


PERFORMANCE EVALUATION OF HUMAN ACTIVITY RECOGNITION IN VIDEO SEQUENCES

Kalawitigoda Gamage Manosha Chathuramali

(138051X)

 University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
Thesis submitted in partial fulfillment of the requirements for the degree
www.lib.mrt.ac.lk
Master of Science

Department of Electronic and Telecommunication Engineering

University of Moratuwa
Sri Lanka

December 2014

Declaration

I declare that this is my own work, and this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or institute of higher learning, and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part, in print, electronic, or any other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:  University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk Date:

The candidate, whose signature appears above, carried out research for the MSc dissertation under my supervision.

Signature:

Date:

Abstract

In this thesis, we address human behavior recognition, as one of the important topics in computer vision. It finds applications in many areas such as surveillance, military installations, and sports. The problem becomes more challenging, due to the huge intra-class variation, background clutter, occlusions, illumination changes and noise. Human behavior recognition typically requires standard preprocessing steps such as motion compensation, background modeling. The errors of the motion compensation step and background modeling increase the mis-detections. We use JBFM as our background model and optic flow values to compute the motion. We propose two different spatio-temporal feature descriptors, SOF and DTF, which combine both computed motion and appearance based features. We use SVM to recognize human actions, by using different evaluation protocols (test cases). We perform several experiments and compare over a diverse set of challenging videos to address the problem, human behavior recognition by simplifying into three tasks. They are, human action recognition in stationary background, human action recognition in dynamic background, and abnormal activity recognition. Our Experimental results show that the selected framework outperforms state-of-the-art methods in many cases in terms of both recognition rate and computational complexity.

Index terms— Human action recognition, dynamic backgrounds, abnormal activity, silhouette, optic flow, SVM, HOG, HOF, MBH, dense trajectories, BoVW



University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Acknowledgements

I would like to extend my gratitude to my supervisor, Dr. Ranga Rodrigo, Department of Electronic and Telecommunication Engineering, University of Moratuwa. It is a pleasure to have known him, and I consider myself lucky for having been his student.

Additionally, I should thank my progress review panel members, Dr. Lochandaka Ranathunga and Dr. Jayathu Samarawickrama for their valuable suggestions and comments to improve the content of my thesis. I appreciate the commitment shown to conduct the timely progress reviews according to the schedule, despite their busy schedules.

Also, I should thank the Head of the Department, Dr. Ajith Pasqual, Research Coordinator, Dr. Chandika Wawegeedara, and all other members of the Department of Electronic and Telecommunication Engineering, University of Moratuwa, for their support in various matters. I specially thank Prof. Dileeka Dias for giving me the opportunity to present my work at the monthly research seminar and giving valuable suggestions to improve my work.

Further, I need to thank my fellow colleagues, Mr. Dileepa Jayamanne and Mr. Sameera Ramasinghe for their support to make the final outcome of the research more important.

Last, but not least, my warm thanks should go to my family members: My father and mother, who motivated me to start my postgraduate studies and without that none of the research work would have been realistic. I also thank my sister and husband for their valuable sacrifices during the time of my research.

Contents

Declaration	i
Abstract	ii
Acknowledgments	iii
1 INTRODUCTION	1
1.1 Problem Statement	3
1.2 Contribution	3
1.3 Thesis Overview	4
2 LITERATURE REVIEW	5
2.1 Feature Extractors	5
2.1.1 Background Modeling	7
2.1.2 Abnormal Activity Modeling	8
2.2 Feature Learning	9
3 FEATURE DESCRIPTOR	12
3.1 Silhouettes and Optic flow based Features (SOF)	12
3.1.1 Local Features	12
3.1.2 Motion Descriptor	17
3.2 Dense Trajectory Based Features (DTF)	17
3.2.1 Dense Trajectories	17
3.2.2 HOG, HOF and MBH Descriptors	18
4 METHOD: FEATURE REPRESENTATION AND LEARNING	22
4.1 Representation	22
4.1.1 Bag-of-Visual-Words (BoVW)	22
4.2 Learning	24
4.2.1 Supervised Learning	25
4.2.2 Unsupervised Learning	29

5	EXPERIMENTS, DATASETS, and RESULTS	32
5.1	Human Action Recognition	32
5.1.1	Datasets	32
5.1.2	Experimental Results	33
5.2	Dynamic Scenes	36
5.2.1	Introduction	36
5.2.2	Datasets	38
5.2.3	Experimental Results	38
5.3	Abnormal Activity Recognition	42
5.3.1	Introduction	42
5.3.2	Datasets	42
5.3.3	Experimental Results	44
6	CONCLUSIONS	50



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

List of Figures

1.1	Examples of different applications and abstractions of human behavior recognition	1
3.1	A graphical representation of silhouettes and optic flow based descriptor.	13
3.2	Various sources of dynamic behavior	14
3.3	Sample frames from dense trajectories of datasets	20
3.4	Block diagram of HOG descriptor	21
4.1	Visual-word image representation based on vector-quantized key-point features	23
4.2	The difference between Euclidean distance and Mahalanobis distance	26
4.3	Support Vector Machines	27
4.4	Data points shown in cartesian and polar coordinates.	28
5.1	Sample frames from each action in weizmann dataset	33
5.2	Sample frames from each action in UIUC1 dataset	34
5.3	Evaluation protocol	35
5.4	Comparison results of three classifiers for Weizmann and UIUC1 datasets	37
5.5	Sample frames of datasets (dynamic background/dynamic background with object)	39
5.6	Comparison of qualitative results of three datasets	43
5.7	The qualitative results of the <i>Switchlight</i> dataset	44
5.8	Sample frames from some actions in UMN dataset	45
5.9	Sample frames of datasets titled Boat-Sea, Boat-River, Traffic-Bellevue and Airport-Wrong-Direction	49

List of Tables

4.1	Different examples of kernel functions	29
5.1	The variations in the activity datasets	32
5.2	Percent recognition rates for training with regular training sets . .	35
5.3	Percent recognition rates for training with a few training examples	36
5.4	Percent recognition rates using the Weizmann Dataset	36
5.5	Action recognition rate against the four background model	40
5.6	Action recognition rates with dynamic backgrounds in different datasets	40
5.7	Description of the Boat-Sea, Boat-River, Traffic-Bellevue and Airport-Wrong-Direction Datasets	44
5.8	Recognition rates of abnormal activity using the UMN Dataset . .	46
5.9	Recognition rates of abnormal activity using the Boat-Sea, Boat-River, Traffic-Bellevue and Airport-Wrong-Direction Datasets . .	47
5.10	Computational time for training and testing	47

List of Abbreviations

Abbreviation	Description
ROI	Region of interest
MEI	Motion energy image
MHI	Motion history image
HOG	Histograms of oriented gradients
HOF	Histograms of optic flow
MBH	Motion boundary histograms
JBFM	Joint background foreground model
AMM	Approximate mean model
GMM	Gaussian mixture models
FDM	Frame difference model
3DHOG	3D-Histogram of oriented gradients
PCA	Principal component analysis
MRF	Markov random fields
HMM	Hidden markov models
DBN	Dynamic Bayesian networks
SVM	Support vector machines
SOF	Silhouettes and optic flow based features
DTF	Dense trajectory based features
N-cut	Normalized cuts
SIFT	Scale-invariant feature transforms
BoVW	Bag-of-visual-words
NN	Nearest neighbor
LMNN	Large margin nearest neighbors
RBF	Radial basis function
LOO	Leave-one-out
L1AO	Leave one actor out
L1AAO	Leave one actor-action out

L1SO	Leave one sequence out
RR	Recognition rates
FE	Few examples



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk