out. Motion Energy Image can be produced when MHI is thresholded above zero. It is the binary version of Motion History Image (MHI). MEI “*Eτ*” can be represented as:

$$E\tau(x,y,t) = \bigcup_{i=0}^{\tau-1} D(x,y,t \quad -1) \tag{3.10}$$

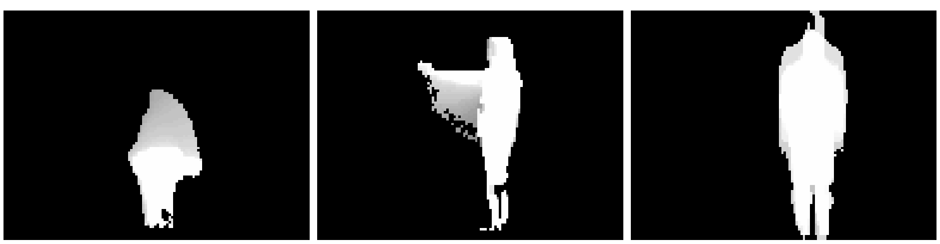
Where *D(x,y,t)* is a binary image sequence and τ is the duration which is very crucial in defining the temporal extent of a movement.

Motion Energy Image (MEI) drops the temporal features which are very important in analyzing the motions but Motion History Image (MHI) takes it into account. So, Motion History Image (MHI) has been used in this research for human motion analysis instead of Motion Energy Image (MEI). Examples of Motion History Image (MHI) and Motion Energy Image (MEI) are shown in Figure 3.11 where a person is performing a Bending down motion in the video clip.

(a) (b)

Figure 3.11: "Bending down" motion (a) Motion Energy Image (MEI) (b) Motion History Image (MHI)

Motion History Image (MHI) is created for each type of action in both categories without Background Subtraction approach as well as with Background Subtraction approach. For example, without Background Subtraction approach, Motion History Images are created for these five types of actions using Isodata thresholding method as shown in Figure 3.12. Similarly with Background Subtraction approach, Motion History Images for these five types of actions using Isodata threshold technique are shown in Figure 3.13.



(a) (b) (c)

(d) (e)

Figure 3.12: Motion History Images created using "Isodata Threshold" method and *without* Background Subtraction approach (a) Bending down (b) Gun Shot (c) Jumping up (d) Kicking front (e) Punching forward

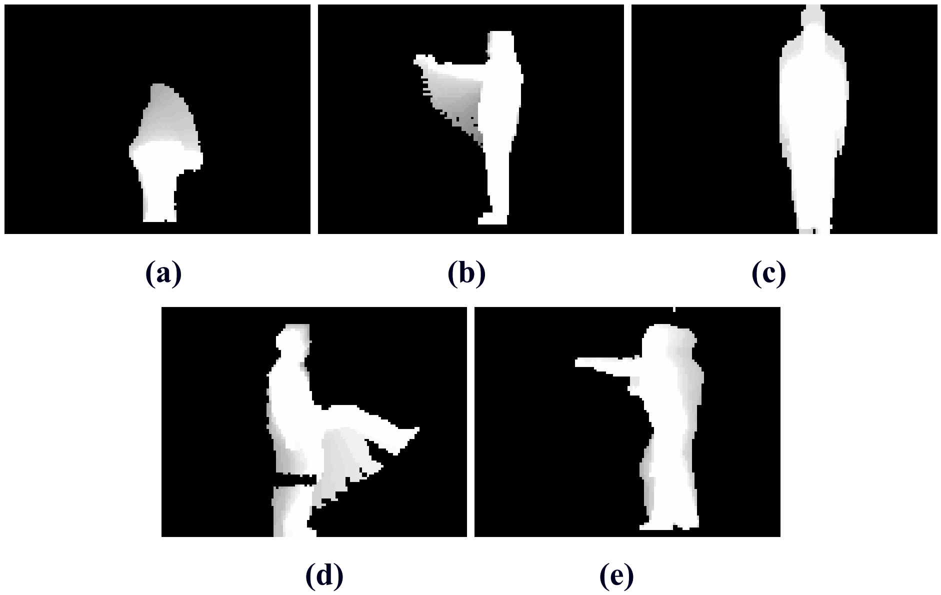


Figure 3.13: Motion History Images created using "Isodata Threshold" method and *with* Background Subtraction (a) Bending down (b) Gun Shot (c) Jumping up (d) Kicking front (e) Punching forward

3.3.5 Hu Moment Invariants

Nixon [73] defined Moments as an image shape's layout (pixels' arrangement of an image). It represents high-level description of the image for example area, density, compactness, irregularity etc. Image moments are the example of statistical measures for object recognition and certain particular weighted averages of the image pixels' intensities. Image moments are the global narration of shape of the image. The statistical second-order moment describes the rate of change in a shape's area. Moments are more associated with statistical pattern recognition than with model-based vision because of a major assumption which is: there is an unoccluded view of the target shape. Moment provides a global description with invariance characteristics and compact description in order to avoid noise.

There are two types of shape descriptors: counter-based shape descriptor and region-based shape descriptor [74]. Hu Moment Invariants are the examples of contour-based shape descriptor.

3.3.5.1 Raw Moments

The two-dimensional Cartesian moment is allied with an order that starts from zero up to higher orders. For a two-dimensional continuous function *f(x,y)*, the moment of order p and q is defined as:

$$M_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)dxdy \tag{3.11}$$

Where p, q = 0,1,2,... For discrete images, the Equation 3.11 becomes:

$$M_{pq} = \sum_{x}\sum_{y} x^p y^q I(x,y)\Delta A \tag{3.12}$$

This may be computed by considering the image as a probability density function (PDF), e.g., dividing Equation 3.12 by $\sum_{x}\sum_{y} I(x,y)$.

These descriptors have a uniqueness property [73]. A uniqueness theorem by Papoulis, states that if *f(x,y)* is piecewise continuous and has nonzero values only in a finite part of the xy plane, moments of all orders exist, and the moment sequence (M_{pq}) is uniquely determined by *f(x,y)* [9]. Conversely, (M_{pq}) determines *f(x,y)* uniquely. These moments are the descriptors and can be used to reconstruct a shape. The zero-order moments M_{00} which represent the total mass of the function are:

$$M_{00} = \sum_{X}\sum_{Y} I(x,y)\Delta A \tag{3.13}$$

. The two first order moments, M_{01} and M_{10} are given as under:

$$M_{10} = \sum_{X}\sum_{Y} xI(x,y)\Delta A \tag{3.14}$$

$$M_{01} = \sum_{X}\sum_{Y} yI(x,y)\Delta A \tag{3.15}$$

In general, the centre of mass $(\bar{x}, \bar{y})$ can be calculated from the ratio of the first-order to the zero-order components [73] as

$$\bar{x} = \frac{M_{10}}{M_{00}} \quad , \quad \bar{y} = \frac{M_{01}}{M_{00}} \tag{3.16}$$

3.3.5.2 Central Moments

The zero-order *centralized* moment represents again the shape's area. Central moments can be defined as:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y)\, dxdy \tag{3.17}$$

Where: $\bar{x} = \frac{M_{10}}{M_{00}}$ and $\bar{y} = \frac{M_{01}}{M_{00}}$ are the components of the Centroid. If *f(x,y)* is a digital and discrete image, then the Equation 3.17 becomes:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \tag{3.18}$$

The central moments of order up to 3 are given as under:

$$\mu_{00} = M_{00} \tag{3.19}$$

$$\mu_{01} = 0 \tag{3.20}$$

$$\mu_{10} = 0 \tag{3.21}$$

$$\mu_{11} = M_{11} - \bar{x} M_{01} = M_{11} - \bar{y} M_{10} \tag{3.22}$$

$$\mu_{20} = M_{20} - \bar{x} M_{10} \tag{3.23}$$

$$\mu_{02} = M_{02} - \bar{y} M_{01} \tag{3.24}$$

$$\mu_{21} = M_{21} - 2\bar{x} M_{11} - \bar{y} M_{20} + 2\bar{x}^2 M_{01} \tag{3.25}$$

$$\mu_{12} = M_{12} - 2\bar{y}M_{11} - \bar{x}M_{02} + 2\bar{y}^{2}M_{10} \tag{3.26}$$

$$\mu_{30} = M_{30} - 3\bar{x}M_{20} + 2\bar{x}^{2}M_{10} \tag{3.27}$$

$$\mu_{03} = M_{03} - 3\bar{y}M_{02} + 2\bar{y}^{2}M_{01} \tag{3.28}$$

Neither of the first-order centralized moments has any descriptive capability because they both have zero values. The third Moment μ_3 measure the skewness of the histogram. If the measure is 0, histogram is symmetric, positive by histogram skewed to right and negative for histogram skewed to the left [7, 8]. It can be represented as:

$$\mu_{pq} = \sum_{m}^{p}\sum_{n}^{q}\binom{p}{m}\binom{q}{n}\left(-\bar{x}\right)^{(p-m)}\left(-\bar{y}\right)^{(q-n)}M_{mn} \tag{3.29}$$

Central moments are invariant to translational but not to rotation. By constructing a covariance matrix, information about image orientation can be derived with the help of second order central moments. The second order central moments are:

$$\mu'_{20} = \mu_{20}/\mu_{00} = M_{20}/M_{00} - \bar{x}^{2} \tag{3.30}$$

$$\mu'_{02} = \mu_{02}/\mu_{00} = M_{02}/M_{00} - \bar{y}^{2} \tag{3.31}$$

$$\mu'_{11} = \mu_{11}/\mu_{00} = M_{11}/M_{00} - \bar{x}\bar{y} \tag{3.32}$$

The covariance matrix of the image *I(x,y)*, created by using the second order central moments is given as:

$$\operatorname{cov}[I(x,y)] = \begin{bmatrix} \mu'_{20} & \mu'_{11} \\ \mu'_{11} & \mu'_{02} \end{bmatrix} \tag{3.33}$$

The eigenvectors of this matrix correspond to the major and minor axes of the image intensity, so the orientation can be extracted from the angle of the eigenvector associated with the largest eigenvalue [9]. This angle θ can be represented as:

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\mu'_{11}}{\mu'_{20} - \mu'_{02}}\right) \tag{3.34}$$

The eigenvalues of the covariance matrix can be defined as:

$$\lambda_i = \frac{\mu'_{20} + \mu'_{02}}{2} \pm \frac{\sqrt{4\mu'^2_{11} + (\mu'_{20} - \mu'_{02})^2}}{2} \tag{3.35}$$

These are proportional to the squared length of the eigenvector axes. Thus the relative difference in magnitude of eigenvalues is an indication of how stretched out the image is.

3.3.5.3 Scale Invariant Moments

In order to build up invariance to scale, *normalized central moments* are required. *Normalized central moments* can be defined as (Hu, [75]):

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{3.36}$$

Where $\gamma = \frac{p+q}{2} + 1 \qquad \forall p + q \geq 2$ (3.37)

3.3.5.4 Rotation Invariant Moments

Hu in [75] computed seven values from *normalized central moments*, which are invariant to translation, scale, mirroring, and rotation. In terms of *normalized central moments*, seven Hu Moments are:

$$\phi_1 = \eta_{20} + \eta_{02} \tag{3.38}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2 \tag{3.39}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{3.40}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{3.41}$$

$$\varphi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{3.42}$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})] \tag{3.43}$$

$$\varphi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{3.44}$$

The first two Hu Moment Invariants, φ_1 and φ_2, are second-order moments, for which $p + q = 2$. Remaining five Hu Moment Invariants are third-order moments, because $p + q = 3$. The first-order moments are of no outcome because they are zero [73].

First Hu Moment Invariant 'φ_1' is roughly proportional to moment of inertia around the image's centroid, if the pixels' intensities are illustrated as physical density. So first Hu Moment Invariant 'φ_1' can be used as direction independent feature in the distinction of objects. The last Hu Moment Invariant 'φ_7' is skew invariant, which enables it to distinguish mirror images of otherwise identical images. However, these invariant moments are not orthogonal. These seven Hu Moment Invariants form the basis of Classification. Some examples of calculated Hu Moment Invariants of all types of motions (Bending down, Gun Shot, Jumping up, Kicking front, Punching forward) are shown in Table 3.1.

Table 3.1: Some examples of Hu Moment Invariants

Motion Types	**Seven values of Hu Moment Invariants**						
	φ_1	**φ_2**	**φ_3**	**φ_4**	**φ_5**	**φ_6**	**φ_7**
Bending down	1.5044	3.8909	8.0055	9.0722	17.6536	11.0177	18.8647
Gun Shot	1.1046	2.9100	4.2105	5.9017	12.1141	8.2181	11.0099
Jumping up	1.3986	3.4193	7.9524	8.4125	16.5953	10.1223	20.2222
Kicking front	1.3950	4.0373	5.6712	10.9245	23.5363	12.9662	19.2224
Punching forward	1.0240	2.3740	5.3008	6.3076	12.2072	7.6905	12.9869

3.3.6 Training & Testing Datasets Formation

Video clips of all types of actions are divided into two sets, one for training dataset and the other for testing dataset. For example, in 56 clips of Bending down action 35 video clips are used as training and 21 video clips are used as testing. After the completion of the process of calculating Hu Moment Invariants from Motion History Images (MHIs) of all types of actions (video clips), a training dataset has been developed from these Hu Moment Invariants. Similarly the calculated seven-value set of Hu Moment Invariants from selected testing video clips behaves like a testing dataset. Each testing dataset is then compared with training dataset using different classification approaches and where the closest distance is found, the testing data will lie in that specified group of training dataset. The training datasets and testing datasets for the Unit-I are shown in Table 3.2.

Table 3.2: Training & Testing Datasets' sizes

Motion Types	Training Dataset	Testing Dataset
Bending Down	35	21
Gun Shot	35	21
Jumping Up	35	21
Kicking Front	35	21
Punching Forward	45	21
Total Video Clips	**185**	**105**

3.3.7 Recognition & Classification

For recognition and classification, classifiers used in this Unit-I of research work are: Mahalanobis Distance (MD) classifier, Quadratic Bayes Gaussian (QBG) classifier, Linear Bayes Gaussian (LBG) classifier, K-Nearest Neighbor (KNN) classifier with k = 1, 3 and 5 and Fuzzy K-Nearest Neighbor (FKNN) classifier with k = 1, 3 and 5. These classification techniques are discussed in detail as follows:

3. 3.7.1 Mahalanobis Distance (MD) Classifier

Mahalanobis Distance (MD) is also known as statistical distance. MD can be defined as: it is the calculation of the distance between a vector y and the mean m_x of a vector population, multiplied by the inverse of the covariance matrix, "C_x" of that population [7]. In mathematical format, it becomes:

$$d(\mathbf{y},\mathbf{m}_\mathbf{x}) = (\mathbf{y}-\mathbf{m}_\mathbf{x})^T \mathbf{C}_\mathbf{x}^{-1}(\mathbf{y}-\mathbf{m}_\mathbf{x}) \tag{3.45}$$

Where mean vector m_x of a vector population is

$$m_x = \frac{1}{K}\sum_{k=1}^{K} x_k \tag{3.46}$$

and $n \times n$ covariance matrix, C_x of the population is

$$\mathbf{C}_\mathbf{x} = \frac{1}{K-1}\sum_{k=1}^{K}(\mathbf{x}_k-\mathbf{m}_\mathbf{x})(\mathbf{x}_k-\mathbf{m}_\mathbf{x})^T \tag{3.47}$$

The consumption time for original Mahalanobis Distance (MD) classifier is $O(n^2)$ for n-dimensional feature vectors [76]. In the implementation of Mahalanobis Distance (MD) classifier, the operation of inverse matrix is a time consuming process. This operation is optimized by MATLAB's matrix right division operator (/) [8].

3. 3.7.2 Quadratic Bayes Gaussian (QBG) Classifier

According to Bayes Theorem, the probability that an unknown sample is a member of ω_i given feature x is: the probability density function of pattern vectors with feature x being a member of class ω_i multiplied by the probability that a randomly selected pattern vectors will be a member of class ω_i. That is:

$$p(\omega_i / x) = \frac{p(x/\omega_i)\times P(\omega i)}{p(x)} \tag{3.48}$$

Where;

$$p(x) = \sum_{j=1}^{\omega} p(x/\omega_j) \tag{3.49}$$

The use of Bayes rule directly is unpractical because to calculate $P(\mathbf{x} \mid \omega_i)$ need so much data to get the relative frequencies of each groups for each measurement. It is more realistic approach to assume the distribution and then get the probability theoretically [9].

If there is multivariate normal distribution for each group with covariance estimates stratified by group, then Quadratic Bayes Gaussian (QBG) classifier formula can be given as:

$$f_i = \ln p(\omega_i) - \frac{1}{2}\ln |\mathbf{C}_i| - \frac{1}{2}(\mathbf{x}_k - \mu_i)^T \mathbf{C}_i^{-1}(\mathbf{x}_k - \mu_i) \tag{3.50}$$

Where C_i is covariance matrix, μ is the mean vector of pattern population of class ω_i and $| C_i |$ is determinant if covariance matrix C_i. Quadratic Bayes Gaussian (QBG) function measures the covariance matrix C_k, for each class k where $k = 1, 2, \ldots, K$. The decision functions in Equation 3.50 are hyperquadrics (quadratic functions in n-dimensional space), because no terms higher than the second degree in the components of x appear in the Equation 3.50 [7].

3. 3.7.3 Linear Bayes Gaussian (LBG) Classifier

Linear Bayes Gaussian (LBG) function can be extracted from Equation 3.50, if all covariance matrices become equal for $i = 1, 2, \ldots W$. In other words, for Linear Bayes Gaussian (LBG) function, each group has multivariate normal distribution and has the same covariance matrix. The linear decision functions are hyperplanes for $i = 1, 2, \ldots W$ and is expressed as:

$$f_i = \ln p(\omega_i) + \mu_i \mathbf{C}^{-1}\mathbf{x}_k^T - \frac{1}{2}\mu_i \mathbf{C}^{-1}\mu_i^T \tag{3.51}$$

3. 3.7.4 K-Nearest Neighbor (KNN) Classifier

K-Nearest Neighbor (KNN) classifier classifies an unknown sample by finding the distance to its nearest neighbors among a set of known samples. The distance metric used in KNN classifier is inappropriate as long as it applies constantly and consistently to all training dataset samples. The Euclidean Distance is used as a measure of closeness in this technique which is L2 Norm Form. The Euclidean distance between two *n*-dimensional vectors x and y is:

$$D_E(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} \tag{3.52}$$

3. 3.7.5 Fuzzy K-Nearest Neighbor (FKNN) Classifier

Fuzzy K-Nearest Neighbor (FKNN) classifier is based on Fuzzy Set theory, which represents how closely testing dataset represents each type of class in training dataset [77]. Fuzzy Set theory assigns each sample a value that shows "how closely it represents each given class in training dataset". In this way, it replaces the traditional hard set theory which is based on "is a testing sample a class member or not a class member?" x's membership in class i can be expressed as [77]:

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij} \left[\frac{1}{\|x - x_j\|^{\frac{2}{m-1}}} \right]}{\sum_{j=1}^{k} \left[\frac{1}{\|x - x_j\|^{\frac{2}{m-1}}} \right]} \tag{3.53}$$

Where:

$\sum_{j=1}^{k} u_{ij} \left[\frac{1}{\|x - x_j\|^{\frac{2}{m-1}}} \right]$ = represents an unknown testing dataset's membership in each class that is assigned its K-nearest known neighbors' memberships in those classes

$\sum_{j=1}^{k} \left[\frac{1}{\|x - x_j\|^{\frac{2}{m-1}}} \right]$ = is a function of the neighbors' distances from the unknown testing dataset.

Unit-II

3.4 Methodology of Unit-II

In Unit-II, Motion History Images (MHIs) has been developed from already extracted frames, after the thresholding and segmentation process, using temporal template methodology [1]-[3]. Hu Moment Invariants calculated from MHIs, are used to develop training dataset.

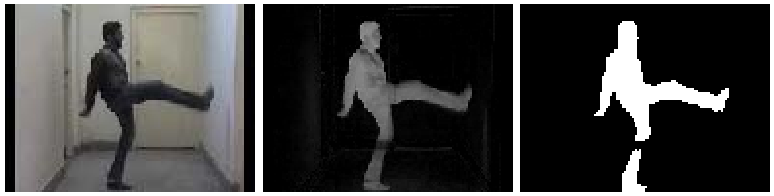
In the context of training dataset, Unit-II is divided into three modules; Module-I, Module-II and Module-III. Module-I deals with training dataset of 105x7 elements. In Module-II, a larger training dataset of 200x7 elements has been developed. On the basis of these training datasets, the process of recognition and classification has been made. While in Module-III, a new simple approach for recognition and classification has been proposed named Trained Table based Recognition & Classification (TTRC) [78]. TTRC is based on a simple trained data table. This table is trained on the basis of the features of Hu Moment Invariants as calculated from MHIs.

3.4.1 Module-I and Module-II

The algorithmic steps involved in Module-I and Module-II of Unit-II are similar to Unit-I and are illustrated as follows:

1. The input data in the format of 'avi file' are given to the system which is collected from a stream of video frames of surveillance camera.
2. Static image frames are created from 'avi file' video data at the rate of 10 frames per second.
3. Known background is subtracted from each frame using pixel-by-pixel background subtraction approach.
4. In 4^{th} step, Thresholding process separates foreground points $(f(x, y) > T)$ from background points $(f(x, y) < T)$ using Otsu's threshold method.

For example, in a video clip of "Kicking front" motion, 19 image frames are extracted at the rate of 10 fps. The image Frame No. 19, its background subtracted image and threshold image are shown in Figure 3.14.



(a) (b) (c)

Figure 3.14: "Kicking front" motion: (a) Image frame no. 19 (b) Image frame after Background Subtraction. (c) Image frame after "Otsu's Threshold" method

5. All these thresholded images are combined in single static image forming Motion History Image (MHI).
6. Seven φ values of Hu Moment Invariants" are computed from normalized central moments as described in detail in Unit-I. These Hu Moment Invariants are invariant to translation, mirroring, scaling and rotation.
7. Hu Moment Invariants computed from Motion History Image, are used to develop the training datasets and testing datasets.

The training dataset of Module-I consists of 105x7 elements. 21 video clips for each type of motions have been collected: Bending down, Gun Shot, Jumping up, Kicking front and Punching forward. After calculating the Hu Moment Invariants from MHIs created from these video clips, a training dataset of 105x7 elements has been developed. Similarly, testing dataset consists of total 90 video clips. Training and Testing datasets for Module-I is given in detail in Table 3.3.

Training datasets of Module-II consists of 200x7 elements. 40 video clips for each type of motion have been collected (total 200 video clips). After calculating Hu Moment Invariants, training datasets of 200x7 elements are developed. Testing dataset remains the same as shown in Table 3.4.

Table 3.3: Training & Testing Datasets for Module-I

Motion Types	Training Data	Testing Data
Bending down	21	16
Gun Shot	21	15
Jumping up	21	15
Kicking front	21	16
Punching forward	21	28
Total Video Clips	**105**	**90**

Table 3.4: Training & Testing Datasets for Module-II

Motion Types	Training Data	Testing Data
Bending down	40	16
Gun Shot	40	15
Jumping up	40	15
Kicking front	40	16
Punching forward	40	28
Total Video Clips	**200**	**90**

8. From these Hu Moment Invariants based training datasets and testing datasets, recognition and classification has been performed in terms of success rate, storage capacity and time & speed. For recognition and classification, classifiers used in this Unit-II are the same as used in Unit-I which are: Mahalanobis Distance (MD) classifier, Quadratic Bayes Gaussian (QBG) classifier, Linear Bayes Gaussian (LBG) classifier, K-Nearest Neighbor (KNN) classifier with k = 1, 3 and 5 and Fuzzy K-Nearest Neighbor (FKNN) classifier with k = 1, 3 and 5.

3.4.2 Module-III

In Module-III, The proposed approach, Trained Table based Recognition & Classification (TTRC) [78] has been explored which is based on a simple trained table.

A trained table of size 3x7 elements has been developed instead of using a large training dataset. This trained table has been created from the feature & behavior study of Hu Moment Invariants. Such values of Hu Moment Invariants are found for each type of motion that behaves distinctively. For example, a maximum and minimum value of φ_7 (7^{th} value of Hu Moment Invariants) for each type of motion has been calculated. Maximum value of φ_7 represents Bending down, Jumping up and Kicking front motions while minimum value of φ_7 represents Gun Shot and Punching forward motions. A table of such type of distinctive values is created which is called Trained Table (TT). On the basis of that Trained Table, the process of recognition and classification has been performed. In this way the motion types can be distinguished from each other and ultimately recognized and classified. The performance of TTRC is compared with other recognition and classification approaches (discussed in Unit-I) in terms of time & speed, storage capacity and accuracy & success rate. The steps involved in algorithm of Trained Table based Recognition & Classification (TTRC) are described as under [78]:

1. Input data as avi file to the system.
2. Create frames from this file at the rate of 10 fps.
3. Subtraction of known background from each created frames using pixel-by-pixel background subtraction approach.
4. Find threshold by using Otsu's Threshold method. Convert each image frame into binary image by using that threshold.
5. Create Motion History image (MHI) from thresholded images.
6. Calculate Hu Moment Invariants (Seven φ values) from this MHI.
7. For the process of recognition and classification, a table is trained. For that purpose first find maximum and minimum values of each column of the trained table, e.g.

   $MaxVC_5$ = Maximum value of Column 5 of Trained Table

   $MinVC_7$ = Minimum value of Column 7 of Trained Table
8. Compute Euclidean Distance $D_{E1}(x, y)$ between $MaxVC_4$ and $MaxVC_1$ from the trained table.

$$D_{E1}(x, y) = \sqrt{(MaxVC_4 - MaxVC_1)^2}$$

Where:

$D_{E1}(x, y)$ = First Euclidean Distance

$MaxVC_1$ = Maximum value of Column 1 of Trained Table

$MaxVC_4$ = Maximum value of Column 4 of Trained Table.

9. Calculate Euclidean Distance $D_{E2}(x, y)$ between φ_4 and φ_1.

$$D_{E2}(x, y) = \sqrt{(\varphi_4 - \varphi_1)^2}$$

Where:

$D_{E2}(x, y)$ = Second Euclidean Distance

φ_1 = First Hu Moment Invariant

φ_4 = Forth Hu Moment Invariant

10. if ($\varphi_7 > MinVC_7$ && $\varphi_5 > MaxVC_5$)

if ($\varphi_3 < MaxVC_3$)

"Kicking front Motion is recognized."

elseif ($\varphi_2 > MinVC_2$)

"Bending down Motion is recognized."

else

"Jumping up Motion is recognized."

end if

elseif ($D_{E1}(x, y) > D_{E2}(x, y)$)

if ($\varphi_3 > MinVC_3$)

"Punching forward Motion is recognized."

else

"Gun Shot Motion is recognized.

end if

end if

Unit-III

3.5 Methodology of Unit-III

In Unit-III, a new method for motion recognition & classification called Moment Invariant based Recognition and Classification (MIRC) [6] has been proposed. Moment Invariant based Recognition and Classification is based on values of Hu Moment Invariants themselves and the Euclidean Distances between these values of same image frames as well as other image frames. Performance of this new type of recognition and classification approach MIRC is compared with other classification methods (discussed earlier) using temporal template motion detection approach [1]–[3], in terms of success rate, time and speed. The steps involved in algorithm of Moment Invariant based Recognition and Classification (MIRC) are [6]:

1. Input avi file to the system and create frames from this file at the rate of 10 fps.
2. Subtract known background from each created frames using the pixel-by-pixel background subtraction approach.
3. Compute threshold level by using Otsu's Threshold method. Convert each frame into binary image by using that threshold level.
4. Calculate Hu Moment Invariants of each binary thresholded image frame.
5. Compute Euclidean Distance $D_{E1}(x, y)$ between φ_2 and φ_1 for each binary image frame. The Euclidean distance between two n-dimensional vectors x and y is:

   $$D_E(x, y) = [(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2]^{1/2}$$

6. Find the difference measure between $(D_{E1}(x, y))_L$ and $(D_{E1}(x, y))_F$.

   $$\Delta_E = (D_{E1}(x, y))_L - (D_{E1}(x, y))_F$$

   Where:

   Δ_E = Difference between Euclidean Distances.

   $(D_{E1}(x, y))_F$ = Euclidean Distance b/w φ_2 and φ_1 for first binary image frame.

   $(D_{E1}(x, y))_L$ = Euclidean Distance b/w φ_2 and φ_1 for last binary image frame.

7. Compute Euclidean distance $D_{E2}(x, y)$ between φ_4 and φ_3 for each binary image.
8. Find the difference between the values of φ_3 of first and last binary image frames.

$$\Delta_{MI} = (\varphi_3)_L - (\varphi_3)_F$$

Where:

Δ_{MI} = Difference between 3rd MI (φ_3).

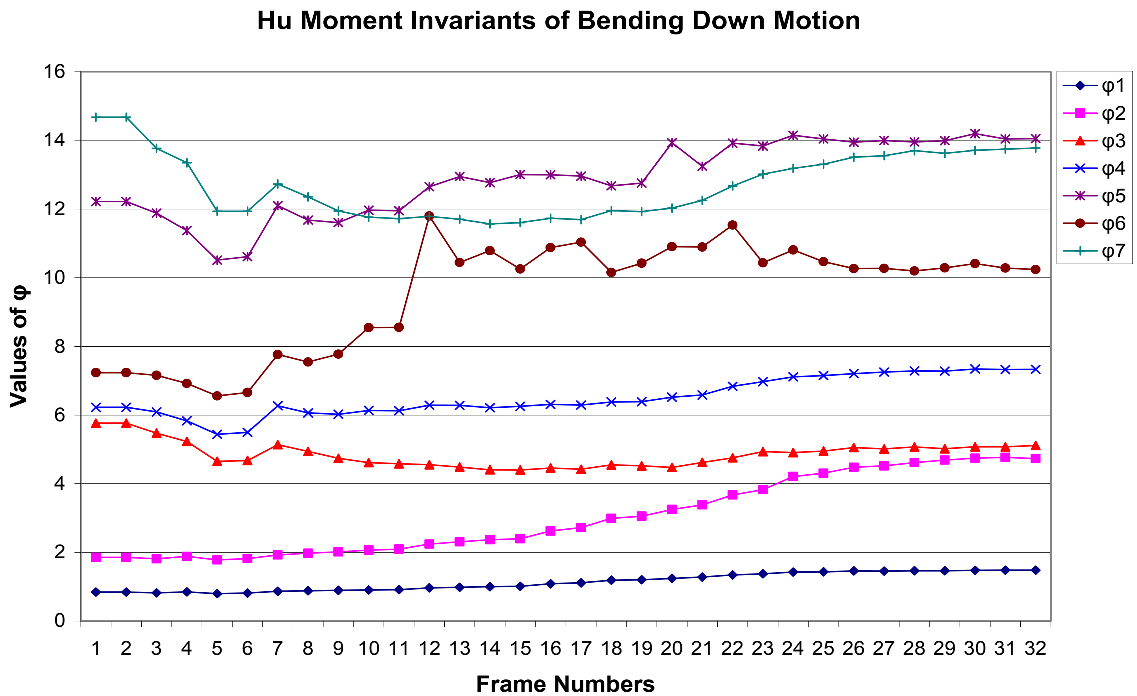
$(\varphi_3)_F$ = 3rd MI (φ_3) of first binary image frame.

$(\varphi_3)_L$ = 3rd MI (φ_3) of last binary image frame.

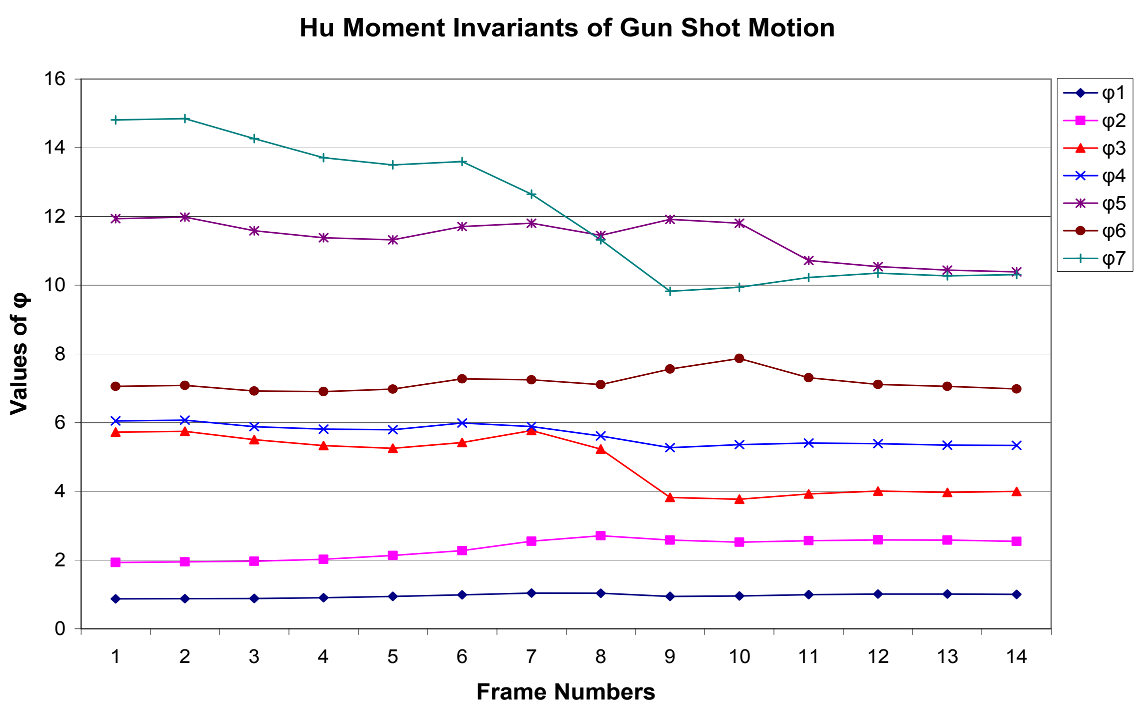
9. if $((D_{E1}(x, y))_L >= (D_{E1}(x, y))_F * 2.5)$

```
if ((φ3)L > (φ2)L)
        "Bending Down Motion found."
else
        "Kicking Front Motion found."
end if
else
    if (ΔE > 0.2)
        "Gun Shot Motion found."
    else
        if (ΔMI > 1)
            "Punching Forward Motion found."
        else
            if (φ7 > φ6 && DE2(x, y) < 1)
                "Jumping Up Motion found."
            end if
        end if
    end if
end if
```

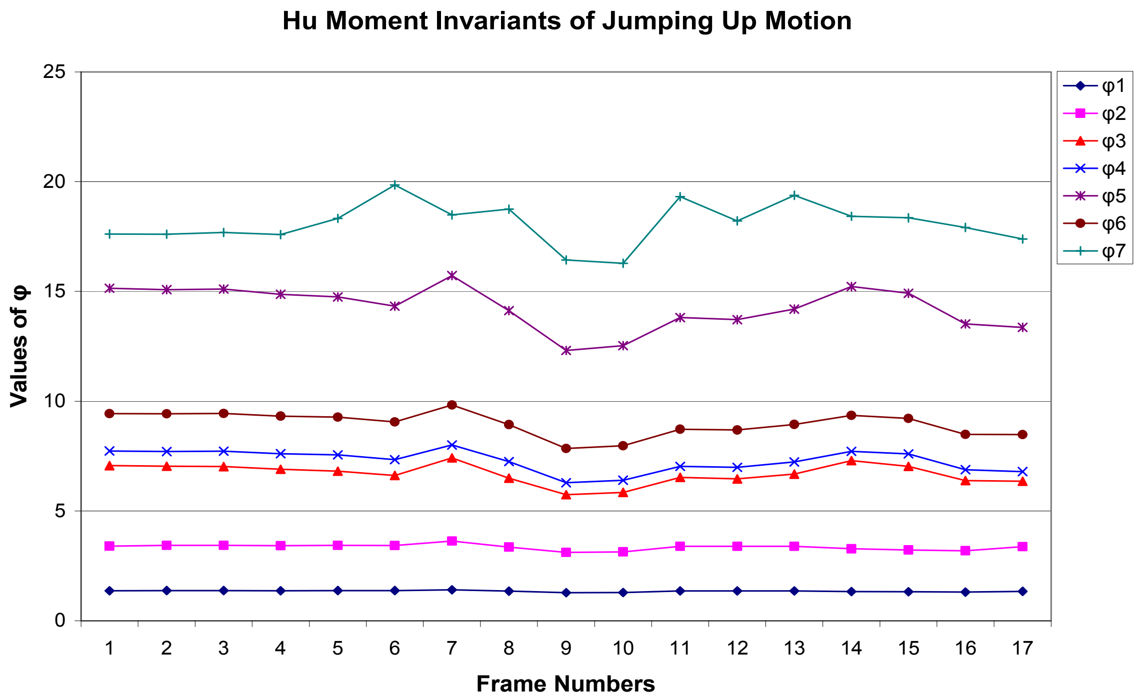
Some examples of Hu Moment Invariants of all thresholded image frames of all types of motion video clips are shown in Figure 3.15, which shows graphically seven values of the Hu Moment Invariants of each thresholded image.


Hu Moment Invariants of Bending Down Motion
Values of φ
Frame Numbers
φ1
φ2
φ3
φ4
φ5
φ6
φ7
0
2
4
6
8
10
12
14
16
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32

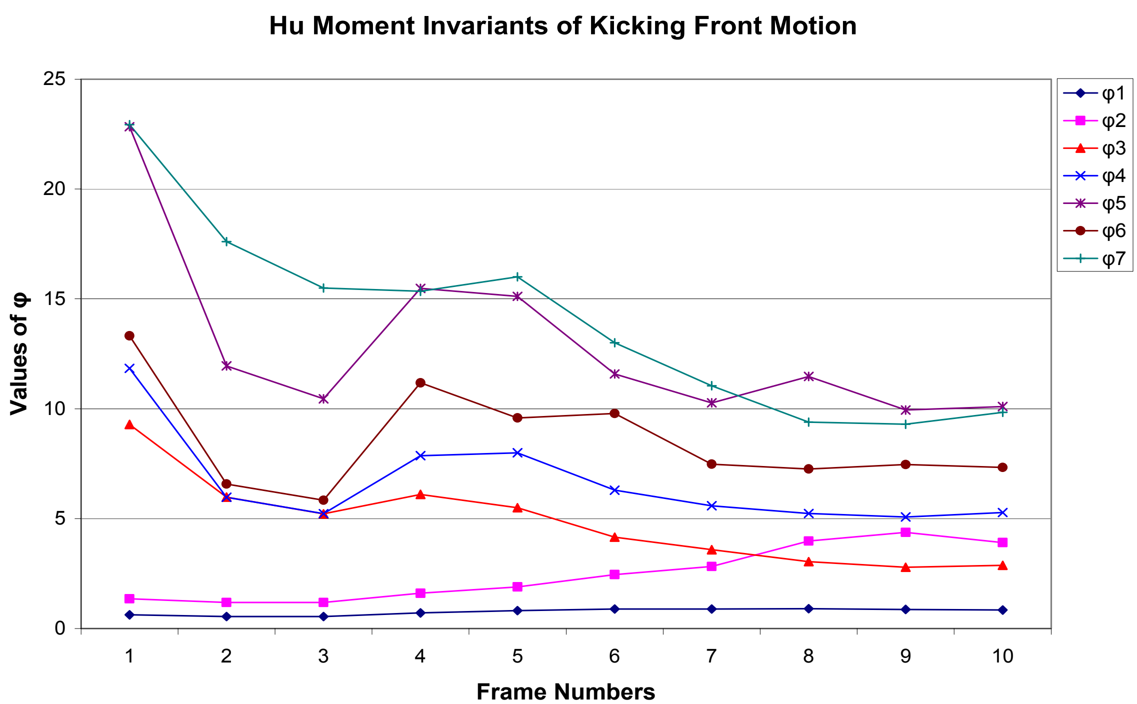
(a)


Hu Moment Invariants of Gun Shot Motion
Values of φ
Frame Numbers
φ1
φ2
φ3
φ4
φ5
φ6
φ7
0
2
4
6
8
10
12
14
16
1 2 3 4 5 6 7 8 9 10 11 12 13 14

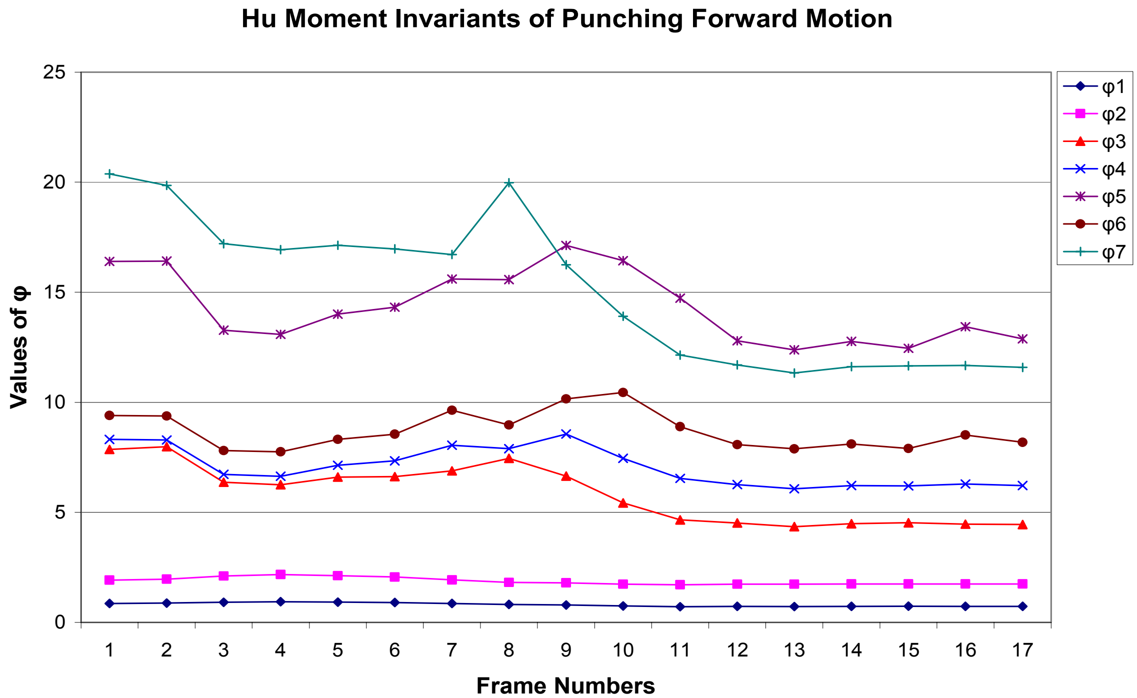
(b)


Hu Moment Invariants of Jumping Up Motion
Values of φ
Frame Numbers
φ1
φ2
φ3
φ4
φ5
φ6
φ7
0
5
10
15
20
25
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

(c)


Hu Moment Invariants of Kicking Front Motion
Values of φ
Frame Numbers
φ1
φ2
φ3
φ4
φ5
φ6
φ7
0
5
10
15
20
25
1
2
3
4
5
6
7
8
9
10

(d)



(e)

Figure 3.15: Hu Moment Invariants of all created thresholded image frames of motion types: (a) Bending down (b) Gun Shot (c) Jumping up (d) Kicking front (e) Punching forward

Moment Invariant based Recognition and Classification (MIRC) is compared with other classification methodologies which involve Temporal Template methodology [1]-[3]. The training datasets and testing datasets involved in the process of Temporal Template methodology are the same as used in Unit-I (Table 3.2).

