UNIVERSITI PUTRA MALAYSIA

SHAPE-BASED MULTI-VIEW HUMAN ACTION RECOGNITION USING DISTANCE-BASED-MATRIX-REGARDLESS-OF-ROW-PRIORITY CLASSIFIER

ALIHOSSEIN ARYANFAR

FSKTM 2016 5

SHAPE-BASED MULTI-VIEW HUMAN ACTION RECOGNITION USING DISTANCE-BASED-MATRIX-REGARDLESS-OF-ROW-PRIORITY CLASSIFIER

By

ALIHOSSEIN ARYANFAR

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

June 2016

DEDICATION

This thesis is especially dedicated
To my wife for all of her love and support
To my mother with love and eternal appreciation
To my father who could not see this thesis completed.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in Fulfilment of the requirement for the Degree of Doctor of Philosophy

SHAPE-BASED MULTI-VIEW HUMAN ACTION RECOGNITION USING DISTANCE-BASED-MATRIX-REGARDLESS-OF-ROW-PRIORITY CLASSIFIER

By

ALIHOSSEIN ARYANFAR

June 2016

Chairman : Associate Professor Razali Yaakob, PhD
Faculty : Computer Science and Information Technology

The recognition of human activity (or action) in videos has elicited significant attention in recent years given its potential use in many real-life applications. Human Action Recognition (HAR) is typically applied in fields such as human–computer interaction, surveillance, content-based video retrieval, and sports event analysis.

HAR is a complex process because characteristics such as gender, height, body shape, and age considerably affect the visual reproduction/representation of captured actions. In practical applications, changes in viewpoint are common and fundamentally unavoidable given the inherent limitations of camera technology or the inevitable dynamism of human motion. When such changes are implemented, the recognition rate of current HAR approaches dramatically decreases. This problem is typically mitigated by the use of cameras equipped with multiple fields of views, which provide richer information than that derived from single-view cameras. Even with such innovations, nonetheless, ensuring accurate correlation and acquiring multi-view learning data remain complicated challenges.

This work proposes four methods to advance the field of HAR. The Shape-based features are extracted from frames silhouette by using proposed Global Silhouette Shape Representation (GSSR) method. This GSSR is suitable given that silhouettes present spatial information on actions over time. Concatenation, as a data fusion technique, is also applied to create a multi-view feature vector from a combination of single-view feature vectors. In other words, a matrix of multi-view features is generated for each action. Maximum-Distance-among-Feature-Vectors (MDFV) technique, as a frame selection method, is employed to choose a subset of frames (or feature vectors) with the maximum difference among them. This strategy is based on the removal of frames with mostly similar features. Relevant and suitable features are selected using Binary Particle Swarm Optimization (BPSO) technique. This research likewise develops a Distance-based-Matrix-Regardless-of-Row-Priority (DMRRP) classifier, which is driven by the idea that if two action sequences depict motion performed by the same or different individuals, then the sum (or mean) of the minimum distances between each individual frame of sequence 1 and all the frames

of sequence 2 reflects the similarity between the two actions. This classifier can recognize actions captured from different views.

Finally, this study evaluates the performance of a proposed Multi-View Human Action Recognition Based On Shape-Based Feature Extraction and Distance-Based Classifier (MHARSD) in single- and multi-view HAR. To evaluate this approach, an experiment that involves two publicly available multi-view HAR datasets (i.e., MuHAVi and IXMAS) is conducted to determine the quality of recognition that the method produces for different actions. MHARSD supports the recognition of a wide range of human actions. In all evaluations, it exhibits a recognition accuracy higher than that achieved by 2D multi-view HAR state-of-the-art methods.

Abstrak tesis dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk Ijazah Doktor Falsafah

PENGECAMAN TINDAKAN MANUSIA MULTIPANDANGAN BERASASKAN BENTUK MENGGUNAKAN JARAK MATRIKS BERASASKAN PENGKELASAN TANPA MENGHIRAUKAN KEUTAMAAN BARIS

Oleh

ALIHOSSEIN ARYANFAR

Jun 2016

Pengerusi : Profesor Madya Razali Yaakob, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

Di dalam tahun kebelakangan ini, pengecaman aktiviti manusia (atau tindakan) dalam video telah diberi perhatian untuk digunakan dalam banyak aplikasi kehidupan sebenar. Pengecaman Tindakan Manusia (HAR) kebiasaannya digunakan dalam bidang seperti interaksi manusia-komputer, pengawasan, dapatan semula video berasaskan-kandungan dan analisa acara sukan.

HAR adalah proses yang kompleks kerana ciri-ciri seperti jantina, ketinggian, bentuk badan dan umur secara langsung memberi kesan kepada penghasilan-semula atau perwakilan-semula visual tindakan yang dirakam. Dalam aplikasi yang sebenar, perubahan dalam sudut pandangan adalah perkara biasa dan secara asasnya tidak dapat dielakkan disebabkan oleh had yang terdapat dalam teknologi kamera atau pergerakan manusia yang tidak dinamik. Apabila perubahan ini dilaksanakan, kadar pengecaman bagi pendekatan HAR semasa akan berkurangan secara dramatik. Masalah ini biasanya diatasi dengan penggunaan kamera yang dilengkapi dengan pelbagai bidang pandangan, yang memberikan lebih banyak maklumat daripada kamera pandangan-tunggal. Walaupun dengan inovasi tersebut, memastikan korelasi yang tepat dan mendapatkan data pembelajaran pandangan-pelbagai masih lagi satu cabaran yang rumit.

Kajian ini mencadangkan empat kaedah untuk menambahbaik bidang HAR. Ciri-ciri berasaskan-bentuk diekstrak daripada bebayang dalam bingkai dengan menggunakan kaedah Perwakilan Bentuk Bebayang Global (GSSR) yang dicadangkan. GSSR ini adalah sesuai diberikan bebayang mewakili maklumat ruang ke atas tindakan dari masa ke masa. Penggabungan, sebagai teknik menggabungkan data, juga digunakan untuk membina satu vektor ciri pandangan-pelbagai daripada gabungan vektor-vektor ciri pandangan-tunggal. Dengan kata lain, ciri-ciri matriks pandangan-pelbagai dihasilkan untuk setiap tindakan. Teknik Maksimum-Jarak-antara-Ciri-Vektor (MDFV), sebagai kaedah pemilihan bingkai, digunakan untuk memilih subset bingkai (atau vektor ciri) dengan perbezaan yang besar di antara mereka. Strategi ini adalah berdasarkan kepada penyingkiran bingkai dengan kebanyakan

ciri-cirinya hampir sama. Ciri-ciri yang berkaitan dan sesuai dipilih menggunakan teknik Pengoptimuman Gugusan Zarah Binari (BPSO). Kajian ini juga membangunkan pengkelas Matriks Berasaskan Jarak Tanpa Menghiraukan Keutamaan Baris (DMRP), yang didorong oleh idea bahawa jika dua jujukan tindakan yang menggambarkan gerakan yang dilakukan oleh individu yang sama atau berbeza, maka jumlah (atau purata) jarak minimum di antara setiap bingkai individu jujukan 1 dan semua bingkai jujukan 2 memberikan kesan persamaan antara kedua-dua tindakan. Pengkelas ini boleh mengecam tindakan yang dirakam dari pandangan yang berbeza.

Akhir sekali, kajian ini menilai prestasi Pengecaman Tindakan Manusia Pandangan-pelbagai berasaskan Pengekstrakan Ciri Berasaskan Bentuk dan Pengkelasan Berasaskan Jarak (MHARSD) yang dicadangkan dalam pandangan-tunggal dan pelbagai HAR. Untuk menilai pendekatan ini, satu eksperimen yang melibatkan dua set data pandangan-pelbagai HAR umum (iaitu MuHAVi dan IXMAS) dijalankan untuk menentukan kualiti pengecaman yang dihasilkan bagi tindakan yang berbeza. MHARSD menyokong pengecaman pelbagai tindakan manusia. Dalam semua penilaian, ia mempamerkan ketepatan pengecaman lebih tinggi daripada yang telah dicapai oleh kaedah HAR pandangan-pelbagai 2D yang terkini.

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Razali Yaakob for the continuous support of my Ph.D. study, research and scholarship, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. Besides my supervisor, I would like to thank the rest of my thesis committee: Dr. Alfian Abdul Halin, and Assoc. Prof. Md. Nasir Sulaiman, for their encouragement, insightful comments, patience, motivation, and immense knowledge. I am also thankful to the faculty of computer science and information technology, UPM.

