

A THESIS REPORT
ON
**Spatio Temporal Interest Keypoints and spatial
distribution gradients based HAR**

**SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR
THE AWARD OF THE DEGREE OF**

MASTER OF TECHNOLOGY

IN

SIGNAL PROCESSING AND DIGITAL DESIGN

SUBMITTED BY

JAYA GAUTAM

2K14/SPD/06

UNDER SUPERVISION OF

Dr. DINESH KUMAR VISHWAKARMA

ASSISTANT PROFESSOR



**DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY, DELHI (INDIA)**

JUNE 2016

DECLARATION

I hereby declare that the work presented in this report, titled “**Spatio Temporal Interest Keypoints and spatial distribution gradients based HAR**”, in partial fulfillment for the award of the degree of M.Tech in Signal Processing and Digital Design, submitted in the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, is original and to the best of my knowledge and belief, it has not been submitted in part or full for the award of any other degree or diploma of any other university or institute, except where due acknowledgement has been made in the text.

Jaya Gautam

Roll No. 2K14/SPD/06

M.Tech Signal Processing and Digital Design

Date:

CERTIFICATE

This is to certify the research work embodied in this dissertation entitled **“Spatio Temporal Interest Keypoints and spatial distribution gradients based HAR”** submitted by Miss Jaya Gautam, Roll no. 2K14/SPD/06 student of Master of Technology in Signal Processing and Digital Design under Department of Electronics and Communication Engineering, Delhi Technological University, Delhi is a bonafide record of the candidate’s own work carried out by her under my guidance. This work is original and has not been submitted in part or full for award of any other degree or diploma to any university or institute.

(Dr. Dinesh K. Vishwakarma)

Assistant Professor

Dept. of Electronics and Communication Engineering

Delhi Technological University, Delhi, India.

Date:

ACKNOWLEDGEMENT

I express my sincere thanks and deep sense of gratitude to my guide, **Dr. Dinesh Kumar Vishwakarma**, Assistant Professor, Department of Electronics & Communication Engineering, Delhi Technological University, whose encouragement, the initial to final level enabled me to develop an understanding of the subject. His suggestions and ways of summarizing the things made me go for independent studying and trying my best to get the maximum in my topic, this made my circle of knowledge very vast. I am highly thankful to him for guiding me in this project.

I am also grateful to Prof. Prem R. Chadha. HOD, Department of Electronics & Communication Engineering, Delhi Technological University for his immense support.

Finally, I take this opportunity to extend my deep appreciation to my family, for their endless support during the crucial times of the completion of my project.

Jaya Gautam

Roll No. 2K14/SPD/06

M.Tech Signal Processing and Digital Design

ABSTRACT

Human activity recognition is a formidable topic of machine learning and computer vision research. The aim of action recognition is to analyse the events occurring during the on-going activity from video data. A dependable HAR system is capable of recognizing human actions based upon the uniqueness of the activities and has several applications include video surveillance systems, human computer interaction which involves communication between humans and machine, content-based video annotation and retrieval, video summarization, biometrics and in health care domain.

In past decade, an expeditious proliferation of video cameras has resulted in an enormous outburst of video content. The area of analysing human activity from video data is growing faster and received rapid importance due to surveillance, security, entertainment and personal logging. The activity recognition is an area compiled with several challenges at each level of processing. The low level processing contains pre-processing challenges, robustness against errors. Mid level processing has space and time-invariant representations challenges whereas high level processing has semantic representation problems. In this work, a new hybrid technique is proposed for human action and activity recognition in video sequences. The work is demonstrated on widely used databases i.e. KTH, Weizmann, Ballet and a multi view dataset IXMAS to show the accuracy of the adopted method. The videos are segmented using texture based segmentation followed by calculating the average energy image (AEI). The extreme points are calculated from difference of Gaussians images to find the key points of AEI images. The vocabulary of these points is created

using vector quantization which is unique for each class of dataset. Then spatial distribution gradients are calculated which are combined with key point descriptors to act as a unique feature vector. These features are classified using support vector machine (SVM) and hidden markov model (HMM) for accurate recognition.

Keywords— Human activity recognition, average energy image, spatial distribution gradients, spatio temporal interest points.

Table of Contents

DECLARATION.....	i
CERTIFICATE.....	ii
ACKNOWLEDGEMENT.....	i
ABSTRACT	ii
Table of Contents	iv
LIST OF FIGURES	vii
LIST OF TABLE	ixx
CHAPTER 1.....	Error! Bookmark not defined.
INTRODUCTION.....	Error! Bookmark not defined.
1.0 OVERVIEW OF HAR.....	Error! Bookmark not defined.
1.1 MOTIVATION FOR HAR	4
1.2 FRAMEWORK OF HAR.....	5
1.3 CHALLENGES IN HAR	8
1.3.1 Variation in view point.....	8
1.3.2 Occlusion	9
1.3.3 Execution Rate	10
1.3.4 Variation in body measures	10
1.3.5 Camera motion.....	10
1.3.6 Cluttered background.....	10
1.4 OUTLINE OF THESIS.....	11