Similarly the same classifiers are used for comparison purposes which are: Mahalanobis Distance (MD) classifier, Quadratic Bayes Gaussian (QBG) classifier, Linear Bayes Gaussian (LBG) classifier, K-Nearest Neighbor (KNN) classifier with k = 1, 3 and 5 and Fuzzy K-Nearest Neighbor (FKNN) classifier with k = 1, 3 and 5. Performance comparison between MIRC and above-mentioned classifiers has been made in terms of accuracy & success rate, time and speed.

3.6 Chapter Summary

Chapter 3 explores the methodologies of different approaches and techniques used in this research work. In other words, it illustrates the algorithmic steps of methodologies of all Units: Unit-I, Unit-II and Unit-III. Two new recognition and classification approaches TTRC and MIRC have also been explained in this chapter. It expresses the complete research in a decomposed mode and clarifies all the theoretical issues and matters involved in this research work.

4

Results & Discussions

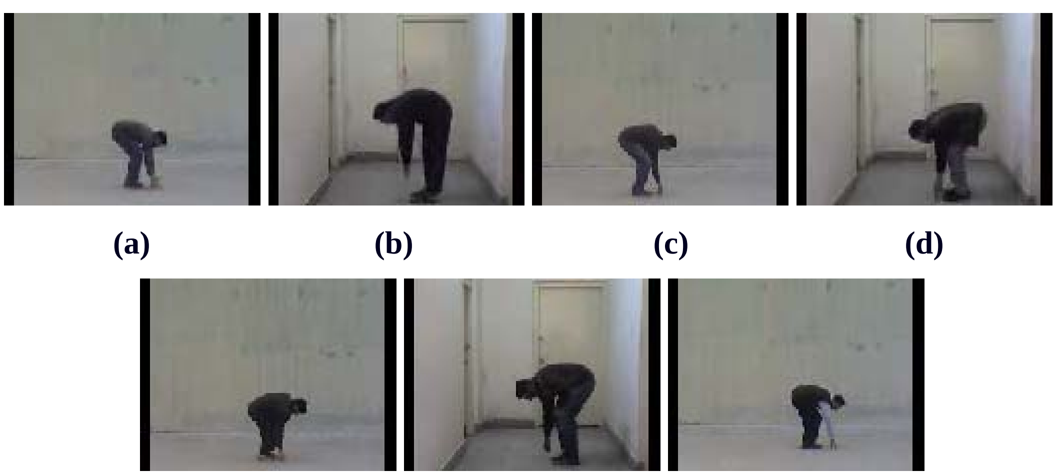
In this chapter, experimental results and the performance analysis of the experiments of all Units are carried out. Furthermore, final evaluation of the recognition and classification approaches is also elaborated. At the end, conclusion has been made based on these experimental results. Before going into details, first assumption and precondition has been explained, and then results are discussed in tabular as well as graphical format.

4.1 Assumptions and Preconditions

Experimental environments, assumptions and preconditions under which all types of video clips are collected (used in Unit-I, Unit-II and Unit-III) are [6, 78]:

- A single stationary surveillance camera is used.
- Known and static background is already provided to the developed system.
- Different backgrounds are used in collecting the training datasets and testing datasets as shown in Figure 4.1.
- Training datasets and testing datasets are independent from each other.

- Training datasets and testing datasets have been collected at different lighting conditions and at separate distances.
- Each type of motion is performed completely in video clips in order to reduce the ambiguity.
- Actors perform actions that are independent from other actions performed by other actors.
- Different environmental scenarios have been used during the recording of video clips, so that environmental barriers' effect can be checked to some extent. For example in Figure 4.1, some samples of "Bending down" motion have been shown which are recorded at different environments.
- There is only one person in field of view at any given time.
- Actors are dressed up that resemble the background and some have dressed up that contrast with background as shown in Figure 4.1 (g) where the dress of an actor at arm resembles the background.
- Each subject performs each action five to six times in different sessions so that independence can be maintained between actions.
- Each action must be unrelated & dissimilar from other.
- Minimum frame extraction rate from video clip is 10 frames per second.



(a) (b) (c) (d)

(e) (f) (g)

Figure 4.1: Different types of "Bending down" motion video clips

Experimental results of each Unit (Unit-I, Unit-II and Unit-III) have been explained in detail one by one as follows:

4.2 Experimental Results of Unit-I

Unit-I deals with classifiers' efficiency ratio at different dynamic thresholding techniques with Background Subtraction (BGS) approach as well as without Background Subtraction (BGS) approach.

4.2.1 Module-I of Unit-I *without* BGS

In First module of Unit-I, system behavior is checked without Background Subtraction (BGS) at different automatically global threshold methods instead of Trial & Error Threshold method. Four global threshold techniques are used which are Otsu's Threshold Method, Isodata Threshold Method, Image Histogram Shape based Threshold Method and Maximum Entropy based Threshold Method. The experimental results of Module-I of Unit-I have been shown in Table 4.1, Table 4.2, Table 4.3 and Table 4.4 as under. These tables elaborate the successful detection and success ratio of different classifiers at above-mentioned thresholding methods without Background Subtraction (BGS) approach.

Table 4.1: Successful detections under "Otsu's Threshold" method *without* Background Subtraction approach

Motion Types	Successful Detections								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Bending down	20	21	19	20	20	20	21	21	21
Gun Shot	16	12	9	16	11	10	11	10	11
Jumping up	19	20	19	19	20	19	19	20	19
Kicking front	20	18	18	20	18	18	20	14	20
Punching fwd.	18	18	18	18	18	18	20	20	20
Total (105)	**93**	**89**	**83**	**93**	**87**	**85**	**91**	**85**	**91**
Ratio (%)	**88.6**	**84.8**	**79**	**88.6**	**82.8**	**80.9**	**86.7**	**80.9**	**86.7**

Table 4.2: Successful detections under "Isodata Threshold" method *without* Background Subtraction approach

Motion Types	Successful Detections								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Bending down	20	21	19	20	20	20	21	21	21
Gun Shot	16	12	9	16	11	10	11	10	11
Jumping up	19	20	19	19	20	19	19	20	19
Kicking front	21	19	19	21	19	19	21	15	21
Punching fwd.	18	18	18	18	18	18	20	20	20
Total (105)	**94**	**90**	**84**	**94**	**88**	**86**	**92**	**86**	**92**
Ratio (%)	**89.5**	**85.7**	**80**	**89.5**	**83.8**	**81.9**	**87.6**	**81.9**	**87.6**

Table 4.3: Successful detections under "Image Histogram Shape based Threshold" method *without* Background Subtraction approach

Motion Types	Successful Detections								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Bending down	18	20	20	18	20	20	14	21	16
Gun Shot	9	9	8	9	9	9	6	4	7
Jumping up	19	17	17	19	18	18	18	19	18
Kicking front	18	17	17	18	17	17	20	15	20
Punching fwd.	16	16	16	16	16	16	16	20	15
Total (105)	**80**	**79**	**78**	**80**	**80**	**80**	**74**	**79**	**76**
Ratio (%)	**76.2**	**75.2**	**74.3**	**76.2**	**76.2**	**76.2**	**70.5**	**75.2**	**72.4**

Table 4.4: Successful detections under "Maximum Entropy based Threshold" method *without* Background Subtraction approach

Motion Types	Successful Detections								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Bending down	20	20	20	20	20	20	21	21	21
Gun Shot	14	10	8	14	10	9	10	9	13
Jumping up	20	20	21	20	20	20	19	20	20
Kicking front	20	19	18	20	19	18	20	15	20
Punching fwd.	17	16	17	17	17	17	20	19	19
Total (105)	**91**	**85**	**84**	**91**	**86**	**84**	**90**	**84**	**93**
Ratio (%)	**86.7**	**80.9**	**80**	**86.7**	**81.9**	**80**	**85.7**	**80**	**88.6**

4.2.1.1 Performance Evaluations of Thresholding Methods *without* BGS

In Module-I of Unit-I without Background Subtraction (BGS) approach; it is found that Isodata Global Threshold method gives maximum performance as compared to the other thresholding methods. Success rate of Otsu threshold method is quite better than Image Histogram based threshold method and Maximum Entropy based threshold method but it is a little less comparable to Isodata Threshold method. In case of Maximum Entropy based threshold method, Quadratic Bayes Gaussian (QBG) classifier performs the best as compared to other classification methods. Image Histogram Shape based threshold method gives comparatively poor results as shown in Table 4.5 and Figure 4.2.

4.2.1.2 Performance Evaluations of Classifiers *without* BGS

K-Nearest Neighbor (KNN) and Fuzzy K-Nearest Neighbor (FKNN) Classifiers with K = 1, classify successfully in almost all Global Threshold methods. Other classifiers like K-Nearest Neighbor (KNN) Classifier with K = 3 & 5, Fuzzy K-Nearest Neighbor (FKNN) Classifier with K = 3 & 5, Mahalanobis Distance (MD) Classifier, Linear Bayes Gaussian (LBG) Classifier perform the best under Isodata Global Threshold method as compared to other thresholding techniques as shown in Table 4.5 and Figure 4.2. Quadratic Bayes Gaussian (QBG) classifier shows its better results under Isodata Global Threshold method with efficiency ratio *87.6 %* but gives its superlative result under Entropy based threshold method with efficiency ratio *88.6 %* as shown in Table 4.5 and Figure 4.2.

In the whole scenario of performance evaluations of classifiers and threshoding methods without BGS, it can be concluded that K-Nearest Neighbor (KNN) classifier with K = 1 and Fuzzy K-Nearest Neighbor (FKNN) classifier with K = 1, classify most successfully with *89.5 %* efficiency ratio under Isodata Global Threshold method as shown in Table 4.5.

The complete state of the Module-I with performance evaluation of thresholding techniques and classifiers without Background Subtraction approach is shown in Table 4.5 and in Figure 4.2.

Table 4.5: Classifiers and Thresholding methods' evaluations *without* Background Subtraction approach

Threshold Techniques	Efficiency Ratio (%) of Classifiers								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Otsu	88.6	84.8	79	88.6	82.8	80.9	86.7	80.9	86.7
Isodata	**89.5**	**85.7**	**80**	**89.5**	**83.8**	**81.9**	**87.6**	**81.9**	87.6
Histogram	76.2	75.2	74.3	76.2	76.2	76.2	70.5	75.2	72.4
Entropy	86.7	80.9	**80**	86.7	81.9	80	85.7	80	**88.6**

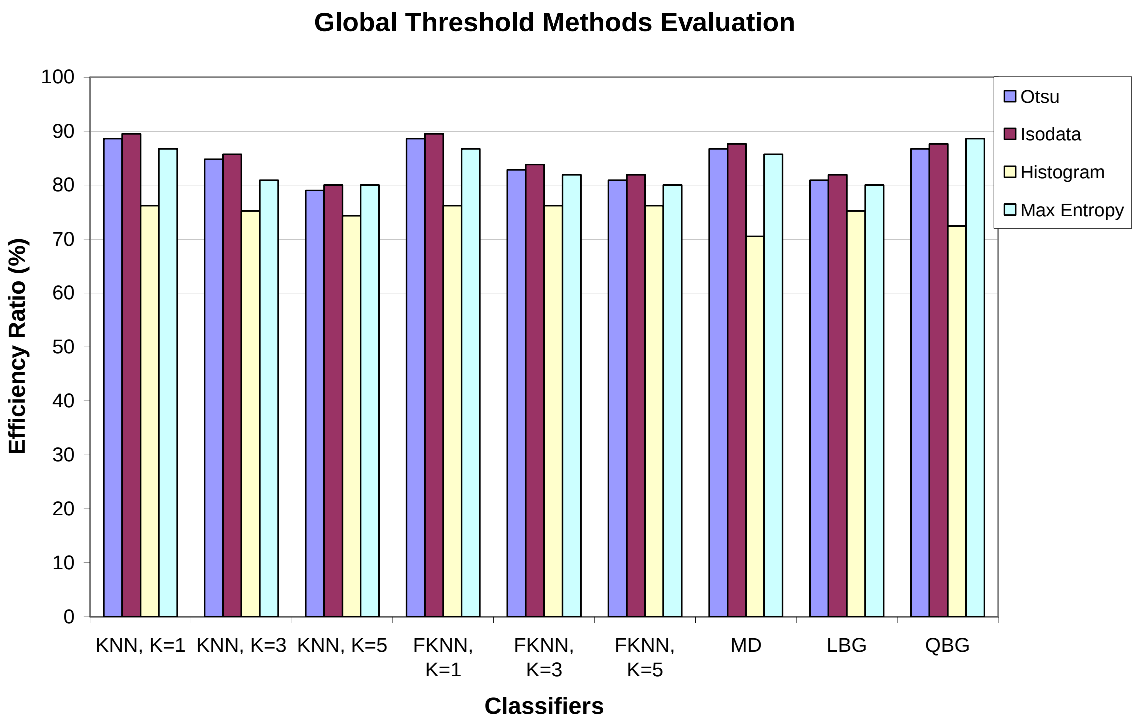


Figure 4.2: Classifiers' performances under different Thresholding techniques *without* Background Subtraction approach

4.2.2 Module-II of Unit-I *with* BGS

In this Module-II, the performances of classification techniques and automatically global threshold methods are evaluated with Background Subtraction (BGS) approach. By using the Background Subtraction (BGS) approach, better results are found for classifiers and thresholding techniques like Otsu's Threshold Method, Isodata Threshold Method and Image Histogram Shape based Threshold Method but Maximum Entropy based Threshold Method produces bad results as shown in following tables Table 4.5, Table 4.6, Table 4.7 and Table 4.8. Similar to Module-I, these tables explain the successful detection and success ratio of classifiers at different thresholding methods with Background Subtraction (BGS) approach.

Table 4.6: Successful detections under "Otsu's Threshold" method *with* Background Subtraction approach

Motion Types	Successful Detections								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Bending down	20	20	19	20	20	20	21	21	21
Gun Shot	18	15	14	18	15	15	11	12	13
Jumping up	19	18	18	19	19	19	18	19	18
Kicking front	20	20	19	20	20	20	20	17	20
Punching fwd.	18	19	19	18	19	19	20	19	20
Total (105)	**95**	**92**	**89**	**95**	**93**	**93**	**90**	**88**	**92**
Ratio (%)	**90.5**	**87.6**	**84.8**	**90.5**	**88.6**	**88.6**	**85.7**	**83.8**	**87.6**

Table 4.7: Successful detections under "Isodata Threshold" method *with* Background Subtraction approach

Motion Types	Successful Detections								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Bending down	20	20	19	20	20	21	21	21	21
Gun Shot	19	15	13	19	15	15	11	12	13
Jumping up	19	18	18	19	19	19	18	19	18
Kicking front	21	20	20	21	20	20	21	18	21
Punching fwd.	19	20	20	19	20	20	21	20	21
Total (105)	**98**	**93**	**90**	**98**	**94**	**95**	**92**	**90**	**94**
Ratio (%)	**93.3**	**88.6**	**85.7**	**93.3**	**89.5**	**90.5**	**87.6**	**85.7**	**89.5**

Table 4.8: Successful detections under "Image Histogram Shape Based Threshold" method *with* Background Subtraction approach

Motion Types	Successful Detections								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Bending down	19	20	21	19	20	21	20	21	20
Gun Shot	20	14	13	20	15	15	12	11	13
Jumping up	18	18	18	18	19	19	18	19	18
Kicking front	19	17	17	19	18	18	20	15	20
Punching fwd.	18	18	18	18	18	18	20	19	20
Total (105)	**94**	**87**	**87**	**94**	**90**	**91**	**90**	**85**	**91**
Ratio (%)	**89.5**	**82.8**	**82.8**	**89.5**	**85.7**	**86.7**	**85.7**	**80.9**	**86.7**

Table 4.9: Successful detections under "Maximum Entropy based Threshold" method *with* Background Subtraction approach

Motion Types	Successful Detections								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Bending down	18	19	18	18	19	18	10	16	10
Gun Shot	21	20	20	21	21	20	6	16	14
Jumping up	5	5	6	5	5	6	5	7	5
Kicking front	19	19	19	19	19	19	20	18	20
Punching fwd.	18	19	18	18	19	18	15	19	15
Total (105)	**81**	**80**	**79**	**81**	**83**	**81**	**56**	**76**	**64**
Ratio (%)	**77.1**	**76.2**	**75.2**	**77.1**	**79**	**77.1**	**53.3**	**72.4**	**60.9**

4.2.2.1 Performance Evaluations of Thresholding Methods *with* BGS

In Module-II of Unit-I with Background Subtraction (BGS) approach, Isodata Global Threshold method shows pre-eminent performance as in Module-I. Success rate of Otsu threshold method is again better than Image Histogram Shape based threshold method and Maximum Entropy based threshold method but less than Isodata Global Threshold method. The point notable in Module-II with Background Subtraction (BGS) is: performance of Image Histogram Shape based threshold method is much more increased which is near about the performance of Otsu threshold method. Maximum Entropy based threshold method gives poor results in Module-II with Background

Subtraction (BGS) approach in contrast with Module-I (without Background Subtraction (BGS) approach) as shown in Table 4.10 and Figure 4.3.

4.2.2.2 Performance Evaluations of Classifiers *with* BGS

K-Nearest Neighbor Classifier and Fuzzy K-Nearest Neighbor classifier with K = 1, like first Module-I of Unit-I, perform excellently under almost all thresholding methods. Fuzzy K-Nearest Neighbor (FKNN) classifier with K = 3, 5 also performs better than other classifiers, unlike Module-I. Mahalanobis Distance (MD) classifier performs a bit less efficiently as compared to FKNN but gives better performance than Linear Bayes Gaussian (LBG) and Quadratic Bayes Gaussian (QBG) classifier as shown in Table 4.10 and Figure 4.3. Mahalanobis Distance (MD) classifier and Quadratic Bayes Gaussian (QBG) classifier give worst results under Entropy based threshold method with efficiency ratio *53.3 %* and *60.9 %* as shown in Table 4.10 and Figure 4.3.

In whole situation, it is concluded that K-Nearest Neighbor Classifier and Fuzzy K-Nearest Neighbor Classifier with K = 1, like Module-I of Unit-I, perform the most excellently with *93.3 %* efficiency ratio under Isodata Threshold method.

The entire state of Module-II with performance evaluation of threshold methods and classifiers with Background Subtraction is shown in Table 4.10 and in Figure 4.3.

Table 4.10: Classifiers and Thresholding methods' evaluations *with* Background Subtraction approach

Threshold Techniques	Efficiency Ratio (%) of Classifiers								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Otsu	90.5	87.6	84.8	90.5	88.6	88.6	85.7	83.8	87.6
Isodata	**93.3**	**88.6**	**85.7**	**93.3**	**89.5**	**90.5**	**87.6**	**85.7**	**89.5**
Histogram	89.5	82.8	82.8	89.5	85.7	86.7	85.7	80.9	86.7
Entropy	77.1	76.2	75.2	77.1	79	77.1	53.3	72.4	60.9

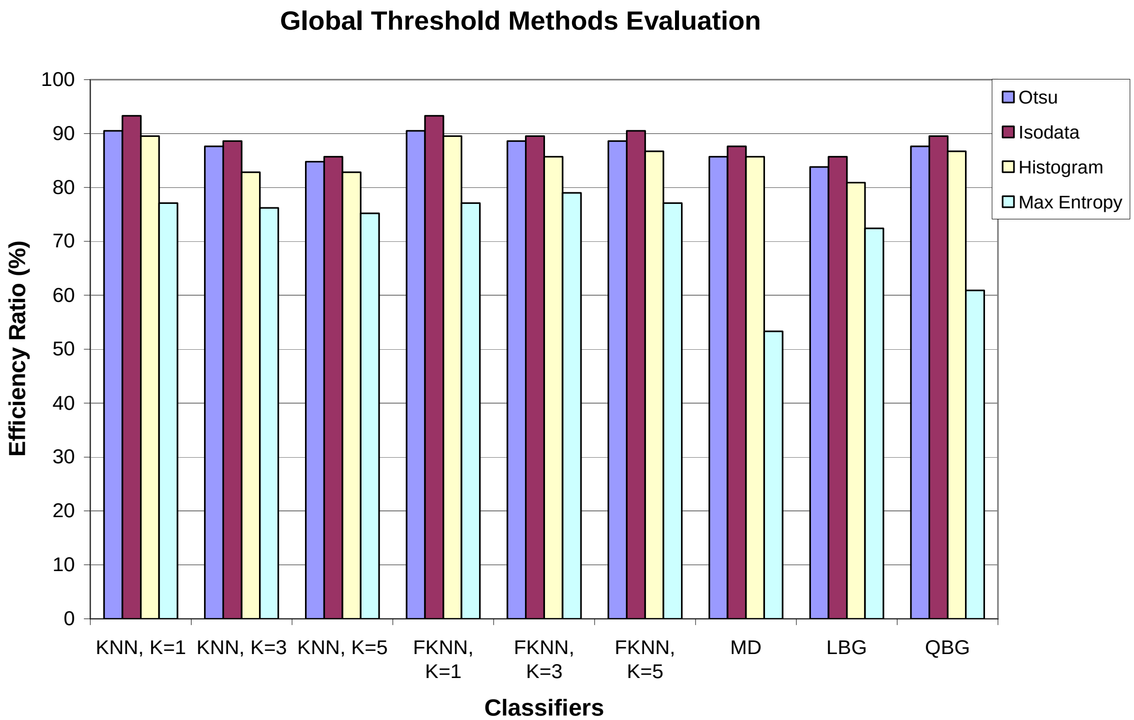


Figure 4.3: Classifiers' performances under different Thresholding techniques *with* Background Subtraction (BGS) approach

4.2.3 Experimental Assessment of Module-I & Module-II of Unit-I

The process of experimental assessment of Module-I and Module-II of Unit-I, has been made with reference to Background Subtraction approach, without Background Subtraction approach, thresholding methodologies and the classification techniques. All this can be explained as follows.

4.2.3.1 Assessment of Background Subtraction Approach

In performance evaluation of Module-I of Unit-I without Background Subtraction and second Module-II of Unit-I with Background Subtraction, it is found that overall performance of thresholding techniques has been increased in Module-II except Maximum Entropy based Threshold Method. Isodata Threshold method gives maximum performance with Background Subtraction (BGS). Under Isodata Global Threshold method, K-Nearest Neighbor Classifier (KNN) with K = 1 and Fuzzy K-Nearest Neighbor Classifier (FKNN) with K = 1 classify testing data most successfully with *93.3 %* ratio as shown in Table 4.10 and Figure 4.3.