I also owe a debt of gratitude to my wife Leila Mohammadpour, who was always there for me throughout all my struggles and frustrations. The constant emotional support I received from her was vital in the completion of my study. Last but not least, I would like to thank my Mother for their unconditional emotional support, throughout my study. In particular, I greatly appreciate my mom's patience and understanding during these years. I thank my father's holy soul who wisely taught me how to be resistance in difficult situations in life.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Razali Yaakob, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Alfian Abdul Halin, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Md. Nasir Sulaiman, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by Graduate Student

I hereby confirm that:

- This thesis is my original work;
- Quotation, illustrations and citations have been duly referenced;
- This thesis has not been submitted previously or concurrently for any other degree at any other instructions;
- Intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- Written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journal, modules, proceeding, popular writing, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in Universiti Putra Malaysia (Research) Rules 2012;
- There is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ________________________ Date: ________________________

Name and Matric No : Alihossein Aryanfar GS31661

Declaration by Member of Supervisory Committee

This is to confirm that:

- The research conducted and the writing of this thesis was under our supervision;
- Supervision responsible as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________

Name of
Chairman of
Supervisory : Associate Professor
Committee : Dr. Razali Yaakob

Signature: ______________________

Name of
Chairman of
Supervisory
Committee : Dr. Alfian Abdul Halin

Signature: ______________________

Name of
Chairman of
Supervisory Associate Professor
Committee : Dr. Md. Nasir Sulaiman

TABLE OF CONTENTS

			Page
ABSTRACT			i
ABSTRAK			iii
ACKNOWLEDGEMENTS			v
APPROVAL			vi
DECLARATION			viii
LIST OF TABLES			xiii
LIST OF FIGURES			xiv
LIST OF APPENDICES			xvi
LIST OF ABBREVIATIONS			xvii
CHAPTER			
1	**INTRODUCTION**		1
	1.1	Background and Motivation	1
	1.2	Problem Statements	2
	1.3	Hypotheses	4
	1.4	Objectives	4
	1.5	Contributions	5
	1.6	Research Scope and Assumptions	5
	1.7	Thesis Organization	6
2	**LITERATURE REVIEW**		7
	2.1	Introduction	7
	2.2	Motions, Events, Actions, and Activities	7
	2.3	Single-view and Multi-view Human Action Recognition	8
	2.4	Human Action Recognition Approaches Categorization	9
	2.5	Human Action Recognition Phases	10
	2.6	Feature Extraction	11
		2.6.1 Global Features	12
		2.6.2 Local Features	14
		2.6.3 Hybrid features	15
		2.6.4 Feature Fusion	15
	2.7	Video Description	16
	2.8	Classification	17
		2.8.1 Exemplar-based Method	17
		2.8.2 Generative Models	18
		2.8.3 Discriminative classifiers	18
	2.9	Datasets and Benchmarks	19
		2.9.1 IXMAS Dataset	21
		2.9.2 MuHAVi Dataset	22
		2.9.3 Weizmann Dataset	24
	2.10	Related Works	25
	2.11	Summary	33
3	**RESEARCH METHODOLOGY**		35
	3.1	Introduction	35
	3.2	Research Methodology Steps	35

	3.3	Literature Review	36
	3.4	General Human Action Recognition Framework	36
	3.5	An Approach Based on Keyposes, 2D Wavelet and C5.0 Classifier	37
	3.6	An Approach Based on Distance Signal, 2D Wavelet and Hierarchical Classifier	38
	3.7	An Approach Based on PMBI, PSO, and Hierarchical SVM	41
	3.8	MHARSD Approach	44
		3.8.1 Proposed framework for approach MHARSD	44
	3.9	Summary	45
4	**MULTI-VIEW HAR USING MHARSD**		46
	4.1	Introduction	46
	4.2	Proposed Framework for Human Action Recognition	46
	4.3	Feature Extraction	48
		4.3.1 Silhouette-Based Feature Extraction	51
		4.3.2 Shape-Based Feature Extraction	51
		4.3.3 Global Silhouette Shape-Representation (GSSR)	53
	4.4	Feature fusion	63
	4.5	Distance Metrics	65
		4.5.1 Distance Metrics for Matrix	67
	4.6	Keyframe Selection	67
		4.6.1 Keyframe Selection Algorithm	68
	4.7	Feature Subset Selection	71
		4.7.1 Evolutionary Feature Subset Selection	72
		4.7.2 Feature Subset Selection Using BPSO	72
	4.8	Classification	75
		4.8.1 Multi-View Human Action Recognition based on DMRRP	76
	4.9	Summary	81
5	**RESULTS AND DISCUSSION**		82
	5.1	Introduction	82
	5.2	Evaluation Methodology and Technical Details	82
	5.3	Result of Initial Approaches (KWCC, DSWHC, and PPHS)	83
	5.4	Evaluation of Various Distance Measurement Techniques	84
	5.5	Evaluation of Various Frame Subset Selection Methods	86
	5.6	Results of the MHARSD on MuHAVi-8	88
		5.6.1 Result of Optimum Value for Number of Keyframes	88
		5.6.2 Preliminary Result	89
		5.6.3 Final Result on MuHAVi-8	93
	5.7	Results of the MHARSD on MuHAVi-14	94
		5.7.1 Result of Optimum Value for Number of Keyframes	94
		5.7.2 Preliminary Result on MuHAVi-14	96

	5.7.3	Final Results on MuHAVi-14	100
5.8		Results of the MHARSD on IXMAS	102
	5.8.1	Result of Optimum Value for Number of keyframes	103
	5.8.2	Preliminary Result on IXMAS	104
	5.8.3	Final Result on IXMAS	107
5.9		Complicated Results of IXMAS, From Single-View to Multi-View	110
5.10		Comparison with State-of-the-Art Methods	113
5.11		Summary	114

6 CONCLUSION AND FUTURE WORKS 116
6.1 Conclusion 116
6.2 Future Works 117

REFERENCES 119
APPENDICES 133
BIODATA OF STUDENT 154
LIST OF PUBLICATIONS 155

LIST OF TABLES

Table		Page
2.1	Public action datasets [82]	20
2.2	Critical review of relative methods	31
2.3	Recognition accuracy of current methods on IXMAS	32
2.4.	Recognition accuracy of current methods on MuHAVi	33
5.1	The accuracy of recognition (%) for initial approaches	83
5.2	Result of DMT on MuHAVi-8 using LOSO-CV	84
5.3	Result of DMT on MuHAVi-8 using LOAO-CV	85
5.4	Result of DMT on MuHAVi-14 using LOSO-CV	85
5.5	Result of DMT on MuHAVi-14 using LOAO-CV	86
5.6	Best result and best keyframe number on MuHAVi-8	88
5.7	Preliminary confusion matrix (table) on MuHAVi-8	92
5.8	Confusion matrix (table) on the MuHAVi-8	94
5.9	Best result and best keyframe number on MuHAVi-14	96
5.10	Confusion matrix of LOSO-CV on the MuHAVi-14 dataset	98
5.11	Confusion matrix of LOAO-CV on the MuHAVi-14 dataset	99
5.12	Confusion matrix (v) on MuHAVi-14	101
5.13	Confusion matrix (table) on MuHAVi-14 (camera 3 LOAO-CV)	102
5.14	Best result and best keyframe number on IXMAS	104
5.15	Confusion matrix of preliminary result on the IXMAS-11	106
5.16	Confusion matrix of preliminary result on IXMAS-13	107
5.17	Comparison of preliminary and final results on IXMAS	108
5.18	Confusion matrix of final result on IXMAS-11	109
5.19	Confusion matrix of final result on IXMAS-13	110
5.20	Preliminary and final results on IXMAS (11 and 13 actions)	112
5.21	Comparison with similar approaches on the MuHAVi -8	113
5.22	Comparison with similar approaches on MuHAVi -14	113
5.23	Comparison with similar approaches on IXMAS (four view angles)	114
5.24	Comparison with similar approaches on IXMAS (five view angles)	114