CHAPTER 2.....	12
RELATED WORK and LITERATURE REVIEW	12
2.1 Pre-Processing	15
2.1.1 Image Smoothing.....	15
2.1.2 Image Sharpening.....	15
2.1.3 Enhancement.....	16
2.2 Foreground Detection	17
2.2.1 Background Subtraction.....	17
2.2.2 Temporal average filter.....	19
2.2.3 Optical flow.....	20
2.2.4 Texture based segmentation.....	21
2.3 Feature Detector.....	22
2.3.1 Canny edge detection.....	22
2.3.2 Scale invariant feature transform.....	26
2.3.3 Harris corner detector.....	29
2.4 Feature Descriptor and Feature Representation	30
2.4.1 Histogram of oriented gradients.....	30
2.4.2 HOG 3D.....	31
2.4.3 Bag of Words.....	32
2.5 Supervised Learning	33

2.5.1 Support Vector Machine.....	34
2.5.2 Hidden Markov Model.....	37
CHAPTER 3.....	40
METHODOLOGY.....	40
3.1 Input video sequences.....	42
3.2 Silhouette Extraction.....	44
3.3 Average energy image (AEI) feature computation.....	46
3.4 SDGs Computation.....	52
3.5 Computation of Spatio Temporal Interest Keypoints.....	54
3.5.1 Scale Space.....	56
3.5.2 Keypoint localisation.....	57
3.5.3 Orientation Assignment.....	57
3.5.4 Keypoint Descriptor.....	58
3.6 Codebook Generation.....	60
3.7 Hybrid Feature Vector.....	60
CHAPTER-4.....	71
Conclusion and Future Scope.....	71
References.....	74

LIST OF FIGURES

Figure 1.1: General Human Action Recognition Framework.....	6
Figure 1.2: HAR framework used in our approach.....	7
Figure 1.3 : Walking action images from i3DPost multiview dataset.....	8
Figure 1.4 : Actions during occlusion.....	9
Figure 2.1 : Figure 2.1: Spatial filter using image sharpening.....	16
Figure 2.2 : Process showing Background subtraction.....	19
Figure 2.3 : Foreground Detection using Optical Flow.....	20
Figure 2.4 : Foreground detection using texture segmentation.....	22
Figure 2.5: The original grayscale image is smoothed with a Gaussian filter to suppress noise.....	23
Figure 2.6 The gradient magnitudes in the smoothed image as well as their directions are determined by applying Sobel operator.....	24
Figure 2.7: Non-maximum suppression. Edge-pixels are only preserved where the gradient has local maxima	25
Figure 2.8: Thresholding of edges. In the second image strong edges are white, while weak edges are grey. Edges with strength below both thresholds are suppressed.....	26
Figure 2.9: Keypoints in hand waving activity.....	29
Figure 2.10: Block Diagram of HOG.....	31
Figure 2.11: HOG3D Block Diagram.....	32
Figure 2.12: Bag of Feature representation.....	33
Figure 2.13: Support Vector Machine Hyperplane.....	34
Figure 2.14: Data points shown in cartesian coordinates, And Data points shown in polar coordinates.....	38
Figure 2.15: Trellis Diagram of HMM.....	39

Figure 3.1: Flow diagram of proposed framework.....	41
Figure 3.2: KTH Dataset.....	42
Figure 3.3: Weizmann dataset sample frames.....	43
Figure 3.4: Ballet dataset actions.....	43
Figure 3.5: IXMAS action sequences at different camera views.....	44
Figure 3.6: The workflow of silhouette extraction.....	45
Figure 3.7: Flow Diagram depicting AEI feature computation.....	48
Figure 3.8: Pixel intensity values of AEI image for hand waving activity.....	49
Figure 3.9: AEI Image Representation.....	50
Figure 3.10: AEI image of different action classes of IXMAS datasets.....	51
Figure 3.11: AEI image of check watch action class of IXMAS datasets at different camera angles.....	51
Figure 3.12: PHOG descriptor.....	52
Figure 3.13: SDGs of various action classes of KTH dataset.....	53
Figure 3.14: Interest points shown in Handwaving activity.....	54
Figure 3.15: Example of good and bad detected feature points. Video is class “running” from KTH dataset.....	55
Figure 3.16: Keypoints shown in different activity classes in KTH dataset..	59
Figure 3.17: Codebook Generation.....	60
Figure 3.18: Hybrid feature vectors for KTH dataset.....	61

LIST OF TABLE

Table 2.1. Examples of kernel functions	37
Table 3.1(a): Confusion matrix for KTH dataset by using SVM classifier.....	63
Table 3.1(b): Confusion matrix for KTH dataset by using HMM classifier.....	63
Table 3.2(a): Confusion matrix for Weizmann dataset by using SVM classifier.....	64
Table 3.2(b): Confusion matrix for Weizmann dataset by using HMM classifier.....	64
Table 3.3(a): Confusion matrix for Ballet dataset by using SVM classifier.....	65
Table 3.3(b): Confusion matrix for Ballet dataset by using HMM classifier.....	65
Table 3.4(a): Confusion matrix for IXMAS dataset by using SVM classifier...	66
Table 3.4(b): Confusion matrix for IXMAS dataset by using HMM classifier.....	67
Table 3.5: Classification accuracy for all the datasets using SVM and HMM classifier.....	68
Table 3.6: Comparison of recognition accuracy with similar state-of-the-art techniques on Weizmann Dataset.....	68
Table 3.7: Comparison of recognition accuracy with similar state-of-the-art techniques on KTH Dataset.....	69
Table 3.8: Comparison with other human action recognition approaches of the state of-the art. The accuracy obtained in the leave-one-actor-out cross validation performed on the Ballet dataset.....	70
Table 3.9: Comparison with other multi-view human action recognition approaches of the state of-the art. The accuracy obtained in the leave-one-actor-out cross validation performed on the IXMAS.....	70