4.2.3.2 Assessment of Threshold Methods

If the performance evaluation of thresholding techniques without Background Subtraction (BGS) and with Background Subtraction (BGS) are summarized, Otsu Threshold Method, Isodata Threshold Method and Image Histogram Shape based Threshold Method perform better with Background Subtraction than without Background Subtraction but Maximum Entropy based Threshold performs better without Background Subtraction technique as shown in Figures 4.4-4.7.

4.2.3.3 Assessment of Classifiers

In the context of classification techniques, K-Nearest Neighbor (KNN) classifier and Fuzzy K-Nearest Neighbor (FKNN) classifier with k = 1 classify more successfully with Background Subtraction (BGS) approach under Otsu's Threshold and Isodata Threshold methods. Similar case has come out for the Linear Bayes Gaussian (LBG) classifier and the Quadratic Bayes Gaussian (QBG) classifier. Mahalanobis Distance (MD) classifier performs equally with Background Subtraction and without Background Subtraction under Isodata Threshold method and it carries out its job better without Background Subtraction approach under Otsu Threshold method. In case of Image Histogram Shape based Threshold, all types of classifiers execute well with Background Subtraction technique, and a totally opposite response has been obtained in case of Maximum Entropy based Threshold method.

Comparative behavioral study of above-mentioned classifiers under dynamic thresholding techniques without Background Subtraction and with Background Subtraction has been shown in Figures 4.4-4.7.

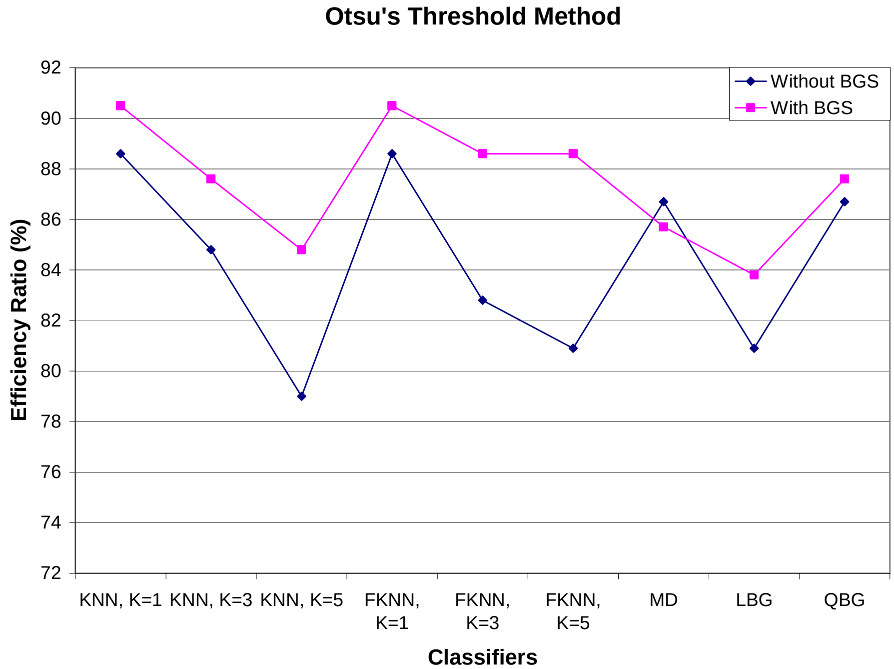


Figure 4.4: Classifiers' performances under "Otsu's Threshold" method *with* BGS & *without* BGS approaches

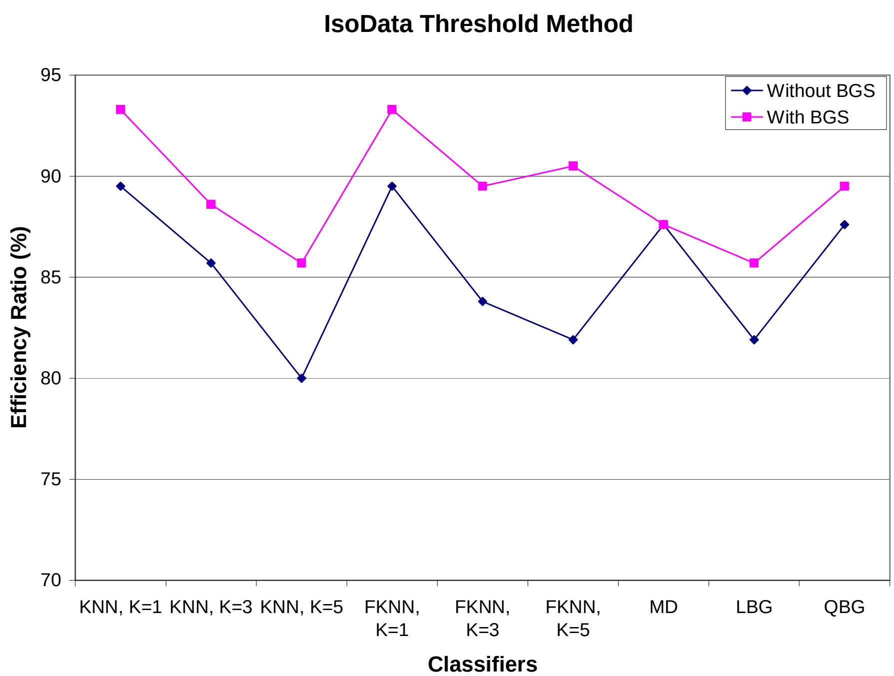


Figure 4.5: Classifiers' performances under "Isodata Threshold" method *with* BGS & *without* BGS approaches

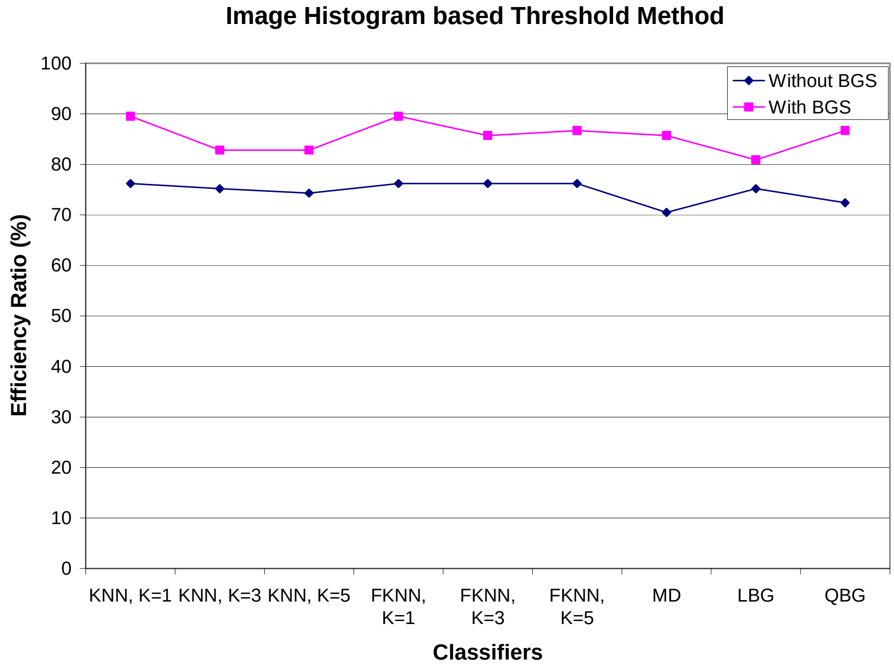


Figure 4.6: Classifiers' performances under "Image Histogram Shape based Threshold" method *with* BGS & *without* BGS approaches

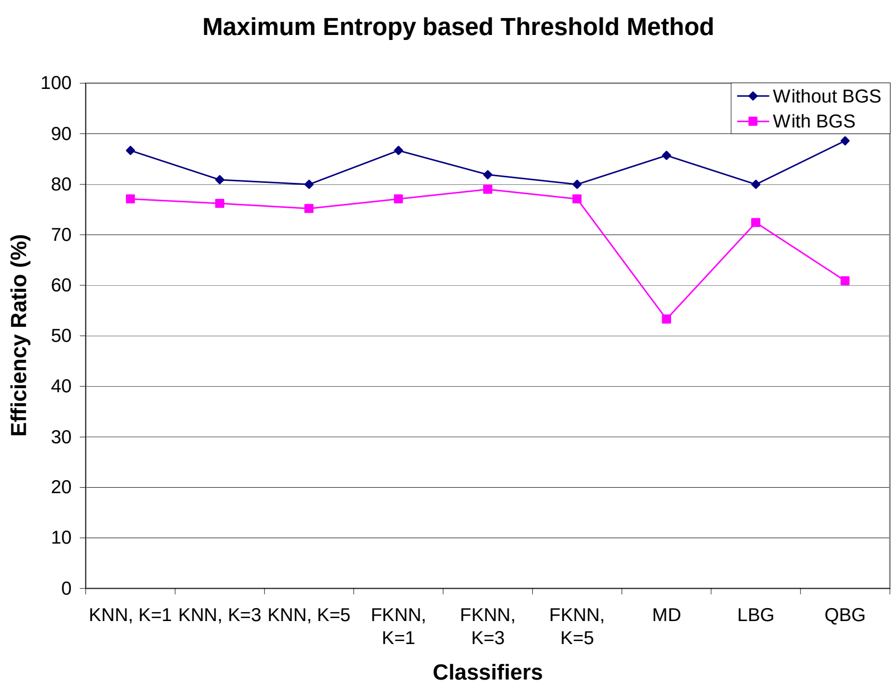


Figure 4.7: Classifiers' performances under "Maximum Entropy based Threshold" method *with* BGS & *without* BGS approaches

4.3 Experimental Results of Unit-II

Unit-II deals with different sizes of training dataset (training dataset of 105x7 elements and training dataset of 200x7 elements) as well as new proposed approach Trained Table based Recognition & Classification (TTRC) [78]. The experimental results are described in tabular form as in Table 4.11, 4.12 and Table 4.13 where successful detection as well as success rate has been shown.

Table 4.11: Classifiers' success rate when Training Dataset = 105x7 elements

Motion Types	Successful Detection When Training Dataset = 105 clips								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Bending down	8	7	8	8	7	7	8	3	9
Gun Shot	15	15	15	15	15	15	14	15	14
Jumping up	15	15	15	15	15	15	15	15	15
Kicking front	14	14	14	14	14	14	16	13	16
Punching fwd.	14	16	16	14	16	16	15	20	16
Total (90)	**66**	**67**	**68**	**66**	**67**	**67**	**68**	**67**	**70**
Ratio (%)	**73.3**	**74.4**	**75.6**	**73.3**	**74.4**	**74.4**	**75.6**	**74.4**	**77.8**

Table 4.12: Classifiers' success rate when Training Dataset = 200x 7 Elements

Motion Types	Successful Detection When Training Dataset = 200 clips								
	KNN			FKNN			MD	LBG	QBG
	1	3	5	1	3	5			
Bending down	11	13	12	11	13	12	13	14	15
Gun Shot	15	15	15	15	15	15	15	15	15
Jumping up	15	15	15	15	15	15	15	15	15
Kicking front	15	15	15	15	15	14	16	13	16
Punching fwd.	25	19	22	25	22	22	24	21	24
Total (90)	**81**	**77**	**79**	**81**	**80**	**78**	**83**	**78**	**85**
Ratio (%)	**90**	**85.6**	**87.8**	**90**	**88.9**	**86.7**	**92.2**	**86.7**	**94.4**

From the Tables 4.11 and 4.12, it is quite obvious that when training dataset increases, success rate of recognition and classification approaches increases. Each classifier performs better in Module-II with training dataset of 200x7 elements as compared to the Module-I with training dataset of 105x7 elements. For example, Mahalanobis Distance (MD) classifier has detected 68 video clips out of 90 total clips when the training dataset is of 105x7 elements while it has detected 83 video clips out of a total of 90 clips when the training dataset is of 105x7 elements.

Similarly the results of Trained Table based Recognition & Classification (TTRC) has been presented in Table 4.13. Trained Table based Recognition & Classification (TTRC) recognize and classify 80 clips out of 90 total clips, success ratio of 88.9 % as shown in Table 4.13.

Table 4.13: TTRC's success rate

Motion Types	Successful Detection
	TTRC
Bending down	15
Gun Shot	15
Jumping up	15
Kicking front	15
Punching fwd.	20
Total (90)	**80**
Ratio (%)	**88.9**

4.3.1 Performance Assessment of Unit-II

Experimental results of Unit-II explained in terms of accuracy & success rate, time & speed and storage & memory capacity are as follows.

4.3.1.1 Performance Assessment in Terms of Success Rate

In terms of accuracy and success rate, performance ratio of recognition and classification approaches in Module-I of Unit-II with training dataset of 105x7 elements, Module-II of Unit-II with training dataset of 200x7 elements and Module-III, Trained Table based Recognition & Classification (TTRC) with 3x7 elements, is shown in the Table 4.14 and Figure 4.8. It has been observed that increase in training dataset causes the increase in success rate. Quadratic Bayes Gaussian (QBG) classifier performs best with successful recognition 85 out of 90 video clips, having *94.4 %* success ratio, in Module-II while success ratio of TTRC is *88.9 %*.

Table 4.14: Recognition & Classification Techniques' success rate

Approaches	Classifiers		Total (90)	Ratio (%)
Training Dataset with 105x7 Elements	KNN	1	66	73.3
		3	67	74.4
		5	68	75.6
	FKNN	1	66	73.3
		3	67	74.4
		5	67	74.4
	MD		68	75.6
	LBG		67	74.4
	QBG		70	77.8
Training Dataset with 200x7 Elements	KNN	1	81	90
		3	77	85.6
		5	79	87.8
	FKNN	1	81	90
		3	80	88.9
		5	78	86.7
	MD		83	92.2
	LBG		78	86.7
	QBG		*85*	*94.4*
Trained Table (3x7)	TTRC		80	88.9

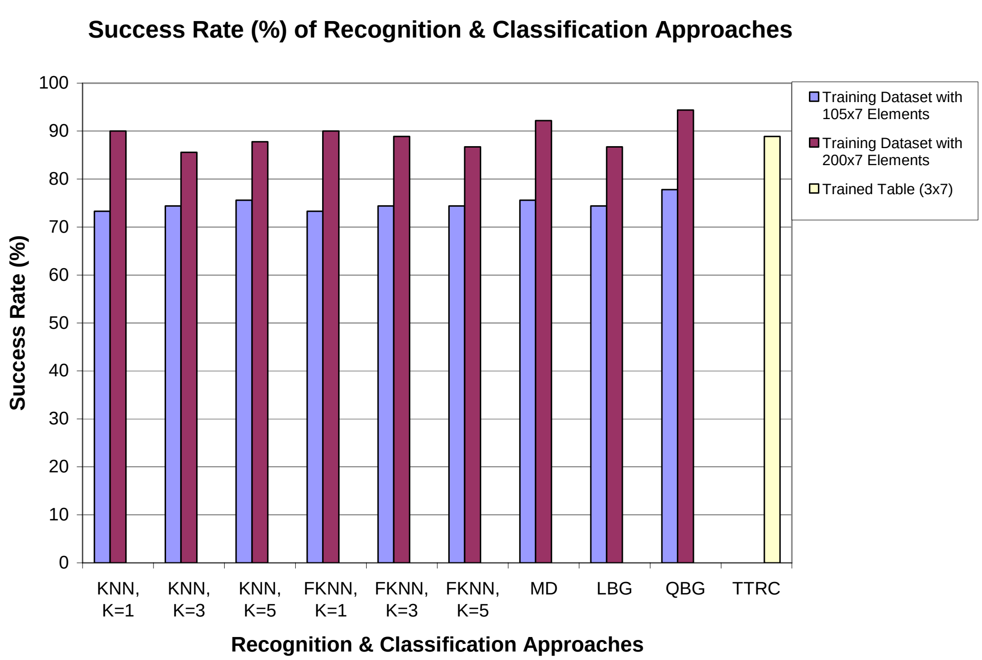


Figure 4.8: Classifiers' performances in terms of success rate

4.3.1.2 Performance Assessment in Terms of Time & Speed

In context of time & speed, efficiency of all recognition & classification techniques is described hereunder in detail. Total Times (taken to execute all 90 video clips) as well as Average Times are calculated in seconds for all classification techniques. The time taken by each classifiers increases with the increase in training dataset; it means classifiers take more time in Module-II with training dataset of 200x7 elements than in Module-I with training dataset of 105x7 elements.

QBG classifier takes less time as compared to other classifiers in Module-I and Module-II. While Trained Table based Recognition & Classification (TTRC) perform best by taking minimum total time of *85.390* seconds for 90 video clips, with average time of *0.948* seconds/clip. The whole scenario of the Module-I, Module-II, Module-III of Unit-II is shown in Table 4.15 and in Figure 4.9.

Table 4.15: Recognition & Classification Techniques' executed time

Approaches	Classifiers		Total Time (s)	Average Time (s)
Training Dataset with 105x7 Elements	**KNN**	**1**	90.326	1.004
		3	87.242	0.969
		5	88.301	0.981
	FKNN	**1**	86.151	0.957
		3	86.336	0.959
		5	86.504	0.961
	MD		86.941	0.966
	LBG		88.081	0.979
	QBG		*86.107*	*0.957*
Training Dataset with 200x7 Elements	**KNN**	**1**	91.796	1.020
		3	89.343	0.993
		5	89.484	0.994
	FKNN	**1**	86.392	0.960
		3	86.496	0.961
		5	86.617	0.962
	MD		87.076	0.968
	LBG		88.894	0.988
	QBG		*86.343*	*0.959*
Trained Table (3x7)	**TTRC**		*85.390*	*0.948*

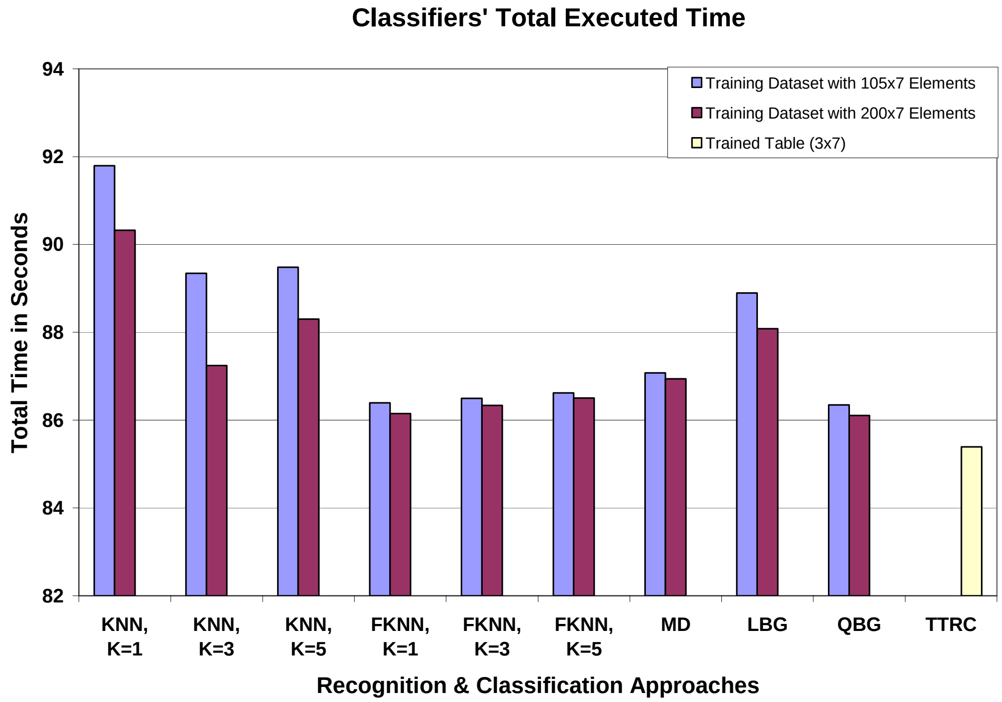


Figure 4.9: Classifiers' performances in terms of time & speed

4.3.1.3 Performance Assessment in Terms of Storage Capacity

With reference to storage and memory capacity, Module-I with training dataset of 105x7 elements takes 5880 bytes memory while Module-II with training dataset of 200x7 elements covers 11200 bytes memory. The increase in size in training dataset means the increase in storage and memory capacity. In case of Module-III, TTRC (Trained Table based Recognition & Classification) takes just 168 bytes memory as shown in Table 4.16 and Figure 4.10.

Table 4.16: Recognition & Classification Techniques' memory capacity

Training Datasets	Total Elements	Total Size in Bytes
Training Dataset with 105x7 Elements	105x7	5880
Training Dataset with 200x7 Elements	200x7	11200
Trained Table with 3x7 Elements	3x7	168

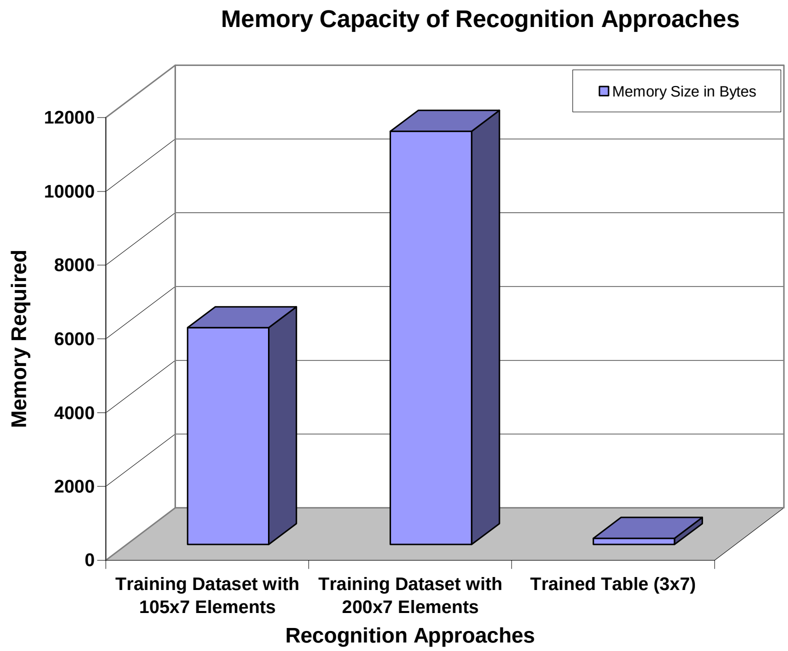


Figure 4.10: Recognition & Classification techniques' performances in terms of memory capacity

4.4 Experimental Results of Unit-III

The results of Unit-III (which deals with temporal template methodology as well Moment Invariant based Recognition and Classification (MIRC) [6]) are shown in Table 4.17 and Table 4.18 in terms of success rate and time (Total Times and Average Times).