LIST OF FIGURES

Figure		Page
1.1	Some applications of HAR [1]	2
2.1	The categories and sub-categories of methods on HAR [27]	9
2.2	Overview of action modelling categorization [25]	10
2.3	The primary stages in human action recognition	11
2.4	Sample images of IXMAS dataset	21
2.5	Sample of IXMAS images with different view angle	22
2.6	Sample images of MuHAVi dataset	23
2.7	Sample images of and extracted silhouettes from Weizmann	24
3.1	The general steps of research methodology	35
3.2	The stages of KWCC	37
3.3	The stages of DSWHC	38
3.4	First steps of feature extraction, (a) sample of images in dataset IXMAS, (b) extracted silhouette, (c) cropped silhouette include holes, (d) after removing holes and (e) extract the borders.	39
3.5	(a) GDA Algorithm to separate 12 actions into two-category, (b) The structure of the proposed classifier	40
3.6	The stages of PPHS	41
3.7	Proposed hierarchical classifier	43
3.8	Proposed MHARSD approach	44
4.1	Framework of the MHARSD	47
4.2	Overall techniques used in the MHARSD	48
4.3	Example of RGB images, silhouettes and ROI from different views	49
4.4	Example of image, depth information, and its silhouette	50
4.5	Structure of shape representation and sample on features [143]	52
4.6	Feature extraction phases using GSSR	54
4.7	Example of a centroid point on a contour and silhouette	56
4.8	Example of a bounding box on a silhouette and contour points	57
4.9	Example of a main ellipse, major and minor axes	59
4.10	Intervals of angular variation	60
4.11	Axes and orientation of an ellipse	61
4.12	Silhouette and smallest convex hull (red)	62
4.13	Silhouette, equivalent circle, and their diameters	63
4.14	Example of feature fusion steps	65
4.15	Flowchart of MDFV	69
4.16	MDFV algorithm	70
4.17	Sample of MDFV-selected keyframes at different thresholds	71
4.18	BPSO-based feature subset selection	74
4.19	Fitness function for BPSO-based feature subset selection	75
4.20	Extracted features from sequence 1 and 2	77
4.21	Comparison of and distance between sequences 1 and 2	78
4.22	Matching the frames of sequence1 and 2	78
4.23	Proposed DMRRP classifier	79
4.24	Flowchart of DMRRP	80
5.1	Comparison of keyframe selection methods on MuHAVi-8	87

5.2	Comparison of keyframe selection methods on MuHAVi-14	87
5.3	Accuracy variation based on the keyframe numbers on MuHAVi-8	89
5.4	Preliminary confusion matrix (heatmap) on MuHAVi-8	91
5.5	Confusion matrix (heatmap) on MuHAVi-8	93
5.6	Accuracy variations according to keyframe number on MuHAVi-8	95
5.7	Preliminary confusion matrix (heatmap) on MuHAVi-14	97
5.8	Confusion matrix (heat map) on MuHAVi-14 (multi-view and camera 4 using LOSO-CV and LOAO-CV; camera 3 using LOSO-CV)	100
5.9	Confusion matrix (heat map) on MuHAVi-14 (camera 3 LOAO-CV)	101
5.10	Accuracy variations according to keyframe number on IXMAS	103
5.11	Confusion matrix of the preliminary results on IXMAS	105
5.12	Confusion matrix of the final results on IXMAS	108

LIST OF APPENDICES

Appendix		Page
A	Introduction	135
B	Result of BPSO on MuHAVi-8	135
C	Result of BPSO on MuHAVi-14	141
D	Result of BPSO on IXMAS	149

LIST OF ABBREVIATIONS

APF	Annealed Particle Filtering
BOF	Bag of Features
BOW	Bag of Words
BPSO	Binary Particle Swarm Optimization
CBP	Correlogram of Body Poses
DFT	Discrete Fourier Transform
DMRRP	Distance based Matrix Regardless of Row Priority
DSWHC	Distance Signal and 2D Wavelet and Hierarchical Classifier
DTW	Dynamic Time Warping
DMT	Distance Measurement Techniques
EA	Evolutionary Algorithms
FV	Feature Vector
GMM	Gaussian Mixture Models
GSSR	Global Silhouette Shape-Representation
KWCC	Keyposes and 2D Wavelet and C 5.0 Classifier
HAR	Human Action Recognition
HBP	Histogram of Body Poses
HMM	Hidden Markov Model
HOG	Histogram of Oriented Gradients
HOR	Histogram of Oriented Rectangles
ISI	Institute for Scientific Information
MBOW	Multi-View Bag of Words
MDFV	Maximum-Distance-among-the-Feature-Vectors
MEI	Binary Motion Energy Image
MHARSD	Multi-View Human Action Recognition Based On Shape-Based Feature Extraction and Distance-Based Classifier
MHI	Motion History Image
PCA	Principal Component Analysis
PPHS	PMBI, PSO, and Hierarchical SVM
ROI	Region of Interest
SC	Sparse Coding
SVM	Support Vector Machines
SVQ	Soft Vector Quantization
VQ	Vector Quantization
STIP	Spatio-Temporal Interest Points

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

Visual perception is the ability to detect and process visible light to interpret the surrounding environment. Human has remarkable ability to perceive human actions purely from visual information. We can localize people and objects, track articulated human motions, and analyze human-object interactions to understand what people are doing and even infer their intents.

In order to perform human action recognition (HAR), vision is probably the most valuable sensor information that can be employed. In the other hand, computer vision is the field of science that aims to duplicate the abilities of the human visual system by electronically perceiving and understanding real world imagery. It includes methods for acquiring, processing, analyzing, and understanding images, videos, and other high-dimensional data from the real world in order to produce numerical or symbolic information.

Action recognition technology aims to recognize the actions and goals of an agent from a sequence of observations of the agent's behavior and the environmental conditions. Actions can be captured using one camera which is called single-view or using multiple cameras which is called multi-view.

Using multiple cameras present effective solutions to the problems related to the limited view. It provides wider view range compared to using single-view on the objects appearing in overlapping camera regions.

Human action recognition and behavior analyses stands out as particularly important, since it can lead to many practical applications in human life. The volume and complexity of generated visual information far exceeds the capacity of human operators to manage, analyze and respond in real-time. It is, therefore, critical to develop efficient and effective automatic methods to analyze video data with the goal of understanding the visual environment. People are doing such analysis every day to find out where other people are, where they are coming from or heading to, how they are moving, and what or who they are interacting with. Through this analysis we are able to identify friendly, suspicious, or threatening behavior, identify social relations among people, get a first impression of the emotional state of other people, and react appropriately to all this information.

A successful analysis of human motion facilitates numerous applications within a broad range of research and industrial fields.

Figure 1.1, shows a brief key applications of human activity recognition [1]. Psychologists have investigated the correlation between how people move and their emotional state as a diagnostic tool. Neuroscientists have investigated how people infer the intension of human motion to understand the cognitive processes of human behavior. Within clinical diagnostics human motion is analyzed to diagnose a number of diseases. In sports human motion is analyzed to optimize performance. The entertainment industry is producing motion analysis tools to allow for new interaction methods with game consoles and to capture human motion for computer graphics special effects.

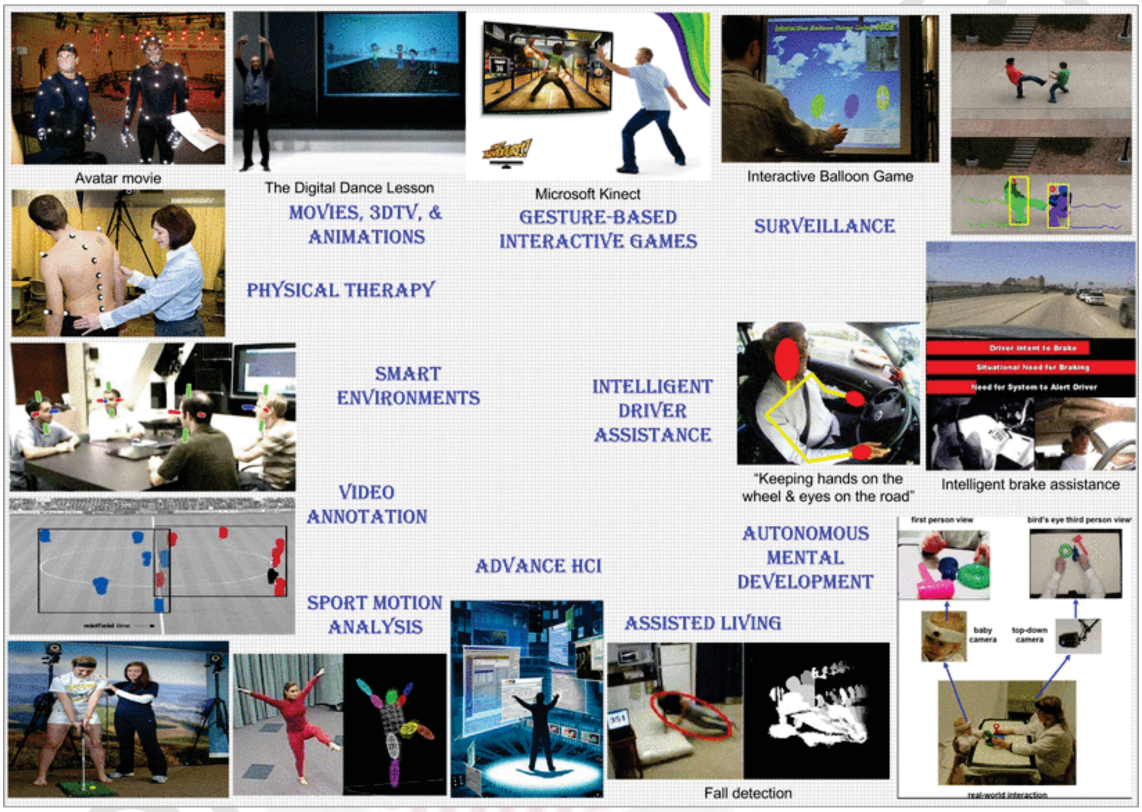

Figure 1.1: Some applications of HAR [1]

Other application include: automated surveillance systems, content based indexing and searching on Web, healthcare monitoring, and human-machine interaction in intelligent environments, robotics, video indexing, and querying, behavioral biometrics (e.g., gait is an emergent biometric aimed essentially to recognize people by the way they walk), bio-mechanics, medicine, monitoring (i.e., patients, elderly people, babies or kids), sports analysis, video entertainment, video games, intelligent transport systems for safety driving.

1.2 Problem Statements

The important research challenges encountered in HAR can be listed out as the occlusion, camera motion and cluttered background, multi-view learning and processing time, and variation in action performing speed and viewpoint [2]. The effects of the partial occlusion can be decreased by using techniques such as the space-time volume[3], or applying multiple camera setup [4], [5]. To overcome the variation in action performing speed, scholars treat temporal variations as intra-class differences [2]. Almost all available benchmark datasets are captured with stationary cameras [2], and many applications operate under the assumption of a static background [6], [7]. Therefore, challenges arising from camera motion and cluttered backgrounds are excluded from this thesis.

An action captured from different viewpoints may significantly differ in appearance. Difficulties arise from perspective distortions and differing viewpoints because dynamic human body movements can be represented in almost unlimited ways [8], [9]. Viewpoint variations can substantially reduce recognition performances [10], [11]. Although multi-views can provide more information than single views [8], [9], [11]–[15], determining a suitable correlation and effectively processing huge data continue to challenge multi-view approaches [8], [9], [11]–[13]. In addition, heavy load and high computational costs are two crucial issues that arise from multi-view data [16], [17]. In the other hand, in multi-view HAR, the features extracted from different view angles should be combined to produce a vector of features. The major problem from which most HAR methods suffer is dimensionality, which reduces recognition accuracy [18].

As indicated in neurobiological studies [19], our brain recognizes human actions by observing only human body postures during action execution. Actions can also be defined on the basis of human silhouettes [18], [20], [21]. Under this approach, a logical consecutive sequence is not always required. This principle underlies dynamic time warping (DTW) [22], which is successfully applied in speech recognition and employed in HAR. The drawback to DTW is that variations in frame rate, execution rate, and/or frame subsampling cause certain classification problems.

The three key phases of general HAR are feature detection, video description, and classification [23]. With respect to these phases, the challenges encountered in multi-view learning, processing time, and viewpoint variation are as follows:

- Current feature extraction methods for multi-view HAR extract high-dimensional features.
- Existing methods of keyframe selection misidentify similar frames as key frames. That is, methods that eliminate repetitive and very closely consecutive frames for multi-view HAR are lacking.
- Irrelevant and redundant information from extracted features degrade the recognition accuracy of classifiers.
- Current classification methods for multi-view HAR fail to achieve satisfactory recognition accuracy.

1.3 Hypotheses

In order to setup the procedure of efficient multi-view HAR the following hypothesis are considered.

- Shape-based feature detection and description method, which are widely used in computer vision especially in object recognition, can be adapted for HAR, in order to appropriately extract the low-dimensional principle characteristics information of actions. The concatenation of human silhouette features that are extracted from different views make for an effective low-cost representation of human body shapes; this representation is independent of camera calibration [18].
- New advent data acquisition systems are able to capture the huge number of frames per second; these frames are very similar to each other. Due to process all available frames at the recognition (classification) phase lead to high computational cost, without considerable affect to final results, therefore a frames subset selection technique can be applied to decrease the computational cost and increase the recognition accuracy by removing the very close consecutive frames.
- Due to the fact that extracted features include irrelevant and redundant information, they have negative affect on the accuracy of recognition. By choosing a relevant subset of features the performance of the recognition would be improved. Optimization techniques can be adopted for feature subset selection from multi-view feature vectors.
- Classifiers have a significant effect on the recognition rate. Therefore, an appropriate classifier is able to differentiate the different actions more appropriately.

1.4 Objectives

This thesis is principally concerned to address the highlighted issues in multi-view HAR to propose a multi-view recognition approach based on distance-based-matrix-regardless-of-row-priority classifier that distinguishes actions by measuring the similarities among the selected shape-based features in an action matrix. They include as follows:

- To propose a shape-based feature detection and description method that appropriately extracts information on the low-dimensional feature.
- To propose a keyframe selection technique that removes repetitive and very closely consecutive frames. Removing similar frames is accomplished by measuring the distance among extracted features from frames.
- To adapt an optimization algorithm that can select a subset of relevant features because of improvements in classifier performance. The optimization technique identifies the features that are necessary to the effective recognition of actions.
- To propose a distance of matrix based regardless row priority classifier that appropriately differentiates dissimilar actions.

1.5 Contributions

A Multi-view HAR approach based on Shape-based feature extraction and Distance-based classifier is proposed (MHARSD). This research contributes to the existing literature in a number of ways:

- This thesis puts forward a global silhouette shape representation (GSSR) method that extracts the 11 shape-based features reflected by a human body's silhouette. GSSR considers only a few features in each feature vector, but this deficiency is compensated by a proposed multi-view HAR based on shape-based feature extraction and a classifier based on distance. In contrast to many methods that require numerous features in evaluation, MHARSD effectively recognizes actions with only a small number of features.
- Actions are performed at different durations, and most of them (e.g., walking or running) are periodic movements. The speed with which action is performed, the duration of performance, and the time at which action is initiated affect the number of frames for each action. To reduce this negative influence and derive a unique number of frames, we develop a maximum-distance-among-feature-vectors (MDFV) method, which selects subset frames (keyframes) from all frames. This strategy is based on the elimination of frames with mostly similar features. In the other words, the second norm of the feature vectors extracted from each frame via other frames is calculated, after which similar frames are eliminated.
- Binary particle swarm optimization (BPSO) is employed on extracted features to improve the performance of the DMRRP classifier. BPSO identifies the relevant features that are necessary to the effective recognition of actions.
- To distinguish actions by measuring the similarities among the features in a matrix of selected features, we propose a distance-based-matrix-regardless-of-row-priority (DMRRP) classifier. DMRRP operates under the principle that if two action sequences depict similar movements performed by the same or different individuals, then the sum (mean) of the minimum distances between each individual frame of sequence 1 and all the frames of sequence 2 reflects the similarity (distance) between the two actions. That is, the sequence of matrix rows is unimportant. Another advantage presented by this classifier is that it does not require time for training.

1.6 Research Scope and Assumptions

Human motion analyses and action recognition are broad research fields. With these considerations in mind, the following assumptions are adopted:

- Background-related challenges are beyond the scope of this thesis, and the extracted silhouette, which available by the datasets are chosen for this work.
- A single individual, who is also the focal point in an image, is being monitored.
- The relationship among human objects, group action detection, and cross-view action recognition are beyond the scope of this thesis.

- A given image reflects a sufficient proportion of the human body, thereby enabling distinguishing postures throughout the entire course of an action.
- Continuous HAR is not under consideration of the scope of this work.

Although the first assumption is inapplicable to some natural videos, in many cases, multi-view cameras are fixed and background subtraction is easy to accomplish. Other assumptions can be formulated, but these are determined by data acquisition and the employed technical setup. For instance, image quality and resolution, as well as distance of cameras to a subject, are relevant in background subtraction, although specific requirements depend on the used method. Furthermore, several related difficulties can be solved with advanced devices that enable human detection and depth-based segmentation (e.g., RGB-D sensors).

The implementation is developed using Matlab software, which is frequently used by researchers. The approach adopted in this work is evaluated on the basis of the IXMAS, MuHAVi-8, and MuHAVi-14 datasets, which are the most complex and popular in multi-view HAR communities. Finally, popular cross-validation techniques (LOAO-CV and LOSO-CV) are used to validate the proposed MHARSD approach, the percentage of average from correct recognitions is considered the final result.