Table 4.17: Classifiers' success rate under Unit-III

Motion Categories	Successful Detection									
	MIRC	KNN			FKNN			MD	LBG	QBG
		1	3	5	1	3	5			
Bending down	19	20	20	19	20	20	20	21	21	21
Gun Shot	21	18	15	14	18	15	15	11	12	13
Jumping up	15	19	18	18	19	19	19	18	19	18
Kicking front	21	20	20	19	20	20	20	20	17	20
Punching fwd.	16	18	19	19	18	19	19	20	19	20
Total (105)	**92**	**95**	**92**	**89**	**95**	**93**	**93**	**90**	**88**	**92**
Ratio (%)	**87.6**	**90.5**	**87.6**	**84.8**	**90.5**	**88.6**	**88.6**	**85.7**	**83.8**	**87.6**

Table 4.18: Classifiers' executed time under Unit-III

Classifiers		Total Time (s)	Average Time (s)
MIRC		*64.743*	*0.617*
KNN	**K=1**	81.250	0.774
	K=3	79.561	0.756
	K=5	79.871	0.760
FKNN	**K=1**	79.380	0.756
	K=3	78.624	0.749
	K=5	78.855	0.751
MD		80.325	0.765
LBG		78.582	0.748
QBG		79.107	0.753

4.4.1 Performance Evaluation of Unit-III

The experimental results reflect efficiency of all types of recognition and classifier techniques. The performances of the proposed system MIRC (Moment Invariant based Recognition and Classification) and the other classifiers under temporal template approach described in this research are represented in terms of accuracy & success rate and time & speed as follows:

4.4.1.1 Performance Evaluation in Terms of Success Rate

In terms of success rate, the K-Nearest Neighbor (KNN) classifier with K = 1 and Fuzzy K-Nearest Neighbor (FKNN) classifier with K = 1 perform best. Both classifiers detect and classify 95 actions out of 105 action clips and have *90.5 %* success ratio each. Quadratic Bayes Gaussian (QBG) classifier recognizes 92 out of 105 video clips, having *87.6 %* success ratio and Linear Bayes Gaussian (LBG) classifier's successful detection is 88 out of 105 video clips, having *83.8 %*. Other classifiers perform nearly equal to Quadratic Bayes Gaussian (QBG) as shown in Table 4.17 and performance ratio graph is shown Figure 4.11.

Proposed recognition and classification methodology MIRC detects and classifies 92 actions and has *87.6 %* success ratio. The success rate of MIRC is nearly equal to the best performing classifiers like KNN and FKNN as illustrated in Figure 4.11.

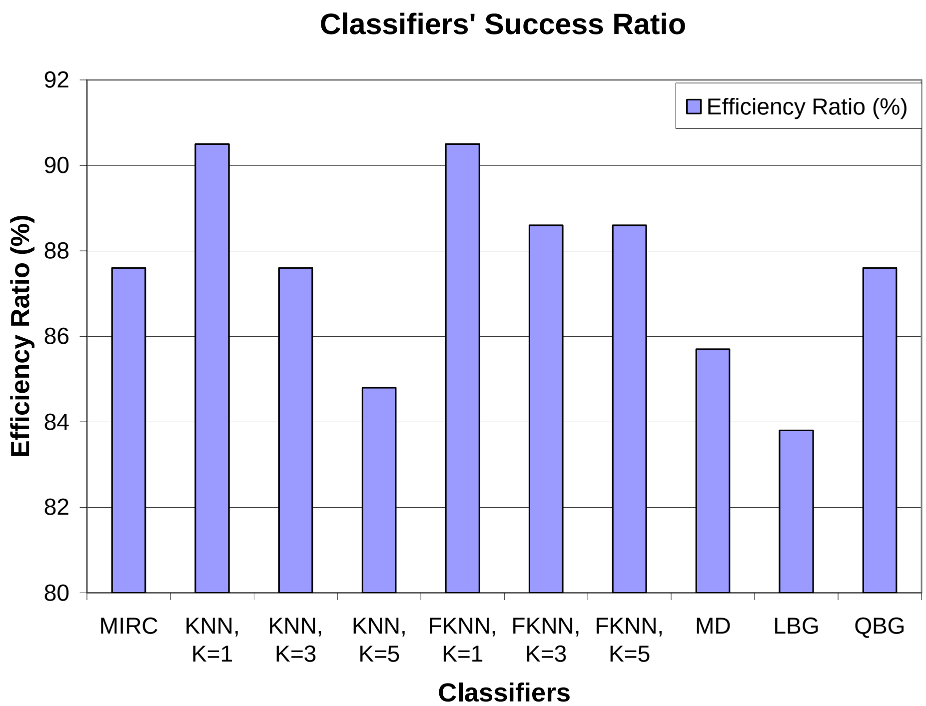


Figure 4.11: Classifiers' performances in terms of success rate

4.4.1.2 Performance Evaluation in Terms of Time & Speed

Performance & efficiency (in context of time) of all recognition & classification techniques is described as under. Total Time (time for all 105 video clips) and Average Time are calculated in seconds for all Recognition & Classification techniques including MIRC, which is represented by Table 4.18 and graphical representation is given in Figure 4.12.

With reference to time & speed, the proposed system MIRC (Moment Invariant based Recognition and Classification) performs best. It takes Total Time, just *64.743* seconds in order to execute 105 video clips, with Average Time of *0.617* seconds per clip, which is very short as compared to other discussed recognition and classification approaches [6]. While the Total Time taken by all other classifiers is near about *80* seconds and the Average Time is near about *0.750* seconds/clip as shown in Table 4.18 and Figure 4.12.

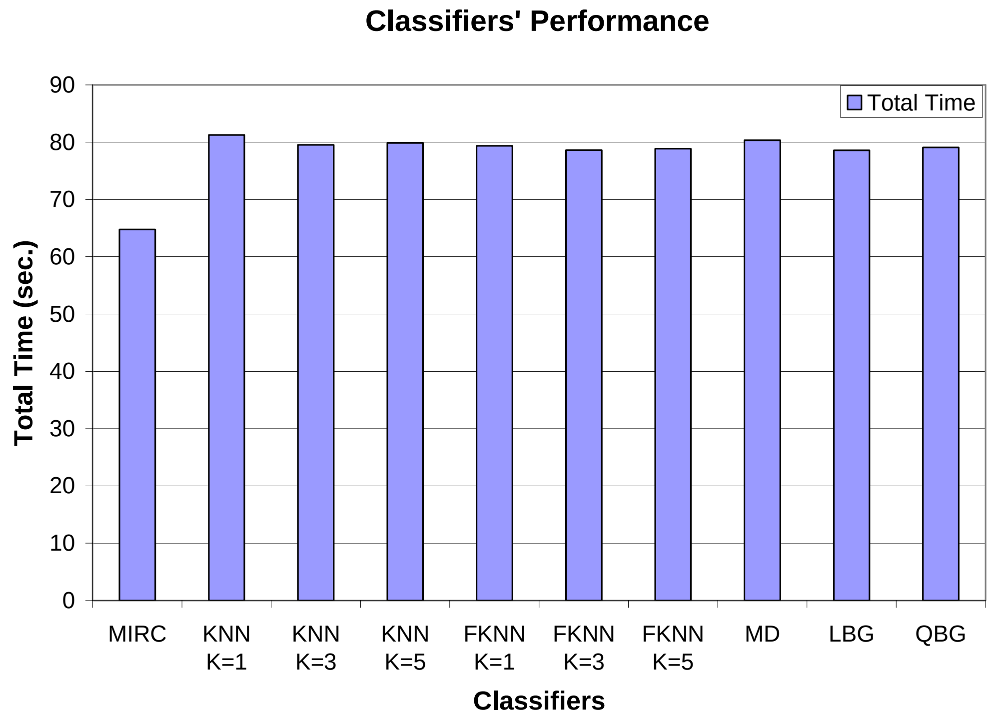


Figure 4.12: Classifiers' performances in terms of time & speed

4.5 Chapter Summary

This chapter probes into the root of the research work and provides experimental results. The comparative performance of used recognition and classification approaches has been discussed in this Chapter. The adopted and proposed techniques with some related issues and problems are explained with the help of tabular and graphical representation.

5

Analysis & Conclusion

In this chapter, the analysis and conclusion of the experiments' results discussed in the previous chapter have been made. The Recognition and Classification's behavior under different global thresholding techniques and Background Subtraction approaches are investigated and then significant, decisive and conclusive statements have been given. Comparisons of all classifiers are also summarized. Furthermore, new proposed algorithms, Trained Table based Recognition and Classification (TTRC) and Moment Invariant based Recognition and Classification (MIRC) are evaluated and compared with other classification techniques in the context of strength, accuracy & success rate, storage capacity and time & speed.

5.1 Analysis – Unit-I

From the experimental results, it is quite obvious how the selection of the suitable threshold level affects on the results in human behavior analysis systems using a view-specific approach. Human action and motion analysis systems focus on the accurate results and on timely manners. Accuracy is one of the vital factors in analysis of the human actions and movements. The basic purpose of the Unit-I is to find out different issues that directly influence the accuracy & success rate of the human motion analysis system.

In the start of Unit-I, the importance of the thresholding is described by conducting an experiment on different threshold levels under Trial and Error Threshold technique. It has been found that at suitable threshold level classifiers show their best results.

In Module-I of Unit-I, research work evaluates the performance of different classifiers like Linear Bayes Gaussian (LBG) classifier, Quadratic Bayes Gaussian (QBG) classifier, Mahalanobis Distance (MD) classifier, K-Nearest Neighbor (KNN) and Fuzzy K-Nearest Neighbor (FKNN) classifier with K = 1, 3, 5 under dynamic threshold techniques without Background Subtraction. It has been determined that using without Background Subtraction approach, performance of KNN and FKNN classifiers with k = 1, is at the top as compared to other classifiers. The best performance of Quadratic Bayes Gaussian (QBG) classifier has also been pointed out under Maximum Entropy based threshold method.

In Module-II of Unit-I, threshold level is calculated with Background Subtraction approach. It is found that all classifiers’ performance increases with Background Subtraction under Otsu threshold method, Isodata threshold method and Image Histogram shape threshold methods. K-Nearest Neighbor (KNN) and Fuzzy K-Nearest Neighbor (FKNN) classifier with K = 1 are found to outperform other classifiers under Isodata threshold method. But in case of Maximum Entropy based threshold method, performance of the classifiers decreases especially the performance of Mahalanobis Distance (MD) and Quadratic Bayes Gaussian (QBG) classifier.

The performance of Mahalanobis Distance (MD) classifier remains the same in case of Isodata threshold method with Background Subtraction approach and without Background Subtraction approach.

It has been concluded that Isodata Threshold method in thresholding techniques and, K-Nearest Neighbor (KNN) and Fuzzy K-Nearest Neighbor (FKNN) with k = 1 among classifiers, give paramount performance in whole scenario. Furthermore, the use of Background Subtraction approach in calculating the threshold levels helps in finding out the best results.

5.2 Analysis – Unit-II

The experimental results show performance of all types of recognition and classification approaches. Quadratic Bayes Gaussian (QBG) is found to outperform other classifiers in both Modules-I and Module-II with success rates of *77.8 %* and *94.4 %* respectively. In the context of time, again Quadratic Bayes Gaussian (QBG) takes fewer times in both Module-I and Module-II, with *86.107* seconds and *86.343* seconds respectively.

It is observed that if training dataset is increased, no doubt, the success rate will be better than the training dataset with fewer elements, but the time required would increase as well. As long as training dataset is increased, the accuracy, success rate and time increase and speed decreases which is a negative aspect for real-time recognition approaches. Larger training dataset also demands a larger memory capacity as well, which is another drawback. So both factors (Success rate and Time) should have to be maintained in human motion analysis system. So there is a need of such a human motion analysis system that has high accuracy, success rate and speed but it consumes less and minimum time.

Proposed recognition and classification approach, Trained Table based Recognition & Classification (TTRC) [78] has success rate nearly equal to other best performing classifiers (like QBG) in Module-II but it has taken minimum time as compared to all other classifiers. It means TTRC performs so efficiently in terms of time & speed which is a critical factor in the real time scenarios. With reference to storage & memory capacity, the proposed system TTRC performs outstandingly. It takes just 168 bytes memory as compared to Module-I and Module-II that require 5880 bytes and 11200 bytes respectively. Another main advantage of TTRC, in addition to time & speed and memory & storage capacity, is that there is no need to develop a huge training dataset. In spite of developing a huge training dataset, a simple table has been trained. On the basis of that Trained Table, recognition and classification has been performed. This simple approach (TTRC) has accuracy and success rate nearly equal to other classifiers, consumes less time and covers less storage space.

In short, it can be said that TTRC performs well when an acceptable & reasonable success rate, short time & high speed and less storage & memory capacity are required in human motion analysis systems.

5.3 Analysis – Unit-III

In Unit-III a new recognition and classification method Moment Invariant based Recognition and Classification (MIRC) is proposed and it is compared with other recognition and classification approaches. These approaches are based on temporal template techniques with Mahalanobis Distance (MD) classifier, Linear Bayes Gaussian (LBG) classifier, Quadratic Bayes Gaussian (QBG) classifier and Fuzzy K-Nearest Neighbor (FKNN) classifier with value of K = 1,3,5. Performance and efficiency is checked in terms of accuracy, time and speed. It is found that MIRC is more efficient than other discussed recognition and classification approaches with reference to time & speed and has success rate nearly equal to other techniques. Briefly speaking, it can be concluded that MIRC performs well when a reasonable success rate, short time and high speed are required in human motion analysis systems [6].

The main advantage in addition to time & speed in proposed system (MIRC) is that there is no need to develop a training dataset. In this research for other classification methodology, a training dataset of total 185 instances is developed. If the training dataset is increased, the success rate will be better, but the time required would increase as well which is a negative aspect of the recognition approaches as illustrated in Unit-II. As long as training dataset increases, the success rate and time increases and speed decreases which is a vital factor for real-time applications [6].

5.4 Chapter Summary

In Unit-I, all thresholding techniques show better results with Background Subtraction (BGS) approach as compared to without Background Subtraction approach except Maximum Entropy based threshold methods. In classifiers, KNN and FKNN classifiers with k = 1 perform superlatively well under Isodata Threshold methods with Background Subtraction (BGS) approach. In Unit-II, It has been determined that proposed approach TTRC is more efficient than other recognition and classification approaches with reference to memory capacity, time & speed and has success rate nearly at the level of other classification techniques. In Unit-III, the result analysis clearly shows that the proposed (MIRC) and adopted approaches have been proved to be very successful in recognition of various types of human motions. The performance of MIRC is more than those of other discussed recognition and classification approaches in the context of time & speed and has success rate nearly equal to other classifiers.

6

Summary & Future Work

6.1 Summary

This research work has been divided into three Units: Unit-I, Unit-II and Unit-III. In Unit-I different classifiers like Fuzzy K-Nearest Neighbor (FKNN) with K = 1, 3, 5, Mahalanobis Distance (MD) classifier, Linear Bayes Gaussian (LBG) classifier and Quadratic Bayes Gaussian (QBG) classifier are compared under different dynamic thresholding techniques like Trial & Error, Otsu, Isodata, Image Histogram shape based, Maximum Entropy based threshold methods without Background Subtraction and with Background Subtraction approaches. Isodata Threshold method with Background Subtraction approach shows the best performance. All thresholding techniques give better results with Background Subtraction approach as compared to without Background Subtraction approach except Maximum Entropy based threshold methods. In short, it can be said that the use of Background Subtraction approach in calculating the threshold levels helps in finding out the best results. In classifiers, FKNN classifiers perform excellently under Isodata Threshold methods with Background Subtraction approach.

Unit-II presents temporal template methodology with different sizes of training datasets, for example, Module-I with training dataset of 105x7 elements and Module-II with training dataset of 200x7 elements. It is observed that success rate as well as time increases with the increase of size of training dataset but the speed decreases. In Module-III (Unit-II), a new simple recognition and classification approach Trained Table based Recognition & Classification (TTRC) has been proposed which is based on simple Trained Table (TT) consisting of 3x7 elements. TTRC's performance is compared with other different classifiers used in Module-I and in Module-II with respect to accuracy, success rate, time & speed and memory & storage capacity. These classifiers are K-Nearest Neighbor (KNN) and Fuzzy K-Nearest Neighbor (FKNN) classifiers with value of K = 1,3,5, Mahalanobis Distance (MD) classifier, Linear Bayes Gaussian (LBG) classifier, Quadratic Bayes Gaussian (QBG) classifier. It is declared that proposed approach TTRC is more proficient and optimal than other recognition and classification approaches in the context of memory capacity, time & speed and has success rate nearly parallel to other classification techniques.

In Unit-III a new recognition and classification method Moment Invariant based Recognition and Classification (MIRC) has been proposed and then its performance is compared with other recognition and classification approaches. These approaches consist of temporal template techniques with Mahalanobis Distance (MD) classifier, Linear Bayes Gaussian (LBG) classifier, Quadratic Bayes Gaussian (QBG) classifier K-Nearest Neighbor (KNN) and Fuzzy K-Nearest Neighbor (FKNN) classifier with value of K = 1,3,5. Performance and efficiency of MIRC is investigated in terms of accuracy, time and speed. It is observed that MIRC is more efficient than other discussed recognition and classification approaches with regard to time & speed and has success rate almost equivalent to other classification techniques.

6.2 Future Work

The discussed research work has been conducted using 2D images frames from a single stationary surveillance camera. The known background is provided to the system. The same research can be further extended with a 3D approach using multiple moving cameras. More moving cameras are available in a restricted environment; in this way there are better probabilities in order to increase the accuracy and success rate. Moreover, with a 3D model there may be other features as well that can be extracted for detecting and recognizing the different types of human motions. This research can be provided a new way by considering the adaptive and continuously changing background and the environment. This may be a challenging factor for the accuracy of the developing human analysis system.

The research work is based on five types of motion video clips, and most of the actions are side views. It would be another area to assess how the change of view angle affects the performances of classification techniques. There is a single object in the video clips used in this research. The same research can be performed when multiple objects are present at the same time inside the movement boundary. Other improvement can be explored in the field of classification by using other new types of classifiers.

Bibliography

[1] J. W. Davis and A. F. Bobick, “The Recognition of Human Movement Using Temporal Templates”, IEEE Trans. on Pattern Analysis And Machine Intelligence, vol 23 pp. 257-267, 2001.

[2] J. W. Davis and A. F. Bobick, “The representation and recognition of action using temporal templates”, Technical Report 402, MIT Media Lab, Perceptual Computing Group, 1997.

[3] A. F. Bobick and J.W. Davis, “Real-time recognition of activity using temporal template”, Proceedings of the IEEE CS Workshop on Application of Computer Vision, pp. 39-42, 1996.

[4] L. Wang, W. Hu and T. Tan, “Recent developments in human motion analysis.” Pattern Recognition (36), pp. 585-601, 2003.

[5] T. B. Moeslund and E. Granum, “A survey of computer vision-based human motion capture”, Computer Vision Image Understnding 81 (3), pp. 231-268, 2001.

[6] H. Yasin, and S.A. Khan, “Moment Invariants based Human Mistrustful & Suspicious Motion Detection, Recognition & Classification”, Tenth International Conference on Computer Modeling and Simulation, uksim2008, pp. 734-739, 2008.

[7] R. C. Gonzalez and R.E. Woods, “Digital Image Processing”, Prentice-Hall, New Jersey, 2002.

[8] R. C. Gonzalez, R.E. Woods and S.L. Eddins, “Digital Image Processing using Matlab”, Pearson Education, New Delhi, India, 2004.

[9] A. A. Farooqi, “Detection of suspicious Human Motion from Video Frame Sequences of Visual Surveillance System Using Image Processing Techniques”, MS Thesis, National University of Science and Technology, Pakistan, 2007.

[10] I. Haritaoglu, D. Harwood, and L.S. Davis, “W4: Who, When, Where, What: A Real Time System for Detecting and Tracking People”. In: Proceedings of the Third Face and Gesture Recognition Conference, pp 222–227, April 1998.

[11] I. Haritaoglu, D. Harwood, and L.S. Davis, “W4: real time surveillance of people and their activities”, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) pp. 809-830, 2000.

[12] I. Haritaoglu, D. Harwood, and L.S. Davis, “Ghost: A Human Body Part Labeling System Using Silhouettes”. In: Proceedings International Conference on Pattern Recognition, pp 77–82, 1998.

[13] C Stauffer and W Grimson, “Learning Patterns of Activity Using Real-Time Tracking”. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(8): pp 747–757, 2000.

[14] C. Wren, A. Azerbayejani, T. Darrell, and A. Pentland, “Pfinder: Real-Time Tracking of the Human Body”. IEEE Trans. on Pattern Analysis and Machine Intelligence, 19(7):780–785, 1997.

[15] F. Porikli and O. Tuzel, “Human Body Tracking by Adaptive Background Models and Mean-Shift Analysis”, Mitsubishi Electric Research Labs, Murray Hill, NJ07974, USA.

[16] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, ”Detecting objects, shadows and ghosts in video streams by exploiting color and motion information”, Proc. of IEEE CIAP, 2001.