1.7 Thesis Organization

The thesis is organized in accordance with the standard thesis and dissertation structure prescribed by Universiti Putra Malaysia and comprises six chapters. Chapter 1 presents the research background, including the problem statement, objectives, scope, and contributions. Chapter 2 provides an overview of related research on HAR in videos. A general taxonomy is presented, and approaches relevant to ours are discussed in appropriate detail in succeeding chapters. Chapter 3 describes the research methodology, and Chapter 4 explains the proposed core approach (MHARSD). Focus is directed toward GSSR, MDFV, and DMRRP. Chapter 5 presents the experimental results and the evaluation parameters of the core approach. MHARSD is evaluated using popular multi-view datasets, namely, IXMAS and MuHAVi. This chapter also presents the interpretation of results and a comparison of our methods with similar 2D approaches. Chapter 6 provides the conclusions drawn from the research and discusses possible future directions.

REFERENCES

M. B. Holte, C. Tran, M. M. Trivedi, and T. B. Moeslund, "Human pose estimation and activity recognition from multi-view videos: Comparative explorations of recent developments," *IEEE Journal on Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 538–552, 2012.

M. Ramanathan, W.-Y. Yau, and E. K. Teoh, "Human Action Recognition With Video Data: Research and Evaluation Challenges," *Ieee Transactions on Human-Machine Systems*, vol. 44, no. 5, pp. 650–663, Oct. 2014.

L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, vol. 29, no. 12, pp. 2247–2253.

A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, Mar. 2001.

Y. Lu, Y. Li, Y. Shen, F. Ding, X. Wang, J. Hu, and S. Ding, "A Human Action Recognition Method Based on Tchebichef Moment Invariants and Temporal Templates," in *2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2012, pp. 76–79.

H. Ragheb, S. Velastin, P. Remagnino, and T. Ellis, "Human action recognition using robust power spectrum features," in *2008 15th IEEE International Conference on Image Processing*, 2008, pp. 753–756.

A. Iosifidis, A. Tefas, and I. Pitas, "Neural representation and learning for multi-view human action recognition," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1–6.

A. A. Liu, N. Xu, Y. T. Su, H. Lin, T. Hao, and Z. X. Yang, "Single/multi-view human action recognition via regularized multi-task learning," *Neurocomputing*, vol. 151, no. P2, pp. 544–553, Mar. 2015.

W. Zhou and Z. Zhang, "Human Action Recognition with Multiple-Instance Markov Model," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 10, pp. 1581–1591, 2014.

A. Iosifidis, A. Tefas, and I. Pitas, "Multi-view human action recognition: A survey," in *Proceedings - 2013 9th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2013*, 2013, pp. 522–525.

F. Zhu, L. Shao, and M. Lin, "Multi-view action recognition using local similarity random forests and sensor fusion," *Pattern Recognition Letters*, vol. 34, no. 1, pp. 20–24, Jan. 2013.

W. Liu and D. Tao, “Multiview hessian regularization for image annotation,” *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2676–2687, 2013.

M. Wang, Y. Gao, K. Lu, and Y. Rui, “View-based discriminative probabilistic modeling for 3D object retrieval and recognition,” *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1395–1407, 2013.

R. Ji, L. Y. Duan, J. Chen, T. Huang, and W. Gao, “Mining compact bag-of-patterns for low bit rate mobile visual search,” *IEEE Transactions on Image Processing*, vol. 23, no. 7, pp. 3099–3113, 2014.

Y. Yang, Q. Liu, R. Ji, and Y. Gao, “Remote Dynamic Three-Dimensional Scene Reconstruction,” *PLoS ONE*, vol. 8, no. 5, p. e55586, 2013.

M. B. Holte, B. Chakraborty, J. Gonz??lez, and T. B. Moeslund, “A local 3-D motion descriptor for multi-view human action recognition from 4-D sSpatio-temporal interest Points,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 553–565, 2012.

T. B. Moeslund, A. Hilton, and V. Kr??ger, “A survey of advances in vision-based human motion capture and analysis,” *Computer Vision and Image Understanding*, vol. 104, no. 2–3 SPEC. ISS., pp. 90–126, Nov. 2006.

A. Iosifidis, A. Tefas, and I. Pitas, “Multi-view action recognition based on action volumes, fuzzy distances and cluster discriminant analysis,” *Signal Processing*, vol. 93, no. 6, pp. 1445–1457, Jun. 2013.

M. a Giese and T. Poggio, “Cognitive neuroscience: Neural mechanisms for the recognition of biological movements,” *Nature Reviews Neuroscience*, vol. 4, no. 3, pp. 179–192, 2003.

N. Gkalelis, A. Tefas, and I. Pitas, “Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1511–1521, 2008.

L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, “Actions as space-time shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2247–2253, 2007.

D. Berndt and J. Clifford, “Using Dynamic Time Warping to Find Patterns in Time Series,” in *12th Int’l Conference on Artificial Intelligence*, 1994, vol. 10, no. 16, pp. 229–248.

H. Xu, Q. Tian, Z. Wang, and J. Wu, “A survey on aggregating methods for action recognition with dense trajectories,” *Multimedia Tools and Applications*, pp. 1–17, 2015.

B. Liebchen and P. Schmelcher, “Interaction induced directed transport in ac-driven periodic potentials,” *New Journal of Physics*, vol. 17, no. 8, pp. 1–30, 2015.

P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008.

S. Hongeng, F. Bremond, and R. Nevatia, "Representation and optimal recognition of human activities," in *Ieee Conference on Computer Vision and Pattern Recognition, Proceedings, Vol I*, 2000, vol. 1, pp. 818–825.

J. K. Aggarwal and M. S. Ryoo, "Human activity analysis," *ACM Comput. Surv.*, vol. 43, no. 3, pp. 1–43, Apr. 2011.

A. B. De Jesús, "Action Recognition in Visual Sensor Networks : A Data Fusion Perspective Rodrigo Cilla Ugarte," UNIVERSIDAD CARLOS III DE MADRID ESCUELA POLITÉCNICA SUPERIOR, 2012.

A. Iosifidis, A. Tefas, and I. Pitas, "View-independent human action recognition based on multi-view action images and discriminant learning," in *2013 IEEE 11th IVMSP Workshop: 3D Image/Video Technologies and Applications, IVMSP 2013 - Proceedings*, 2013, pp. 1–4.

D. Das Dawn and S. H. Shaikh, "A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector," *Visual Computer*, vol. 32, no. 3, pp. 289–306, 2016.

Y. J. Wanare and P. R. Deshmukh, "A Survey on Human Action Recognition using Scale Invariant Feature Transformation," vol. 3, no. 2, pp. 35–38, 2015.

M. Ziaeefard and R. Bergevin, "Semantic human activity recognition: A literature review," *Pattern Recognition*, vol. 48, no. 8, pp. 2329–2345, 2015.

W. Li, Z. Zhang, and Z. Liu, "Action Recognition Based on A Bag of 3D Points.pdf," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 2010, pp. 9–14.

I. Laptev, "On space-time interest points," *International Journal of Computer Vision*, vol. 64, no. 2–3, pp. 107–123, 2005.

J. Liu, S. Ali, and M. Shah, "Recognizing human actions using multiple features," *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, pp. 1–8, Jun. 2008.

N. Ballas, Y. Yang, Z.-Z. Lan, B. Delezoide, F. Preteux, and a Hauptmann, "Space-Time Robust Representation for Action Recognition," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, 2013, vol. c, pp. 2704–2711.

A. Fathi and G. Mori, "Action Recognition By Learning Mid – Level," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, no. 5510, pp. 1–8.

T. Zhang, J. Liu, S. Liu, C. Xu, and H. Lu, “Boosted exemplar learning for action recognition and annotation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 853–866, 2011.

Tianzhu Zhang, J. Liu, Si Liu, Yi Ouyang, and Hanqing Lu, “<![CDATA[Boosted Exemplar Learning for human action recognition]]>,” *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, no. 2, pp. 538–545, Sep. 2009.

K. Guo, *Action recognition using log-covariance matrices of silhouette and optical-flow features*. Boston University, 2012.

R. Cutler and M. Turk, “View-based interpretation of real-time optical flow for gesture recognition,” in *Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998*, 1998, pp. 416–421.

R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, “Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions,” in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, 2009, pp. 1932–1939.

M. B. Holte, T. B. Moeslund, N. Nikolaidis, and I. Pitas, “3D Human Action Recognition for Multi-view Camera Systems,” in *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 2011, pp. 342–349.

J. Yamato, J. Ohya, and K. Ishii, “Recognizing human action in time-sequential images using hidden Markov model,” in *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1992, pp. 379–385.

L. Wang and D. Suter, “Visual learning and recognition of sequential data manifolds with applications to human movement analysis,” *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 153–172, May 2008.

Y. Hu, “Action Detection in Complex Scenes with Spatial and Temporal Ambiguities Objective and Motivation,” in *Challenges*, 2009, no. Iccv, pp. 1–20.

M. C. Roh, H. K. Shin, and S. W. Lee, “View-independent human action recognition with Volume Motion Template on single stereo camera,” *Pattern Recognition Letters*, vol. 31, no. 7, pp. 639–647, May 2010.

C.-H. Fang, J.-C. Chen, C.-C. Tseng, and J.-J. J. Lien, “Human Action Recognition Using Spatio-temporal Classification,” in *COMPUTER VISION - ACCV 2009, PT II*, 2010, vol. 5995, no. II, pp. 98–109.

L. Shao and X. Chen, “Histogram of Body Poses and Spectral Regression Discriminant Analysis for Human Action Categorization,” in *Procedings of the British Machine Vision Conference 2010*, 2010, pp. 88.1–88.11.

L. Shao, D. Wu, and X. Chen, “Action recognition using correlogram of body poses and spectral regression,” in *Proceedings - International Conference on Image Processing, ICIP*, 2011, pp. 209–212.

Z. Liu, K. T. Ng, S. C. Chan, and X. W. Song, “A new multi-view articulated human motion tracking algorithm with improved silhouette extraction and view adaptive fusion,” *Proceedings - IEEE International Symposium on Circuits and Systems*, pp. 713–716, May 2013.

G. Willems, T. Tuytelaars, and L. Van Gool, “An Efficient Dense and Scale-Invariant Spatio-Temporal Interest Point Detector,” in *Computer Vision - Eccv 2008, Pt Ii, Proceedings*, vol. 5303, Springer, 2008, pp. 650–663.

S.-F. Wong and R. Cipolla, “Extracting Spatiotemporal Interest Points using Global Information,” in *Proceedings of IEEE 11th International Conference on Computer Vision*, 2007, pp. 1–8.

A. Yilmaz and M. Shah, “Actions sketch: A novel action representation,” in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005, vol. I, pp. 984–989.

C. Rao, A. Yilmaz, and M. Shah, “View-invariant representation and recognition of actions,” *International Journal of Computer Vision*, vol. 50, no. 2, pp. 203–226, 2002.

H. Qian, Y. Mao, W. Xiang, and Z. Wang, “Recognition of human activities using SVM multi-class classifier,” *Pattern Recognition Letters*, vol. 31, no. 2, pp. 100–111, Jan. 2010.

J. E. Ingvaldsen, ??Zlem ??Zg??Bek, and J. A. Gulla, *Context-aware user-driven news recommendation*, vol. 1542. Spie Press Bellingham^ eWA WA, 2015.

A. U. Haq, I. Gondal, and M. Murshed, “On temporal order invariance for view-invariant action recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 203–211, 2013.

J. Luo, W. Wang, and H. Qi, “Spatio-temporal feature extraction and representation for RGB-D human action recognition,” *Pattern Recognition Letters*, vol. 50, no. SI, pp. 139–148, Dec. 2014.

M. Sivarathinabala, S. Abirami, and R. Baskaran, “View invariant human action recognition using improved motion descriptor,” in *Smart Innovation, Systems and Technologies*, vol. 33, Springer, 2015, pp. 545–554.

X. Wu, D. Xu, L. Duan, and J. Luo, “Action recognition using context and appearance distribution features,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 489–496.

P. Yan, S. M. Khan, and M. Shah, "Learning 4D action feature models for arbitrary view action recognition," *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 1–7, 2008.

And J. M. M. Rodrigo Cilla, Miguel A. Patricio, Antonio Berlanga, "Fusion of Single View Soft k-NN Classifiers for Multicamera Human Action Recognition.pdf," in *Hybrid Artificial Intelligence Systems, Pt 2*, 2010, vol. 6077, pp. 436–443.

G. Srivastava, H. Iwaki, J. Park, and A. C. Kak, "Distributed and lightweight multi-camera human activity classification," in *2009 3rd ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC 2009*, 2009, pp. 433–440.

J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J.-M. Geusebroek, "Visual Word Ambiguity," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 7, pp. 1271–1283, 2010.

B. a Olshausen and D. J. Field, "Sparse coding with an incomplete basis set: a strategy employed by \protect{V1}," *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.

W. E. Vinje, W. E. Vinje, and J. L. Gallant, "Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision," *Science (New York, N.Y.)*, vol. 1273, no. 2000, pp. 1273–1276, 2007.

C. Bao, L. He, and Y. Wang, "Linear spatial pyramid matching using non-convex and non-negative sparse coding for image classification," *2015 IEEE China Summit and International Conference on Signal and Information Processing, ChinaSIP 2015 - Proceedings*. pp. 186–190, 2015.

P. Koniusz, F. Yan, and K. Mikolajczyk, "Comparison of mid-level feature coding approaches and pooling strategies in visual concept detection," *Computer Vision and Image Understanding*, vol. 117, no. 5, pp. 479–492, 2013.

B. Wang, W. Gai, S. Guo, Y. Liu, W. Wang, and M. Zhang, "Spatially regularized and locality-constrained linear coding for human action recognition," *Optical Review*, vol. 21, no. 3, pp. 226–236, 2014.

Z. Gao, H. Zhang, G. P. Xu, Y. B. Xue, and a. G. Hauptmann, "Multi-view discriminative and structured dictionary learning with group sparsity for human action recognition," *Signal Processing*, vol. 112, pp. 83–97, 2015.

P. Bilinski, E. Corvee, S. Bak, and F. Bremond, "Relative dense tracklets for human action recognition," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013*, 2013, pp. 1–7.

O. V. R. Murthy and R. Goecke, "Ordered trajectories for large scale human action recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 412–419.

H. Wang and C. Schmid, “Action Recognition with Improved Trajectories,” in *Computer Vision (ICCV), 2013 IEEE International Conference on*, 2013, pp. 3551–3558.

H. Wang, A. Kl??ser, C. Schmid, and C. L. Liu, “Action recognition by dense trajectories,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, no. 2011, pp. 3169–3176, 2011.

R. Poppe, “A survey on vision-based human action recognition,” *Image and Vision Computing*, vol. 28, no. 6, pp. 976–990, Jun. 2010.

T. K. Vintsyuk, “Speech discrimination by dynamic programming,” *Cybernetics*, vol. 4, no. 1, pp. 52–57, 1972.

D. Hutchison and J. C. Mitchell, “Image Analysis and Recognition,” in *Iciar*, 2009, vol. 5112, pp. 1–976.

K. S. Rao and S. G. Koolagudi, *Robust Emotion Recognition using Spectral and Prosodic Features*. Springer Science & Business Media, 2013.

G. Xiao, Y. Jiang, G. Song, and J. Jiang, “Support-vector-machine tree-based domain knowledge learning toward automated sports video classification,” *Optical Engineering*, vol. 49, no. 12, p. 127003, 2010.

M. Ziaeefard and H. Ebrahimnezhad, “Hierarchical human action recognition by normalized-polar histogram,” in *Proceedings - International Conference on Pattern Recognition*, 2010, pp. 3720–3723.

J. M. Chaquet, E. J. Carmona, and A. Fern??ndez-Caballero, “A survey of video datasets for human action and activity recognition,” *Computer Vision and Image Understanding*, vol. 117, no. 6, pp. 633–659, Jun. 2013.

D. Weinland, E. Boyer, and R. Ronfard, “Action recognition from arbitrary views using 3D exemplars,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2007, pp. 1–7.

S. Singh, S. a. Velastin, and H. Ragheb, “MuHAVi: A multicamera human action video dataset for the evaluation of action recognition methods,” in *Proceedings - IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010*, 2010, pp. 48–55.

R. Polana, N. York, and R. Nelson, “Detecting Activities,” *Journal of Visual Communication and Image Representation*, vol. 5, no. 2, pp. 172–180, 1994.

T. Mori, Y. Kamisuwa, H. Mizoguchi, and T. Sato, “Action recognition system based on human finder and human tracker,” in *Intelligent Robots and Systems, 1997. IROS’97., Proceedings of the 1997 IEEE/RSJ International Conference on*, 1997, vol. 3, pp. 1334–1341.

Y. Kuniyoshi and H. Inoue, “Qualitative recognition of ongoing human action sequences,” in *International Joint Conference on Artificial Intelligence*, 1993, pp. 1600–1609.

N. Amraji, L. Mu, and M. Milanova, “Shape – Based Human Actions Recognition in Videos,” in *HUMAN-COMPUTER INTERACTION: DESIGN AND DEVELOPMENT APPROACHES, PT I*, 2011, vol. 6761, pp. 539–546.

S. Al-Ali, M. Milanova, H. Al-Rizzo, and V. L. Fox, “Human Action Recognition: Contour-Based and Silhouette-Based Approaches,” in *Computer Vision in Control Systems-2*, Springer, 2015, pp. 11–47.