[17] T. Horprasert, D. Harwood, and L. Davis, “A statistical approach for real-time robust background subt. and shadow detection”, Proc. of IEEE ICCV Frame-rate Workshop, 1999.

[18] Y. Kameda and M. Minoh, “A Human Motion Estimation Method Using 3-Successive Video Frames”. In: Proceedings International Conference on Virtual Systems and Multimedia, 1996.

[19] R.T. Collins et al., “A system for video surveillance and monitoring”, VSAM Final Report, CMU-RI-TR-00-12, Technical Report, Carnegie Mellon University, 2000.

[20] A. J. Lipton, H. Fujiyoshi, R.S. Patil, “Moving target classification and tracking from real time video”, Proceedings of the IEEE Workshop on Application of Computer Vision, pp 8-14, 1998.

[21] K. T. Song and J.H. Huang, “Fast optical flow estimation and its application to real time obstacle avoidance”, Proceedings 2001 IEEE International Conference on Robotics and Automation, vol. 3, pt. 3, pp 2891-6, 2001.

[22] H.A. Rowley and J.M. Rehg, "Analyzing articulated motion using expectation maximization", Proceedings of the International Conference on Pattern Recognition, pp. 935-941, 1997.

[23] S Ju, M. Black and Y. Yacoob, "Cardboard People: A Parameterized Model of Articulated Motion". In: Proceedings 2nd International Conference on Automatic Face and Gesture Recognition, Killington, Vermont, pp 38–44, 1996.

[24] C. Bregler and J. Malik, "Tracking People with Twists and Exponential Maps". In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, 1998.

[25] M. Sezgin and B. Sankur., "Survey over image thresholding techniques and quantative performance evaluation", Journal of Electronic Imaging, 13(1), pp. 146-167, 2004.

[26] M. A. Sezan, "A Peak Detection Algorithm and Its Application to Histogram based Image Data Reduction", IEEE trans. Graphic Models and Image Processing SMC-8, 630-632, 1978.

[27] N. Otsu. "A threshold selection method from gray-level histograms", IEEE Transactions on Systems, Man and Cybernetics, 9(1): pp. 62.66, 1979.

[28] T. W. Ridler and S. Calvard, "Picture thresholding using an iterative selection method", IEEE Trans. Syst. Man Cybern. SMC-8, pp. 630–632, 1978.

[29] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram", Graph. Models Image Process. 29, pp. 273–285, 1985.

[30] L. Hertz and R. W. Schafer, "Multilevel Thresholding Using Edge Matching", Computer Vision Graphic Image Processing, 56, pp. 414-419, 1994.

[31] A. S. Abutaleb, "Automatic Thresholding of Grey-level Pictures Using Two Dimensional Entropy", Computer Vision Graphic Image Processing, 47, pp. 22-32, 1989.

[32] M. White and G. D. Rohrer, "Image Thresolding for Optical Character Recognition and Other applications Requiring Character Image Extraction", IBM J. Res. Dev. 27-4, pp. 400-411, 1983.

[33] D. Gavrilla and L. Davis "Towards 3-d model-based tracking and recognition of human movement: a multi-view approach", In International workshop on Automatic Face and Gesture Recognition, 1995.

[34] J. M. Regh and T. Kanade, "Model based tracking of self-occluding articulated objects", In CVPR, 1993.

[35] S Ju, M Black and Y Yacoob, "Cardboard People: A Parameterized Model of Articulated Motion". In: Proceedings 2nd International Conference on Automatic Face and Gesture Recognition, Killington, Vermont, pp 38–44, 1996.

[36] J. W. Davis and A. F. Bobick, "The Representation and Recognition of Human Movement using Temporal Templates". In: Proceedings Computer Vision and Pattern Recognition, pp 928–934, June 1997.

[37] R. Polana and R. Nelson. "Low level recognition of human motion", In Proceedings of IEEE Workshop on Nonrigid and Articulate Motion, 1994.

[38] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification". Wiley, 2001.

[39] C. Cedras and M. Shah, "Motion-based Recognition: A Survey", Image and Vision Computing, pp. 125-159, March 1995.

[40] C. Mohan, A. Papageorgiou and T. Poggio, “Example-Based Object Detection in Images by Components”. IEEE Trans. on Pattern Analysis and Machine Intelligence, 23(4): pp 349–361, 2001.

[41] D.M. Gavrila, “The Visual Analysis of Human Movement: A survey”, Computer Vision and image understanding, vol.73, Issue: 1, pp.82-98, January 1999.

[42] A. J. Lipton, “Local Application of Optic Flow to Analyze Rigid Versus Non-Rigid Motion”. In: Proceedings ICCV Workshop on Frame-Rate Applications, 1999.

[43] R. Rosales and S. Sclaroff, “Inferring Body Pose Without Tracking Body Parts”. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, June 2000.

[44] Q. Cai and J. K. Agarwal, “Tracking Human Motion Using Multiple Cameras”, Proceeding of 13th International Conference on Pattern Recognition, vol. 3, pp. 68-72, August 1996.

[45] D. Gavrila and L. Davis, “3-D Model-Based Tracking of Humans in Action”. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp 73–80, 1996.

[46] K. Grauman, G. Shakhnarovich, and T. Darrell, “Inferring 3D Structure with a Statistical Image-Based Shape Model”. In: Proceedings International Conference on Computer Vision, 2003.

[47] G. Bradski and J. Davis, “Motion Segmentation and Pose Recognition with Motion History Gradients”, Machine Vision and Applications, vol. 13, Issue: 3, pp. 174-184, December 2002.

[48] A. B. Poritz, “Hidden Markov models: a guided tour”, Proceedings of the International Conference on Acoustics Speech and Signal Processing, pp. 7-13, 1988.

[49] M. Brand, N. Oliver, and A. Pentland, “Coupled Hidden Markov Models for Complex Action Recognition”. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp 994–999, 1997.

[50] Y. Guo, G. Xu and S. Tsuji, “Understanding human motion patterns”, Proceedings of the International Conference on Pattern Recognition, pp. 325-329, 1994.

[51] D. Hogg, “Model-based vision: a program to see a walking person, Image and Vision Computing”, 1(1): pp 5-20, 1983.

[52] K. Rohr, “Towards model-based recognition of human movement in image sequences”, CVGIP: Image Understanding 59 (1), pp 94-115, 1994.

[53] B. Freeman and M. Roth. “Orientation histograms for hand gesture recognition”. In Proc. Int. Conf. on Auto. Face and Gesture Recognition, 1995.

[54] Y. Yacoob and M. J. Black, “Parameterized Modeling and Recognition of Activities”. In: Proceedings International Conference on Computer Vision, pp 120–127, 1998.

[55] A. F. Bobick and A. D. Wilson, “A state-based technique for the summarization and recognition of gesture”, In Proc. of 5th Intl. Conf. on Computer Vision, pp 382-388, 1995.

[56] O. Chomat and J. L. Crowley, “Probabilistic Recognition of Activity using Local Appearance”. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, II: pp 104–109, 1999.

[57] G. R. Bradski and J. W. Davis, “Motion Segmentation and Pose Recognition with Motion History Gradients," Machine Vision and Applications, vol. 13:3, pp. 174-184, Jul. 2002.

[58] J. Han and B. Bhanu, “Individual Recognition Using Gait Energy Image," IEEE Trans. on Pattern Anal. and Machine, vol. 28:2, pp. 316-322, Feb. 2006.

[59] T. Xiang and S. Gong, “Beyond Tracking: Modelling Activity and Understanding Behaviour," Int. Journal of Computer Vision, vol. 67:1, pp. 21- 51, Feb. 2006.

[60] A. Sharma, D. K. Kumar, S. Kumar, and N. McLachlan, “Recognition of Human Actions Using Moment Based Features and Artificial Neural Networks”, Proceedings of the 10th International Multimedia Modeling Conference, pp. 368, 2004.

[61] S. M. Seitz and C. R. Dyer, “View-Invariant Analysis of Cyclic Motion”. International Journal of Computer Vision, 25: pp 1–25, 1997.

[62] C. Canton-Ferrer, J.R.Casas and M.Pardas, “Human Model and Motion Based 3D Action Recognition in Multiple View Scenarios”, European Signal Processing Conference EUSIPCO 2006, Firenze (Italy), Sept, 2006.

[63] C. Lo.and H. Don, “3-D Moment Form: Their Construction and Application to Object Identification and Positioning,” IEEE Trans. On Pattern Anal. And Machine, vol. 11:10, pp 1053-1063, Oct, 1989.

[64] L. W. Campbell, D. A. Becker, A. Azerbayejani, A. Bobick, and A. Pentland, “Invariant Features for 3-D Gesture Recognition”, 2nd Int. Conf. on Automatic Face and Gesture-Recognition, Killington, Vermont, 1996.

[65] C. Barron and I. A. Kakadiaris, “Estimating Anthropometry and Pose from a Single Uncalibrated Image”, Computer Vision and Image Understanding, 81, 2001.

[66] H. Fujiyoshi and A. Lipton, “Real-Time Human Motion Analysis by Image Skeletonization”. Fourth IEEE Workshop on Applications of Computer Vision, pp 15–21, 1998.

[67] M. K. Leung and Y.H. Yang. “First sight: A human body outline labeling system. IEEE PAMI, 17(4), pp 359–377, April 1995.

[68] A. Ali and J. K. Aggarwal, “Segmentation and Recognition of Continuous Human Activity”. Proc. IEEE Workshop on Detection and Recognition of Events in Video, July 2001.

[69] A. Elgammal, D. Harwood and L.S. Davis, “Nonparametric background model for background subtraction”, Proceedings of the Sixth European Conference on Computer Vision, 2000.

[70] H.Z. Sun, T. Feng and T.N. Tan, “Robust extraction of moving objects from image sequences”, Proceedings of the fourth Asian Conference on Computer Vision, Taiwan, pp. 961-964, 2000.

[71] T. Pun, ‘‘A new method for gray-level picture threshold using the entropy of the histogram’’, Signal Process. 2 (3), pp 223–237, 1980.

[72] T. Pun, “Entropic thresholding: A new approach”, Comput. Graph. Image Process. 16, pp. 210–239, 1981.

[73] M. S. Nixon and A. S. Aguado, “Feature Extraction and Image Processing”, Newnes, New Delhi, India, 2002.

[74] W. K. Kim and Y. Sung, “A region based shape descriptor using Zernike moments”, Signal Processing Image Commun. 16 pp: 95-102, 2000.

[75] M. Hu. “Visual pattern recognition by moment invariants”, IRE Trans. Information Theory, vol. 8, no. 2, pp. 179-187, 1962.

[76] R. R. Pinho, J. M. R. S. Tavares and M. V. Correia, “Efficient Approximation of the Mahalanobis Distance For Tracking with Kalman Filter”, Internation Journal on Simulation Model 6 2, pp 84-92, 2007.

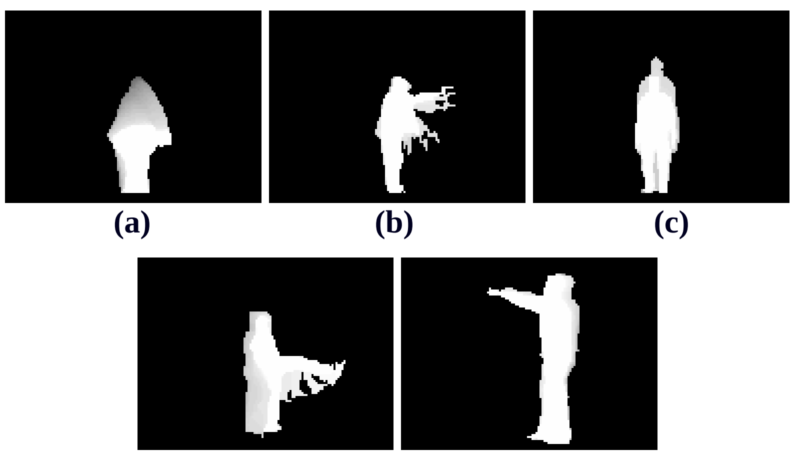
[77] J. M. Keller, M. R. Gray and J. A. Givens, , "A Fuzzy K-Nearest Neighbor Algorithm", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 15, No. 4, pp. 580-585, 1985.

[78] H. Yasin, and S.A. Khan, “Trained Table based Recognition & Classification (TTRC) Approach in Human Motion Reconstruction & Analysis”, 4th International Conference on Emerging Technologies, ICET 2008, pp. 734-739, 2008.

A

Frames & MHIs

Appendix A presents extracted frames of different human motion video clips and then describes the formation of Motion History Image (how as opposed to where motion exist) step by step. Motion History Image (MHI) shows the whole history of the motion. Some examples of Motion History Images (MHIs) are as below.



(a) (b) (c)

(d) (e)

Figure A.1: Motion History Images: (a) Bending down (b) Gun Shot (c) Jumping up (d) Kicking front (e) Punching forward

The steps involved in the formation of Motion History Image (MHI) are shown in Figures A.2 to A.11.

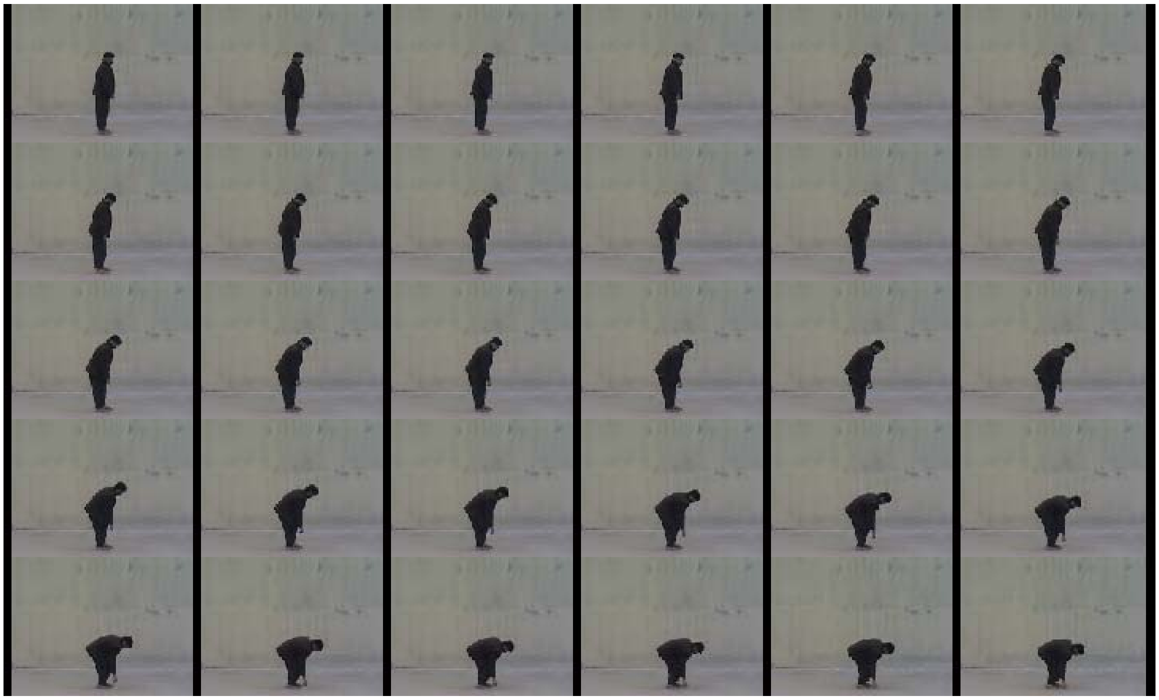


Figure A.2: Frames extracted from "Bending down" motion video clip

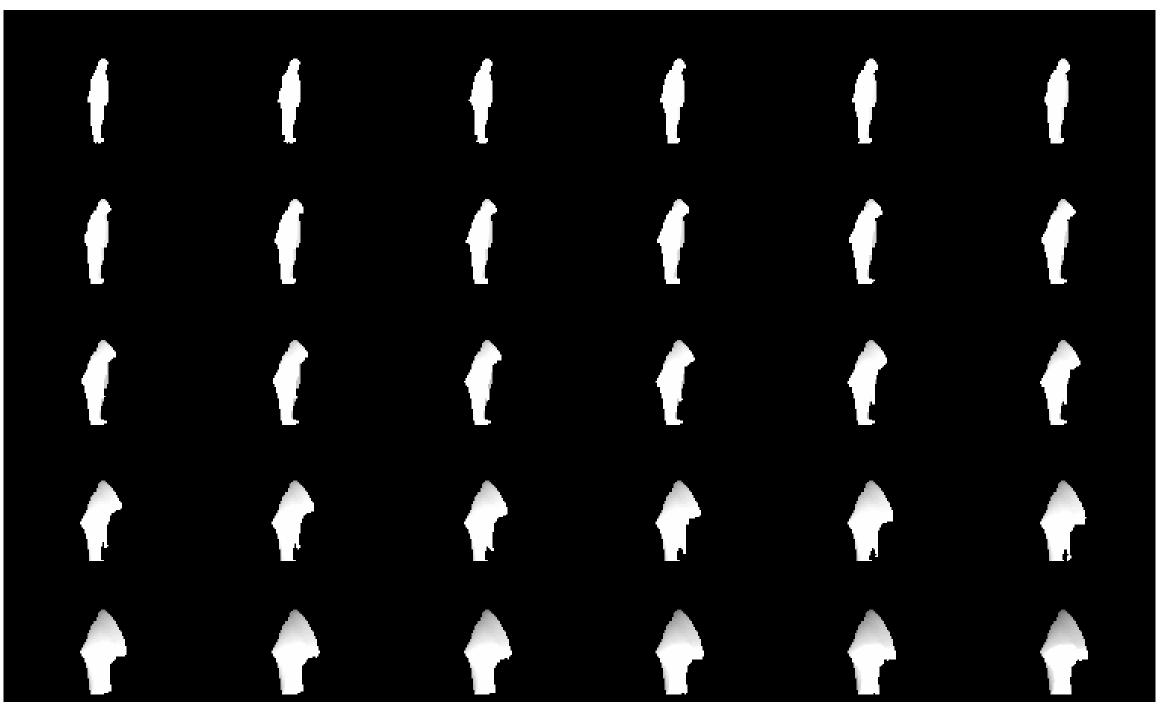


Figure A.3: Motion History Image (MHI) formation from extracted frames of "Bending down" motion video clip

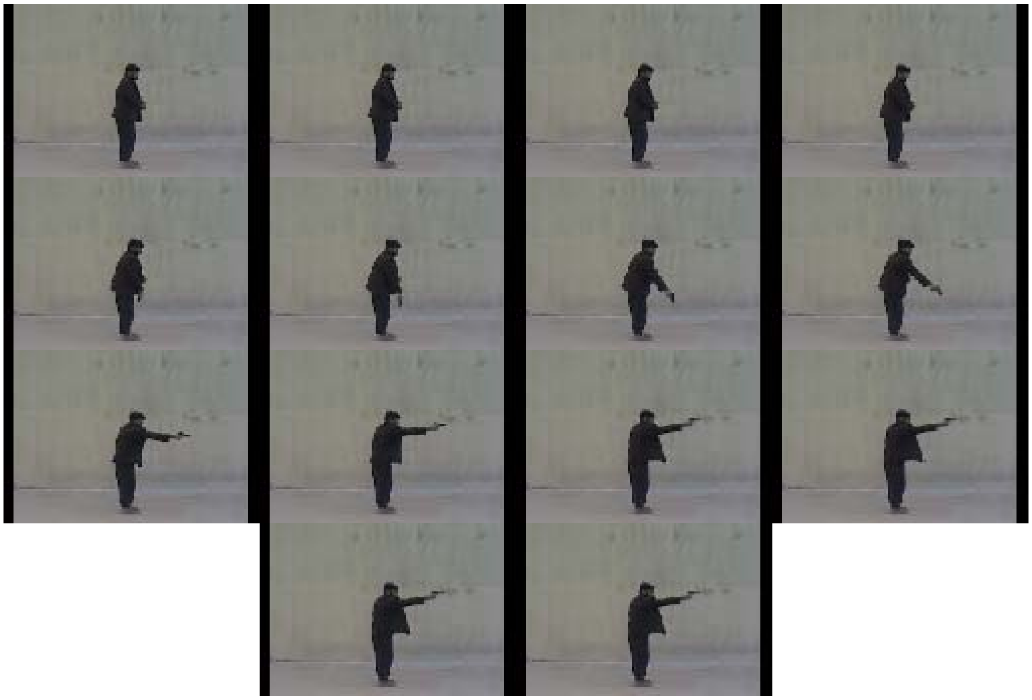


Figure A.4: Frames extracted from "Gun Shot" motion video clip

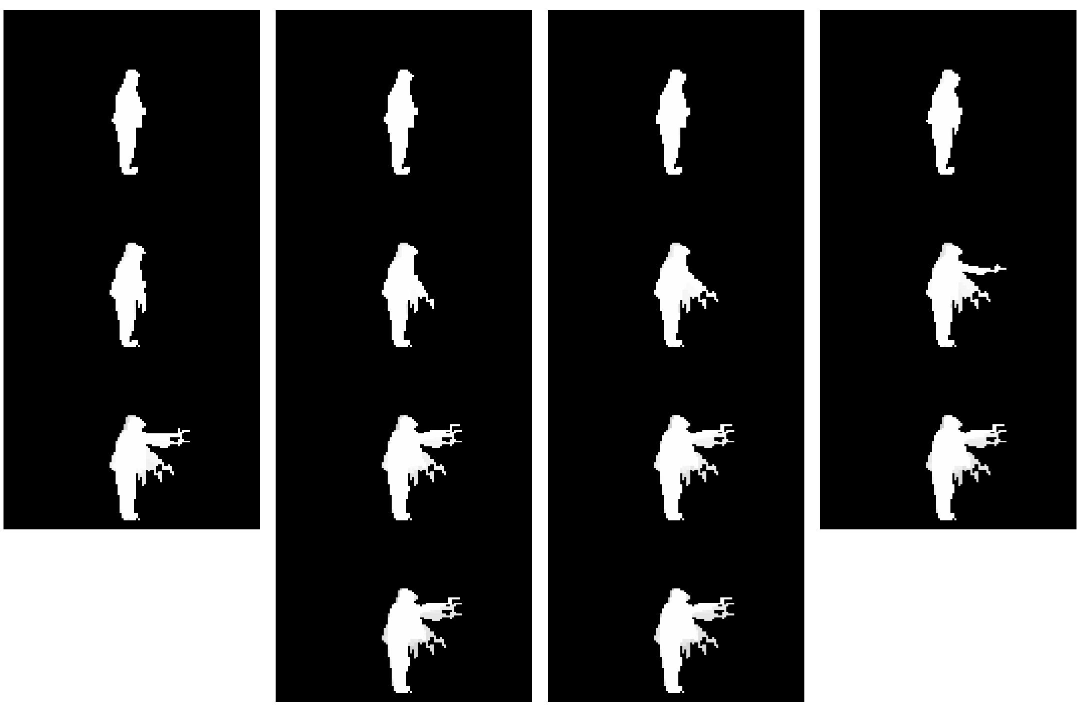


Figure A.5: Motion History Image (MHI) formation from extracted frames of "Gun Shot" motion video clip