S. Althloothi, M. H. Mahoor, X. Zhang, and R. M. Voyles, “Human activity recognition using multi-features and multiple kernel learning,” *Pattern Recognition*, vol. 47, no. 5, pp. 1800–1812, May 2014.

I. Cohen and H. Li, “Inference of human postures by classification of 3D human body shape,” in *2003 IEEE International SOI Conference. Proceedings (Cat. No.03CH37443)*, 2003, pp. 74–81.

A. Yilmaz and M. Shah, “A differential geometric approach to representing the human actions,” *Computer Vision and Image Understanding*, vol. 109, no. 3, pp. 335–351, Mar. 2008.

A. A. Liu, N. Xu, Y. T. Su, H. Lin, T. Hao, and Z. X. Yang, “Single/multi-view human action recognition via regularized multi-task learning,” *Neurocomputing*, vol. 151, no. P2, pp. 544–553, Mar. 2015.

S. Spurlock and R. Souvenir, “Multi-view action recognition one camera at a time,” *2014 IEEE Winter Conference on Applications of Computer Vision, WACV 2014*, pp. 604–609, 2014.

C. Thurau and V. Hlav????, “Pose primitive based human action recognition in videos or still images,” in *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008, pp. 2955–2962.

Y. Du, W. Wang, and L. Wang, “Hierarchial Recurrent Neural Network for Skeleton Based Action Recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1110–1118.

B. X. Nie, C. Xiong, and S.-C. Zhu, “Joint action recognition and pose estimation from video,” *Cvpr*, pp. 1293–1301, 2015.

A. I. Maqueda, A. Ruano, C. R. Del-Blanco, P. Carballeira, F. Jaureguizar, and N. García, “Novel multi-feature Bag-of-Words descriptor via subspace random projection for efficient human-action recognition,” in *AVSS 2015 - 12th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2015, pp. 1–6.

A Iosifidis, a Tefas, and I. Pitas, "Human action recognition based on bag of features and multi-view neural networks," in *Image Processing (ICIP), 2014 IEEE International Conference on*, 2014, pp. 1510–1514.

M. Palhang, M. Faraki, and C. Sanderson, "Log-Euclidean bag of words for human action recognition," *IET Computer Vision*, vol. 9, no. 3, pp. 331–339, 2015.

A. Iosifidis, A. Tefas, and I. Pitas, "Distance-based human action recognition using optimized class representations," *Neurocomputing*, vol. 161, pp. 47–55, Aug. 2015.

S. I. Audin, S. Nath, S. Basak, F. S. Rahman, R. Nath, and S. a Fattah, "A Human Action Recognition Scheme Based on Spatio-Temporal Variation of Region of Interest in Horizontal and Vertical Direction," in *2014 International Conference on Informatics, Electronics & Vision (Iciev)*, 2014.

N. A. Azis, H. J. Choi, and Y. Iraqi, "Substitutive skeleton fusion for human action recognition," in *2015 International Conference on Big Data and Smart Computing, BIGCOMP 2015*, 2015, pp. 170–177.

J. a. Nasiri, N. Moghadam Charkari, and K. Mozafari, "Energy-based model of least squares twin Support Vector Machines for human action recognition," *Signal Processing*, vol. 104, pp. 248–257, 2014.

A. Iosifidis, A. Tefas, and I. Pitas, "Minimum class variance extreme learning machine for human action recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 11, pp. 1968–1979, Nov. 2013.

S. Shariat and V. Pavlovic, "Isotonic CCA for sequence alignment and activity recognition," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2572–2578, Nov. 2011.

J. Wang and H. Zheng, "View-robust action recognition based on temporal self-similarities and dynamic time warping," *2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, pp. 498–502, May 2012.

S. Cherla, K. Kulkarni, A. Kale, and V. Ramasubramanian, "Towards fast, view-invariant human action recognition," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops*, 2008, pp. 1650–1657.

N. Ikizler and P. Duygulu, "Histogram of oriented rectangles: A new pose descriptor for human action recognition," *Image and Vision Computing*, vol. 27, no. 10, pp. 1515–1526, Sep. 2009.

S. A. Rahman, M. K. H. Leung, and S. Y. Cho, "Human action recognition employing negative space features," *Journal of Visual Communication and Image Representation*, vol. 24, no. 3, pp. 217–231, Apr. 2013.

N. Gkalelis, A. Tefas, and I. Pitas, “Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1511–1521, Nov. 2008.

A. A. Chaaraoui, P. Climent-Pérez, and F. Flórez-Revuelta, “Silhouette-based human action recognition using sequences of key poses,” *Pattern Recognition Letters*, vol. 34, no. 15, pp. 1799–1807, Nov. 2013.

E. A. Mosabbeb, K. Raahemifar, and M. Fathy, “Multi-view human activity recognition in distributed camera sensor networks.,” *Sensors (Basel, Switzerland)*, vol. 13, no. 7, pp. 8750–8770, Jan. 2013.

A. Iosifidis, A. Tefas, N. Nikolaidis, and I. Pitas, “Multi-view human movement recognition based on fuzzy distances and linear discriminant analysis,” *Computer Vision and Image Understanding*, vol. 116, no. 3, pp. 347–360, Mar. 2012.

W. Yang, Y. Gao, L. Cao, M. Yang, and Y. Shi, “mPadal: a joint local-and-global multi-view feature selection method for activity recognition,” *Applied Intelligence*, vol. 41, no. 3, pp. 776–790, Jul. 2014.

S. Cheema, A. Eweiwi, C. Thurau, and C. Bauckhage, “Action recognition by learning discriminative key poses,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, no. i, pp. 1302–1309.

A. Eweiwi, S. Cheema, C. Thurau, C. Bauckhage, and F. Iais, “Temporal Key Poses for Human Action Recognition Bonn-Aachen International Center for IT , University of Bonn,” in *History*, 2011, pp. 1310–1317.

F. Martínez-Contreras, C. Orrite-Uruñuela, E. Herrero-Jaraba, H. Ragheb, and S. a. Velastin, “Recognizing human actions using silhouette-based HMM,” in *6th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2009*, 2009, pp. 43–48.

J. Konečný and M. Hagara, “One-shot-learning gesture recognition using hog-hof features,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 2513–2532, 2014.

K. Onishi, T. Takiguchi, and Y. Ariki, “3D human posture estimation using the HOG features from monocular image,” in *2008 19th International Conference on Pattern Recognition*, 2008, pp. 3–6.

L. van der Maaten, E. Postma, and J. van den Herik, “Dimensionality Reduction: A Comparative Review,” Tilburg, Netherlands: Tilburg Centre for Creative Computing, Tilburg University, Technical Report: 2009-005, 2009.

A. R. Ahad, T. Ogata, J. Tan, H. Kim, and S. Ishikawa, “A complex motion recognition technique employing directional motion templates,” *International Journal of Innovative Computing, Information and Control*, vol. 4, no. 8, pp. 1943–1954, 2008.

M. A. R. Ahad, J. K. Tan, H. Kim, and S. Ishikawa, “Motion history image: Its variants and applications,” *Machine Vision and Applications*, vol. 23, no. 2, pp. 255–281, Oct. 2012.

A. A. Chaaraoui, P. Climent-Pérez, and F. Flórez-Revuelta, “An efficient approach for multi-view human action recognition based on bag-of-key-poses,” in *International Workshop on Human Behavior Understanding*, 2012, pp. 29–40.

F. Murtaza, M. H. Yousaf, and S. A. Velastin, “Multi-view Human Action Recognition using Histograms of Oriented Gradients (HOG) Description of Motion History Images (MHIs),” in *2015 13th International Conference on Frontiers of Information Technology (FIT)*, 2015, pp. 297–302.

I. Terms, A. Iosifidis, A. Tefas, and I. Pitas, “MULTI-VIEW HUMAN ACTION RECOGNITION UNDER OCCLUSION BASED ON FUZZY DISTANCES AND NEURAL NETWORKS Alexandros Iosifidis , Anastasios Tefas and Ioannis Pitas Aristotle University of Thessaloniki , Depeartment of Informatics,” no. Eusipco, pp. 1129–1133, 2012.

Z. Gao, H. Zhang, G. P. Xu, Y. B. Xue, and a. G. Hauptmann, “Multi-view discriminative and structured dictionary learning with group sparsity for human action recognition,” *Signal Processing*, vol. 112, no. SI, pp. 83–97, Jul. 2015.

J. Luo, W. Wang, and H. Qi, “Feature extraction and representation for distributed multi-view human action recognition,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 3, no. 2, pp. 145–154, 2013.

A. Iosifidis, A. Tefas, and I. Pitas, “Multi-view Regularized Extreme Learning Machine for Human Action Recognition,” in *Artificial Intelligence: Methods and Applications*, Springer, 2014, pp. 84–94.

X. Ji, Z. Ju, C. Wang, and C. Wang, “Multi-view transition HMMs based view-invariant human action recognition method,” *Multimedia Tools and Applications*, pp. 1–18, 2015.

H. Lin, L. Chaisorn, Y. Wong, A. A. Liu, Y. T. Su, and M. S. Kankanhalli, “View-invariant feature discovering for multi-camera human action recognition,” in *2014 IEEE International Workshop on Multimedia Signal Processing, MMSP 2014*, 2014, pp. 22–24.

J. Feng and J. Xiao, “View-invariant human action recognition via robust locally adaptive multi-view learning,” *Frontiers of Information Technology & Electronic Engineering*, vol. 16, pp. 917–929, 2015.

J.-Y. Nie, *Introduction to Information Retrieval*, vol. 1. Cambridge university press Cambridge, 2012.

M. A. R. Ahad, J. K. Tan, H. Kim, and S. Ishikawa, “Motion history image: Its variants and applications,” *Machine Vision and Applications*, vol. 23, no. 2, pp. 255–281, 2012.

G. Madzarov, D. Gjorgjevikj, and I. Chorbev, “A Multi-class SVM Classifier Utilizing Binary Decision Tree Support vector machines for pattern recognition,” *Electrical Engineering*, vol. 33, no. 1, pp. 233–241, 2009.

Z. H. Feng, J. Kittler, W. Christmas, and X. J. Wu, “Feature level multiple model fusion using multilinear subspace analysis with incomplete training set and its application to face image analysis,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7872 LNCS, Springer, 2013, pp. 73–84.

J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, “Real-time human pose recognition in parts from single depth images,” *Studies in Computational Intelligence*, vol. 411, pp. 119–135, Jun. 2013.

A. A. Chaaraoui and F. Fl??rez-Revuelta, “Optimizing human action recognition based on a cooperative coevolutionary algorithm,” *Engineering Applications of Artificial Intelligence*, vol. 31, no. 2014, pp. 116–125, May 2013.

B. U. Töreyin, U. Güdükbay, and a E. Çetin, “Silhouette-Based Method for Object Classification and Human Action Recognition in Video,” in *Eccv 2006*, 2006, vol. 3979, pp. 64–77.

Z. Z. Htike, S. Egerton, and K. Y. Chow, “Model-free Viewpoint Invariant Human Activity Recognition,” in *Computer*, 2011, vol. I.

J. Carranza, C. Theobalt, M. a. Magnor, and H.-P. Seidel, “Free-viewpoint video of human actors,” *ACM Transactions on Graphics*, vol. 22, no. 3, p. 569, 2003.

M. Ángeles Mendoza, M. Ángeles Mendoza, N. Pérez de la Blanca, and N. Pérez de la Blanca, “HMM-Based Action Recognition Using Contour Histograms,” in *Pattern Recognition and Image Analysis*, Springer, 2007, pp. 394–401.

D. Zhang and G. Lu, “Review of shape representation and description techniques,” *Pattern Recognition*, vol. 37, no. 1, pp. 1–19, 2004.

Y. Mingqiang, K. Kidiyo, and R. Joseph, “A Survey of Shape Future Extraction Techniques,” *Pattern Recognition*, vol. 2008, pp. 43–90, 2008.

X. Huang, L. Zheng, R. Liang, W. Wang, and X. Ma, “Human Action Recognition Based on SVM Using Multiple Features,” in *2012 International Conference on Artificial Intelligence and Soft Computing (Icaisc 2012)*, 2012, vol. 12, pp. 160–165.

G. Guo and A. Lai, “A survey on still image based human action recognition,” *Pattern Recognition*, vol. 47, no. 10, pp. 3343–3361, Oct. 2014.

F. S. Khan, M. A. Rao, J. Van De Weijer, A. D. Bagdanov, A. M. Lopez, and M. Felsberg, "Coloring action recognition in still images," *International Journal of Computer Vision*, vol. 105, no. 3, pp. 205–221, 2013.

P. Yan, S. M. Khan, and M. Shah, "Learning 4D action feature models for arbitrary view action recognition," in *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008, pp. 1–7.

R. Cilla, M. a. Patricio, A. Berlanga, and J. M. Molina, "Human action recognition with sparse classification and multiple-view learning," *Expert Systems*, vol. 31, no. 4, pp. 354–364, Sep. 2014.

R. Gayathri and P. Ramamoorthy, "Feature fusion of palmprint and face biometrics," in *European Journal of Scientific Research*, 2012, vol. 77, no. 4, pp. 457–470.

S. Baysal, M. C. Kurt, and P. Duygulu, "Recognizing Human Actions Using Key Poses," *Pattern Recognition (ICPR), 2010 20th International Conference on*, no. d, pp. 1727–1730, Aug. 2010.

F. Lv and R. Nevatia, "Single View Human Action Recognition using Key Pose Matching and Viterbi Path Searching," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

I. Guyon, S. Gunn, M. Nikravesh, and L. a Zadeh, *Feature extraction: foundations and applications*, vol. 207. Springer, 2008.

E. Cantú-Paz, "Feature Subset Selection, Class Separability, and Genetic Algorithms," in *Genetic and Evolutionary Computation Conference*, 2004, pp. 959–970.

S. C. Yusta, "Different metaheuristic strategies to solve the feature selection problem," *Pattern Recognition Letters*, vol. 30, no. 5, pp. 525–534, 2009.

W. Siedlecki and J. Sklansky, "A Note on Genetic Algorithms for Large- scale Feature Selection," *Pattern Recognition Letters*, vol. 10, no. 5, pp. 335–347, 1989.

H. Liu and H. Motoda, *Feature selection for knowledge discovery and data mining*, vol. 454. Springer Science & Business Media, 1998.

G. John, R. Kohavi, and K. Pfleger, "Irrelevant Features and the Subset Selection Problem.," in *... Learning: Proceedings of the …*, 1994, pp. 121–129.

P. L. Lanzi, "Fast feature selection with genetic algorithms: a filter approach," in *Evolutionary Computation, 1997., IEEE International Conference on*, 1997, pp. 537–540.

Y. Zhang, D. Gong, Y. Hu, and W. Zhang, "Feature selection algorithm based on bare bones particle swarm optimization," *European Journal of Operational Research*, vol. 148, pp. 150–157, 2015.

R. Eberhart and J. Kennedy, “A new optimizer using particle swarm theory,” in *MHS’95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, 1995, vol. 1, pp. 39–43.

M. Mandal, A. Mukhopadhyay, and U. Maulik, “Prediction of protein subcellular localization by incorporating multiobjective PSO-based feature subset selection into the general form of Chou’s PseAAC,” *Medical and Biological Engineering and Computing*, vol. 53, no. 4, pp. 331–344, 2015.

B. Xue, a. K. Qin, and M. Zhang, “An archive based particle swarm optimisation for feature selection in classification,” in *Evolutionary Computation (CEC), 2014 IEEE Congress on*, 2014, pp. 3119–3126.

Eberhart and S. Yuhui, “Particle swarm optimization: developments, applications and resources,” in *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, 2001, vol. 1, pp. 81–86 vol. 1.

R. C. R. Eberhart and Y. Shi, “Comparison between Genetic Algorithms and Particle Swarm Optimization,” in *Ep’98*, 1998, vol. 1447, pp. 611–616.

S. Sumathi and P. Surekha, *Computational Intelligence Paradigms: Theory And Applications Using MATLAB*. CRC Press, 2010.

S. Sumathi and P. Surekha, *Computational Intelligence Paradigms Computational Intelligence Paradigms Theory and Applications*. 2010.

M. Wan, M. Li, G. Yang, S. Gai, and Z. Jin, “Feature extraction using two-dimensional maximum embedding difference,” *Information Sciences*, vol. 274, pp. 55–69, 2014.

R. Cilla, M. a. Patricio, A. Berlanga, and J. M. Molina, “A probabilistic, discriminative and distributed system for the recognition of human actions from multiple views,” *Neurocomputing*, vol. 75, no. 1, SI, pp. 78–87, Jan. 2012.

A a Chaaraoui and F. Florez-Revuelta, “Human Action Recognition Optimization Based on Evolutionary Feature Subset Selection,” in *Gecco’13: Proceedings of the 2013 Genetic and Evolutionary Computation Conference*, 2013, pp. 1229–1236.

Z. Zhang, C. Wang, B. Xiao, W. Zhou, S. Liu, and C. Shi, “Cross-view action recognition via a continuous virtual path,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2690–2697, Jun. 2013.