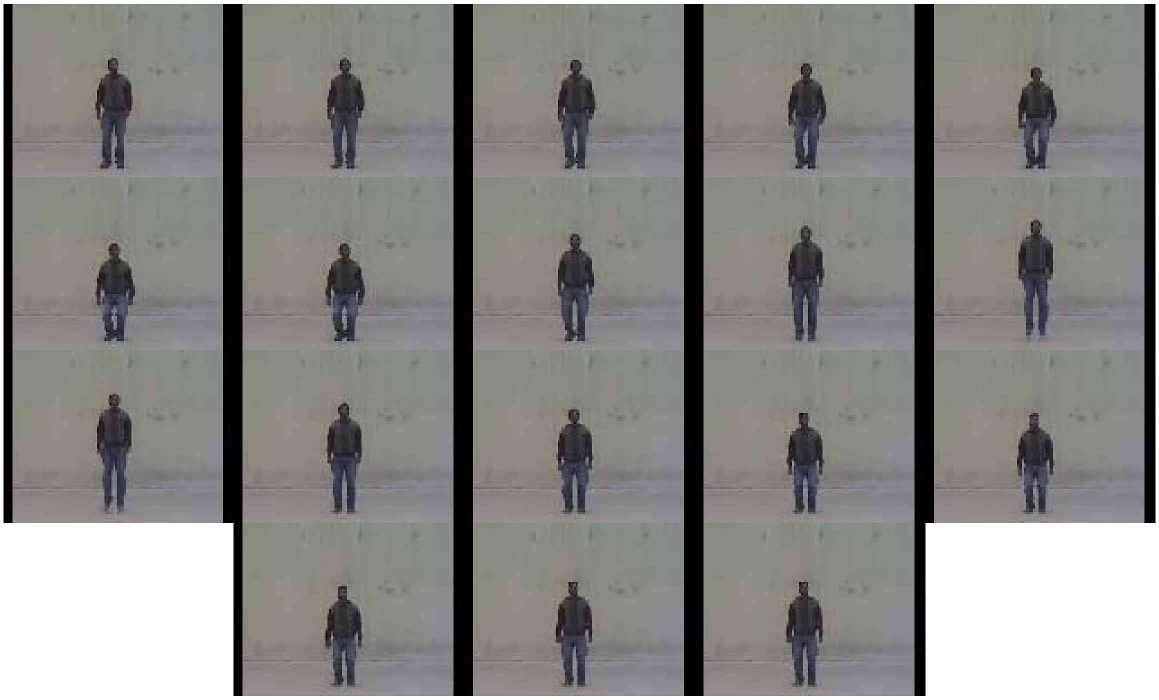


Figure A.6: Frames extracted from "Jumping up" motion video clip

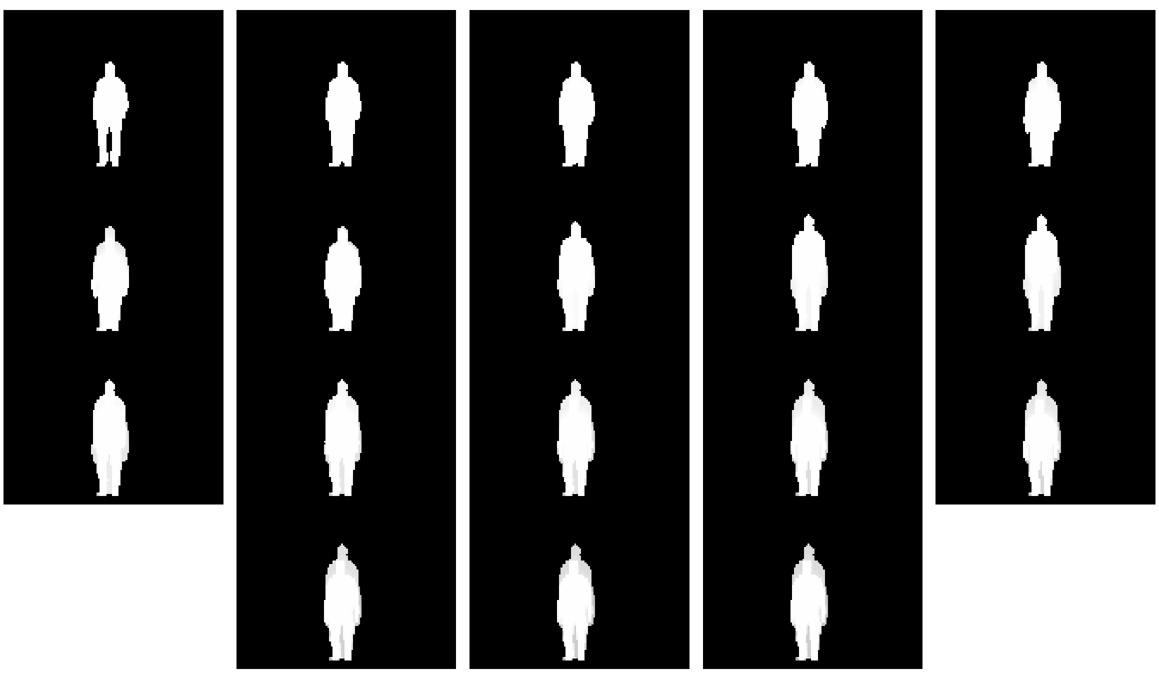


Figure A.7: Motion History Image (MHI) formation from extracted frames of "Jumping up" motion video clip

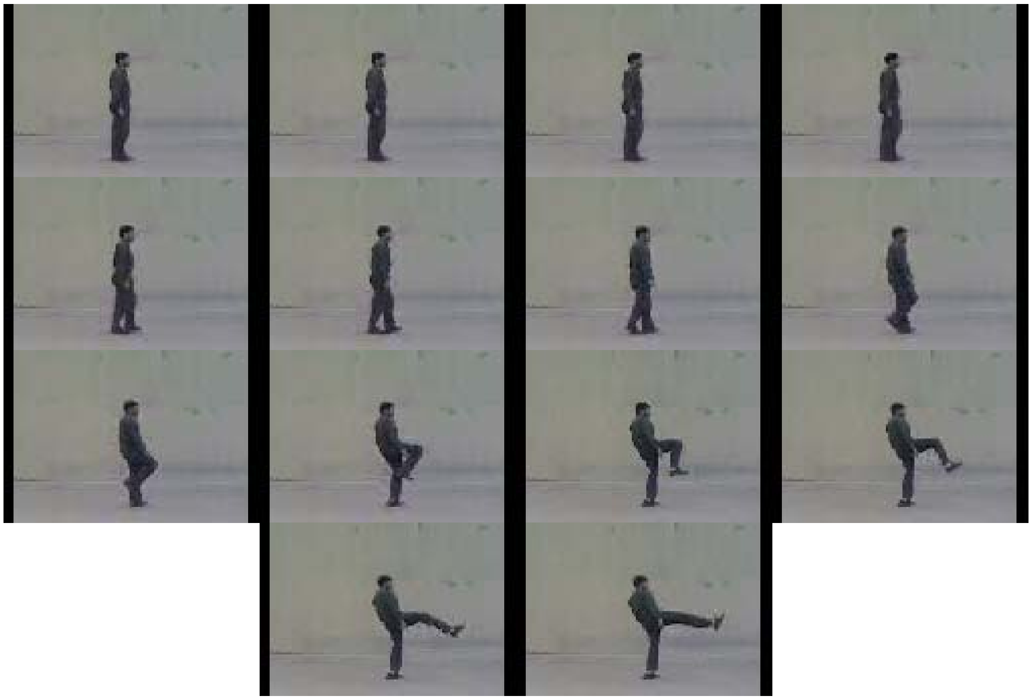


Figure A.8: Frames extracted from "Kicking front" motion video clip

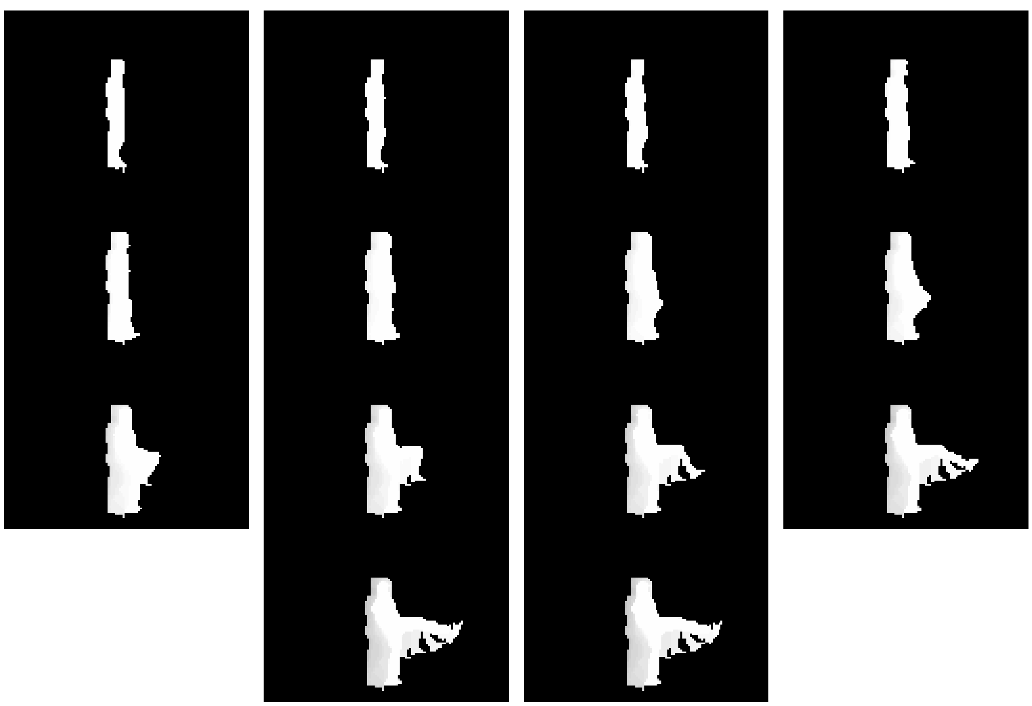


Figure A.9: Motion History Image (MHI) formation from extracted frames of "Kicking front" motion video clip

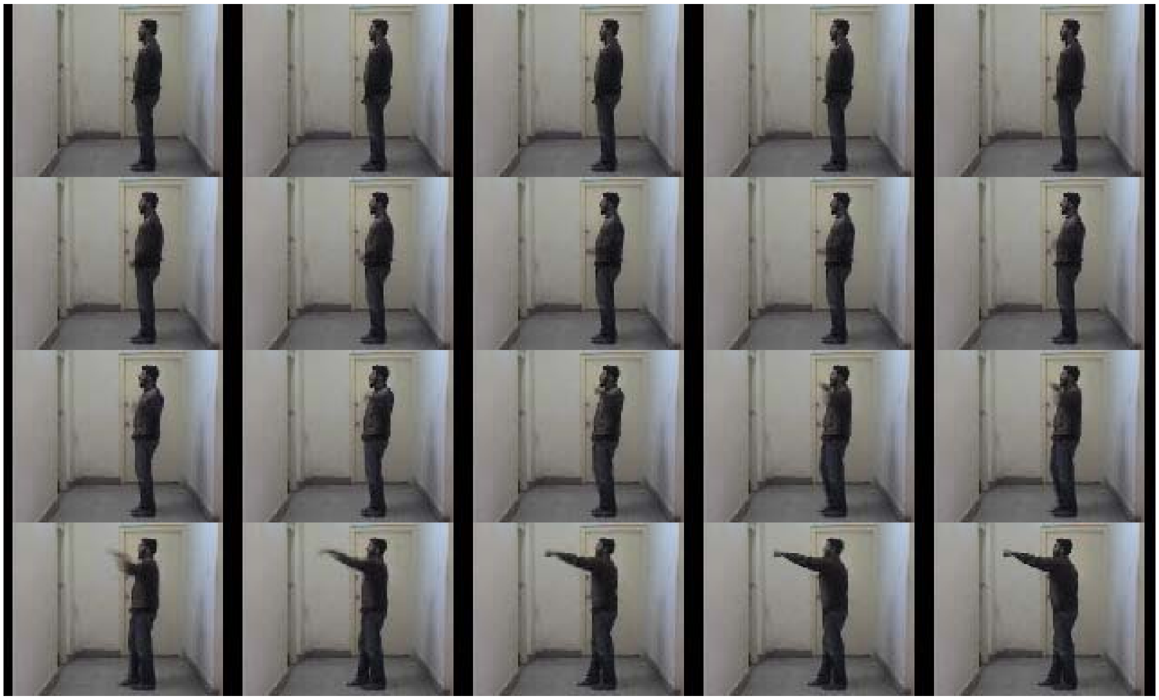


Figure A.10: Frames extracted from "Punching forward" motion video clip

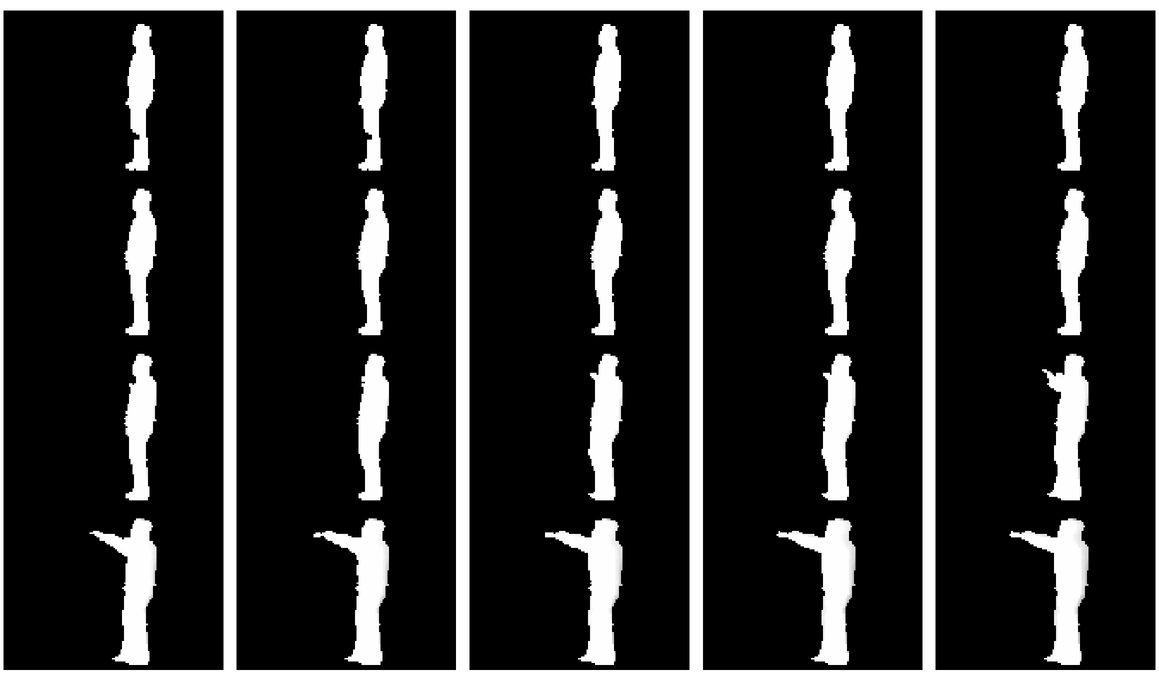
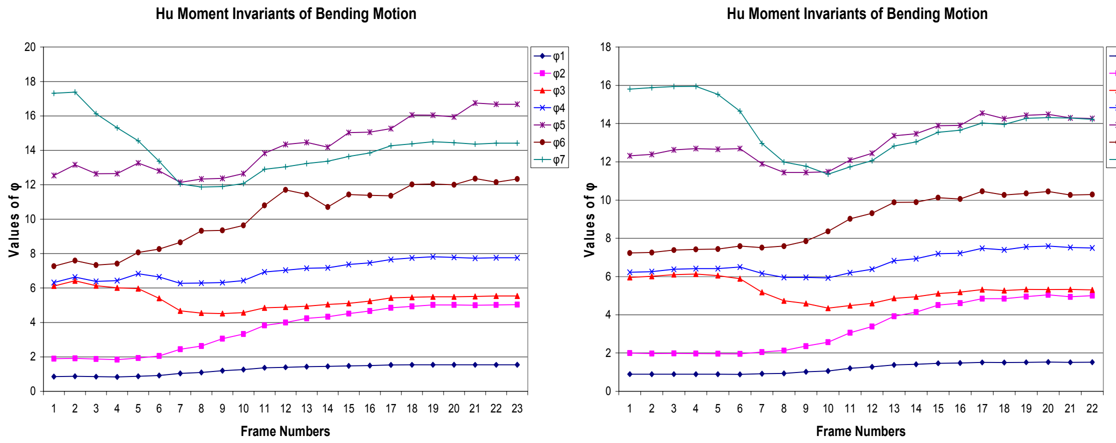


Figure A.11: Motion History Image (MHI) formation from extracted frames of "Punching forward" motion video clip

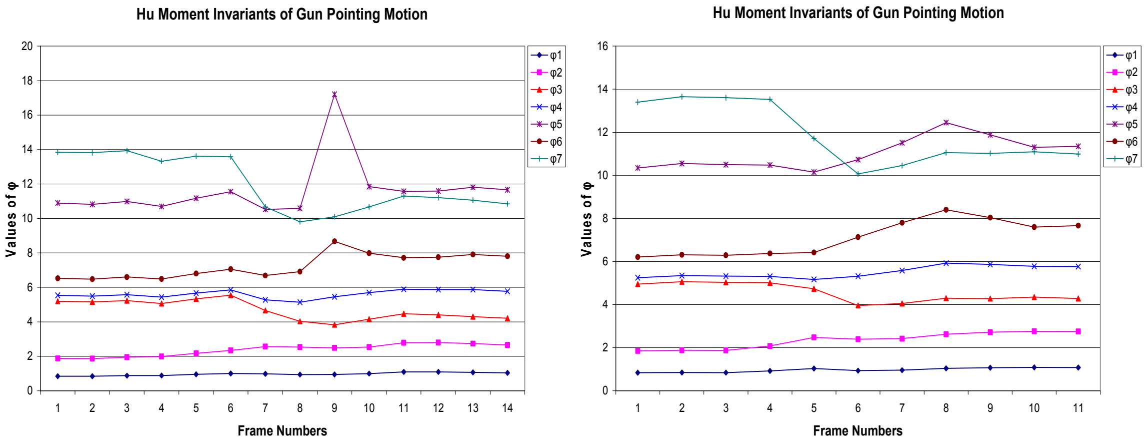
B

Hu Moment Invariants

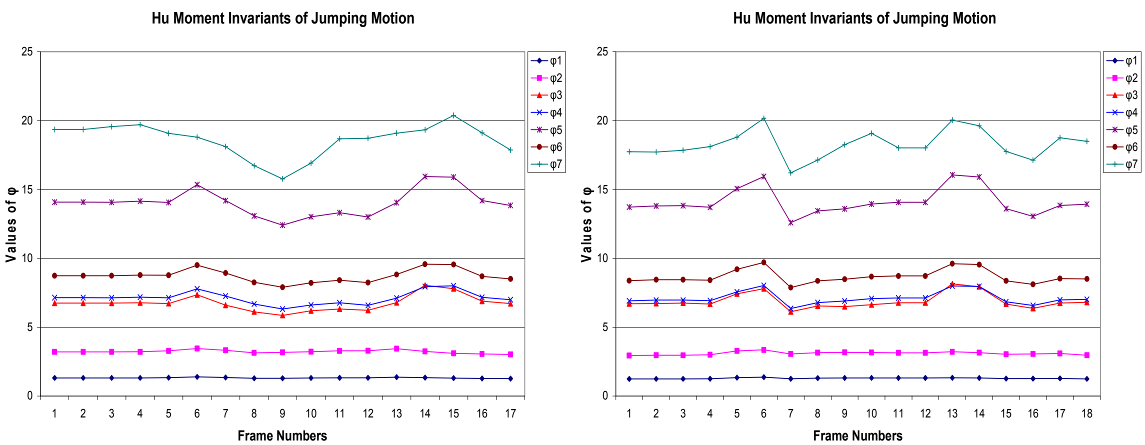
Appendix B shows Hu Moment Invariants calculated for each frame (after thresholding & segmentation) extracted from human motion video clips (Bending down, Gun Shot, Jumping up, Kicking front and Punching forward). These seven values of Hu Moment Invariants become the basis for new proposed methodology “Moment Invariant based Recognition and Classification (MIRC)”. Some examples of Hu Moment Invariants computed from thresholded image frames extracted from motion video clips are shown as below:



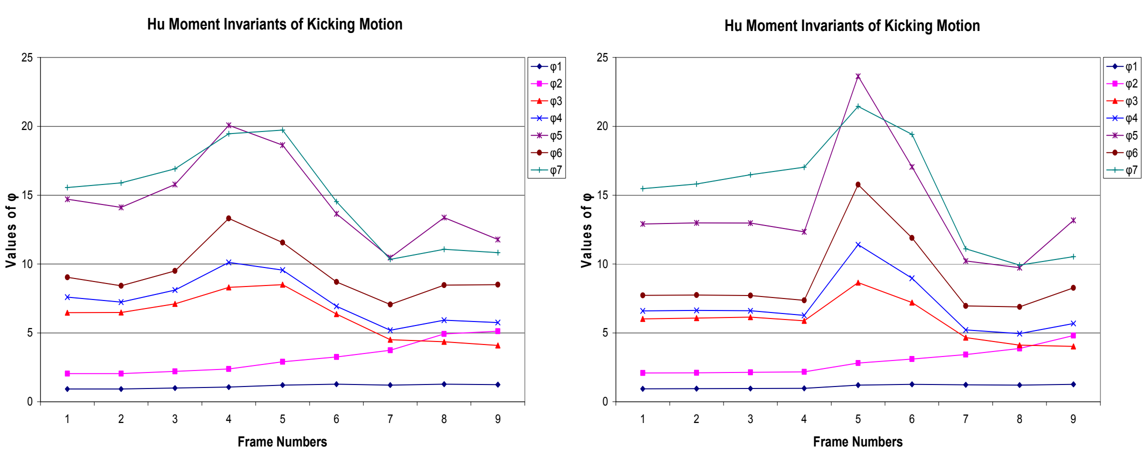
(a)


Hu Moment Invariants of Gun Pointing Motion
Values of φ
Frame Numbers
φ1
φ2
φ3
φ4
φ5
φ6
φ7
Hu Moment Invariants of Gun Pointing Motion
Values of φ
Frame Numbers

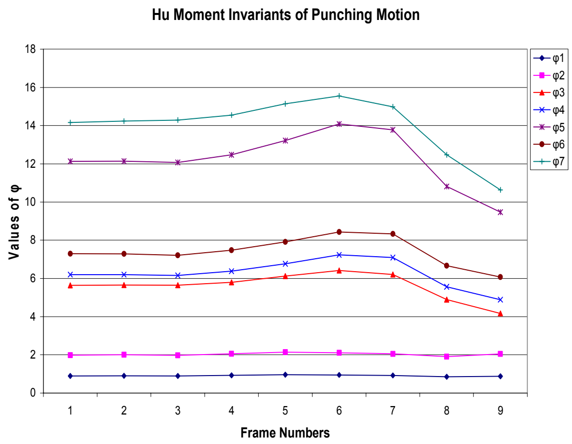
(b)

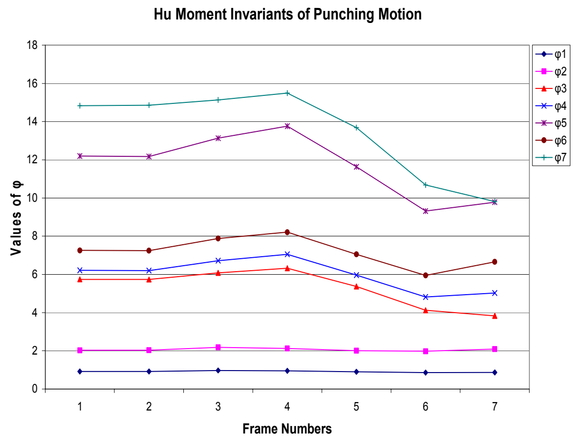

Hu Moment Invariants of Jumping Motion
Values of φ
Frame Numbers
φ1
φ2
φ3
φ4
φ5
φ6
φ7
Hu Moment Invariants of Jumping Motion
Values of φ
Frame Numbers

(c)


Hu Moment Invariants of Kicking Motion
Values of φ
Frame Numbers
φ1
φ2
φ3
φ4
φ5
φ6
φ7
Hu Moment Invariants of Kicking Motion
Values of φ
Frame Numbers

(d)





(e)

Figure B.1: Hu Moment Invariants of all created thresholded image frames of motion types: (a) Bending down (b) Gun Shot (c) Jumping up (d) Kicking front (e) Punching forward

C

PDV and PMV

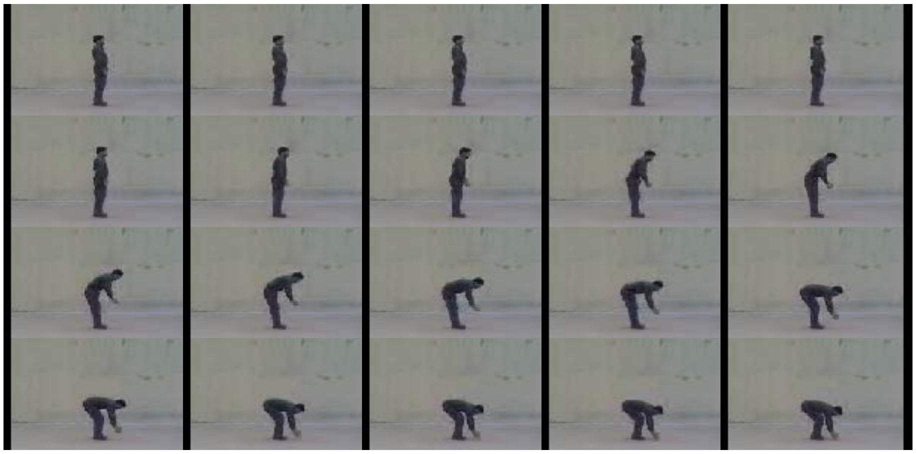
In Appendix C, Pixels' Difference Values (PDV) and Pixels' Motion Values (PMV) are explained. First of all, frames are extracted from the motion video clip (Bending down, Gun Shot, Jumping up, Kicking front and Punching forward) and then pixels' differences between these extracted frames are calculated. The values of pixels, when motion exists in the frame, (PMV) have also been determined as shown below.

C.1 Pixels' Difference Values (PDV)

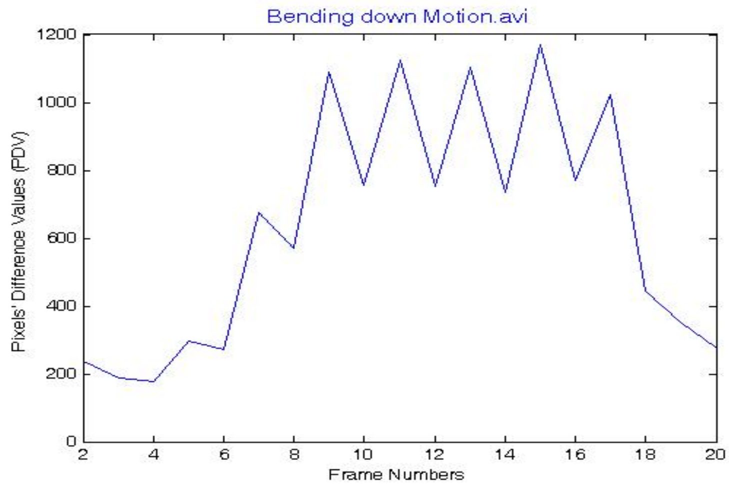
These are the values of pixels that are based on the distances of pixels between the extracted frames of motion video clip. Some examples of Pixels' Difference Values (PDV) for each type of motion have been shown as under.

C.2 Pixels' Motion Values (PMV)

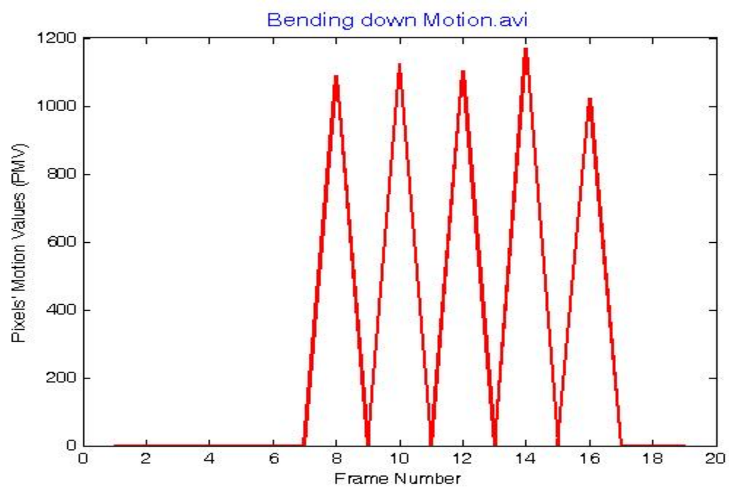
These are the values of pixels that represent the existence of motion in the extracted frames of motion video clip. Pixels' Motion Values (PMV) differentiate between the moving pixels (the pixels whose values have been changed) as well as static pixels (the pixels whose values remain unchanged). Pixels' Motion Values (PMV) graphs have been shown in Figures C.1 to C.5.



(a)

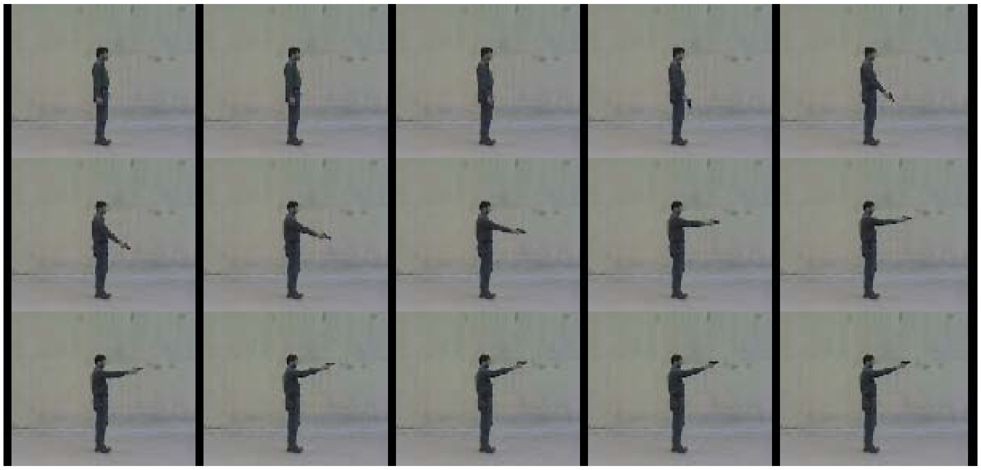


(b)

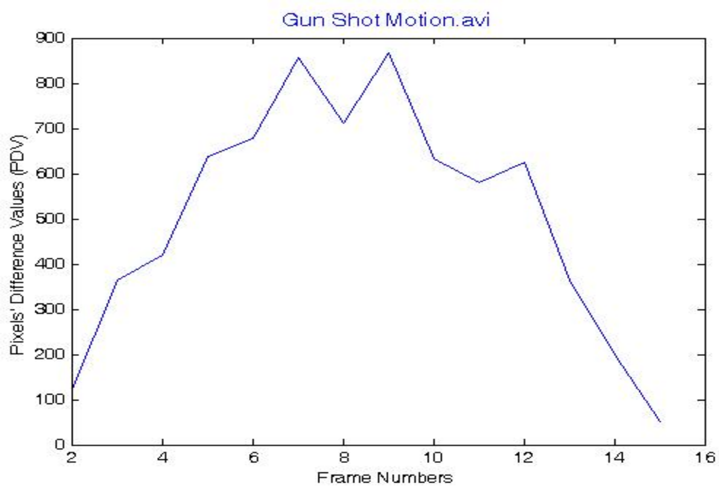


(c)

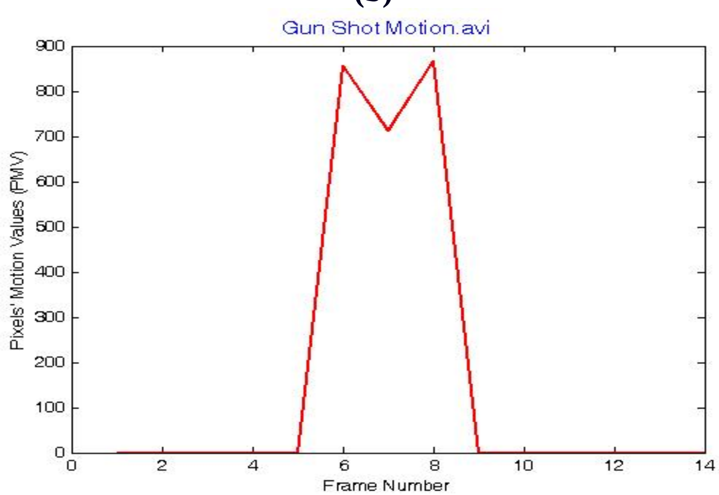
Figure C.1: "Bending down" motion (a) Extracted frames (b) Pixels' Difference Values (PDV) (c) Pixels' Motion Values (PMV)



(a)



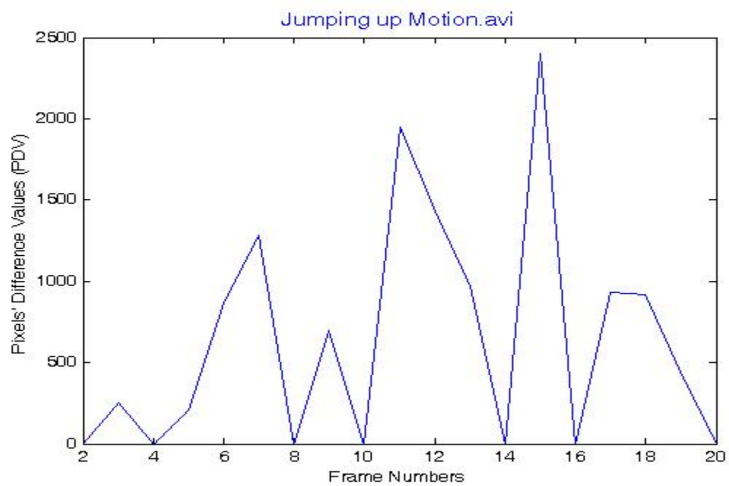
(b)



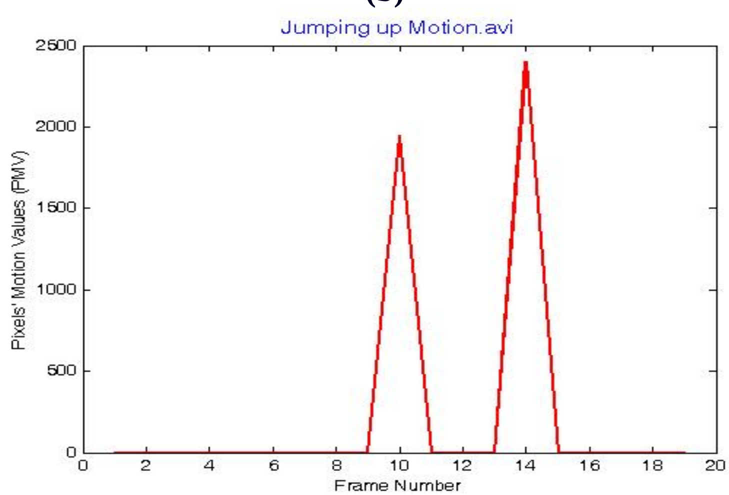
(c)

Figure C.2: "Gun Shot" motion (a) Extracted frames (b) Pixels' Difference Values (PDV) (c) Pixels' Motion Values (PMV)

(a)

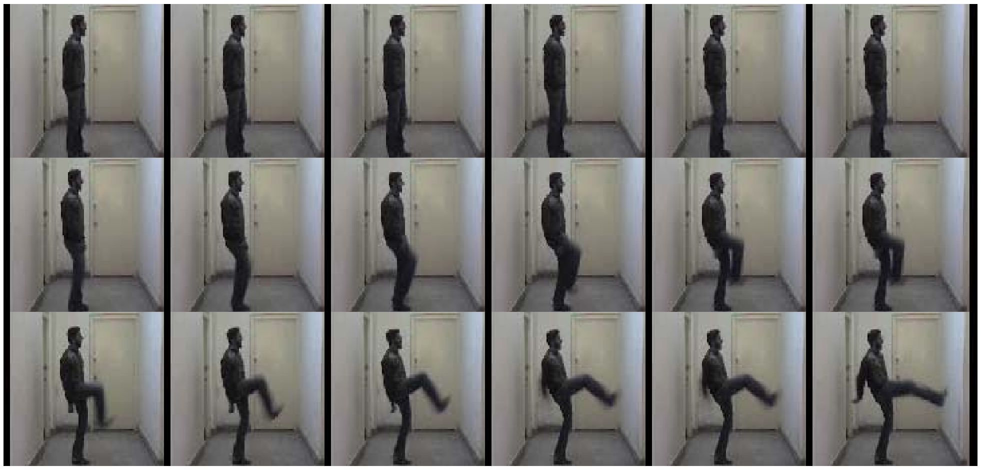


(b)

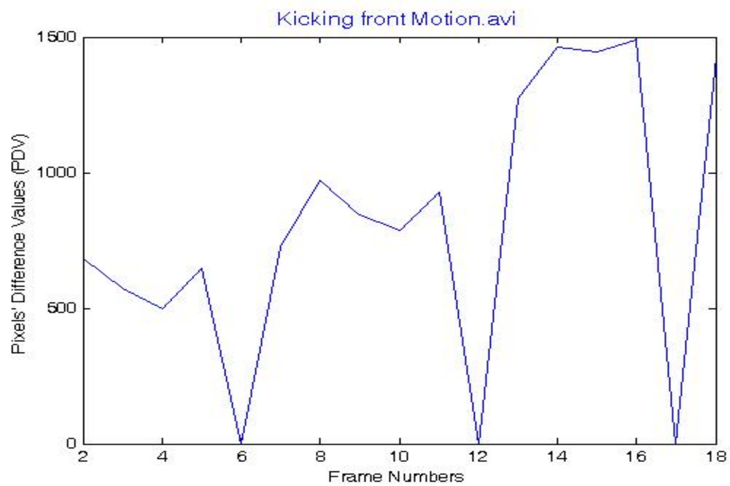


(c)

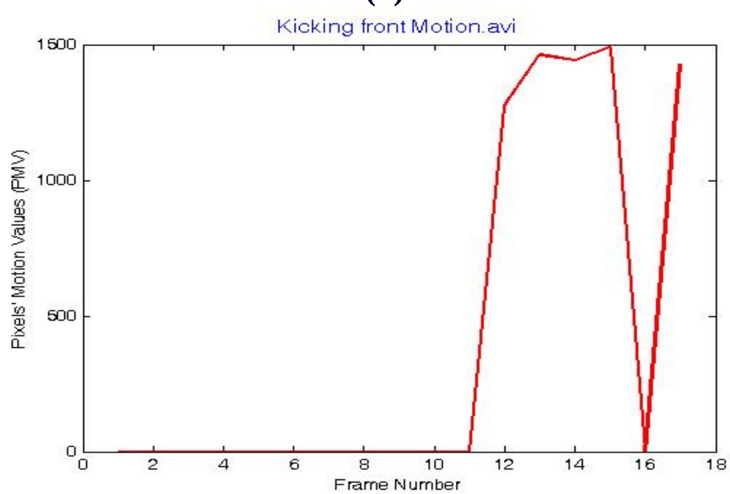
Figure C.3: "Jumping up" motion (a) Extracted frames (b) Pixels' Difference Values (PDV) (c) Pixels' Motion Values (PMV)



(a)



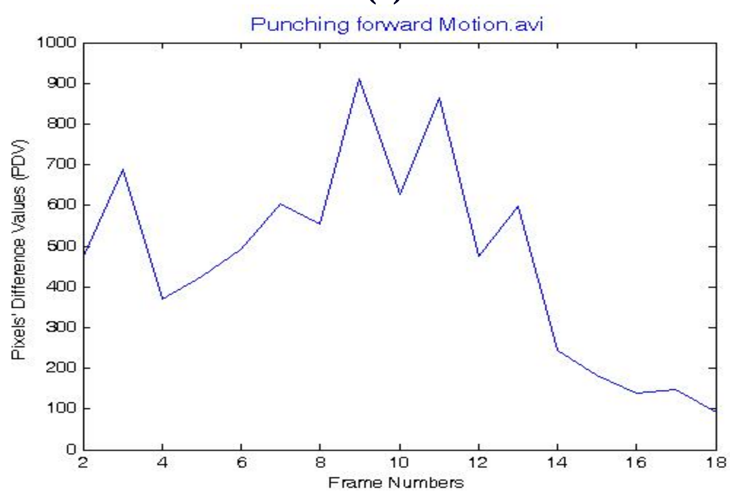
(b)



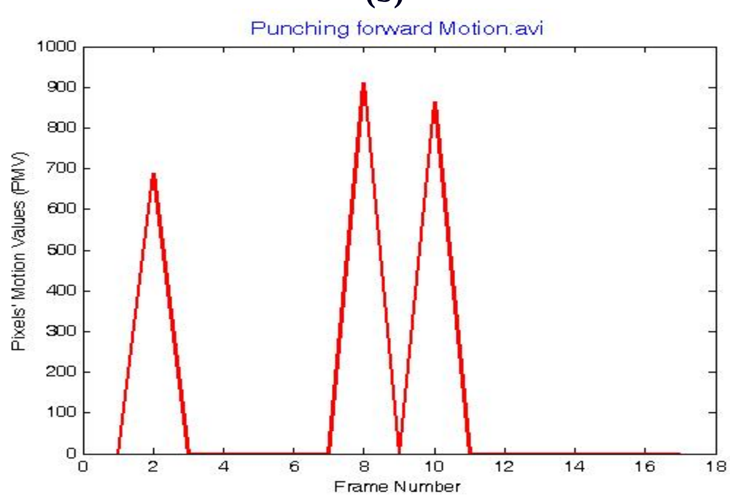
(c)

Figure C.4: "Kicking front" motion (a) Extracted frames (b) Pixels' Difference Values (PDV) (c) Pixels' Motion Values (PMV)

(a)



(b)



(c)

Figure C.5: "Punching forward" motion (a) Extracted frames (b) Pixels' Difference Values (PDV) (c) Pixels' Motion Values (PMV)