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
COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia
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 MENGUNGGANKAN JARAK MATRIKS BERASASKAN PENGKELASAN TANPA MENGHIRAUHAN 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, dapan 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 perbahan 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.
I certify that a Thesis Examination Committee has met on 07 June 2016 to conduct the final examination of Alhossein Aryanfar on his thesis entitled "Shape-Based Multi-View Human Action Recognition Using Distance-Based-Matrix-Regardless-Of-Row-Priority Classifier" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

**Abu Bakar b Md Sultan, PhD**
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

**Lilly Suriani Affendey, PhD**
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

**Norwati Mustapha, PhD**
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

**Xin-She Yang, PhD**
Professor
Middlesex University
United Kingdom
(External Examiner)

---

[ZULKARNAIN ZAINAL, PhD]
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 26 July 2016
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 Committee : Dr. Razali Yaakob

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

Signature: ______________________
Name of Chairman of Supervisory Committee : Dr. Md. Nasir Sulaiman
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
</tr>
<tr>
<td>1.1</td>
<td>Background and Motivation</td>
</tr>
<tr>
<td>1.2</td>
<td>Problem Statements</td>
</tr>
<tr>
<td>1.3</td>
<td>Hypotheses</td>
</tr>
<tr>
<td>1.4</td>
<td>Objectives</td>
</tr>
<tr>
<td>1.5</td>
<td>Contributions</td>
</tr>
<tr>
<td>1.6</td>
<td>Research Scope and Assumptions</td>
</tr>
<tr>
<td>1.7</td>
<td>Thesis Organization</td>
</tr>
<tr>
<td>2</td>
<td>LITERATURE REVIEW</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2.2</td>
<td>Motions, Events, Actions, and Activities</td>
</tr>
<tr>
<td>2.3</td>
<td>Single-view and Multi-view Human Action Recognition</td>
</tr>
<tr>
<td>2.4</td>
<td>Human Action Recognition Approaches Categorization</td>
</tr>
<tr>
<td>2.5</td>
<td>Human Action Recognition Phases</td>
</tr>
<tr>
<td>2.6</td>
<td>Feature Extraction</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Global Features</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Local Features</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Hybrid features</td>
</tr>
<tr>
<td>2.6.4</td>
<td>Feature Fusion</td>
</tr>
<tr>
<td>2.7</td>
<td>Video Description</td>
</tr>
<tr>
<td>2.8</td>
<td>Classification</td>
</tr>
<tr>
<td>2.8.1</td>
<td>Exemplar-based Method</td>
</tr>
<tr>
<td>2.8.2</td>
<td>Generative Models</td>
</tr>
<tr>
<td>2.8.3</td>
<td>Discriminative classifiers</td>
</tr>
<tr>
<td>2.9</td>
<td>Datasets and Benchmarks</td>
</tr>
<tr>
<td>2.9.1</td>
<td>IXMAS Dataset</td>
</tr>
<tr>
<td>2.9.2</td>
<td>MuHAVi Dataset</td>
</tr>
<tr>
<td>2.9.3</td>
<td>Weizmann Dataset</td>
</tr>
<tr>
<td>2.10</td>
<td>Related Works</td>
</tr>
<tr>
<td>2.11</td>
<td>Summary</td>
</tr>
<tr>
<td>3</td>
<td>RESEARCH METHODOLOGY</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>3.2</td>
<td>Research Methodology Steps</td>
</tr>
</tbody>
</table>
3.3 Literature Review

3.4 General Human Action Recognition Framework

3.5 An Approach Based on Keyposes, 2D Wavelet and C5.0 Classifier

3.6 An Approach Based on Distance Signal, 2D Wavelet and Hierarchical Classifier

3.7 An Approach Based on PMBI, PSO, and Hierarchical SVM

3.8 MHARSD Approach

3.8.1 Proposed framework for approach MHARSD

3.9 Summary

4 MULTI-VIEW HAR USING MHARSD

4.1 Introduction

4.2 Proposed Framework for Human Action Recognition

4.3 Feature Extraction

4.3.1 Silhouette-Based Feature Extraction

4.3.2 Shape-Based Feature Extraction

4.3.3 Global Silhouette Shape-Representation (GSSR)

4.4 Feature fusion

4.5 Distance Metrics

4.5.1 Distance Metrics for Matrix

4.6 Keyframe Selection

4.6.1 Keyframe Selection Algorithm

4.7 Feature Subset Selection

4.7.1 Evolutionary Feature Subset Selection

4.7.2 Feature Subset Selection Using BPSO

4.8 Classification

4.8.1 Multi-View Human Action Recognition based on DMRRP

4.9 Summary

5 RESULTS AND DISCUSSION

5.1 Introduction

5.2 Evaluation Methodology and Technical Details

5.3 Result of Initial Approaches (KWCC, DSWHC, and PPHS)

5.4 Evaluation of Various Distance Measurement Techniques

5.5 Evaluation of Various Frame Subset Selection Methods

5.6 Results of the MHARSD on MuHAVi-8

5.6.1 Result of Optimum Value for Number of Keyframes

5.6.2 Preliminary Result

5.6.3 Final Result on MuHAVi-8

5.7 Results of the MHARSD on MuHAVi-14

5.7.1 Result of Optimum Value for Number of Keyframes

5.7.2 Preliminary Result on MuHAVi-14
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.7.3 Final Results on MuHAVi-14</td>
<td>100</td>
</tr>
<tr>
<td>5.8 Results of the MHARSD on IXMAS</td>
<td>102</td>
</tr>
<tr>
<td>5.8.1 Result of Optimum Value for Number of keyframes</td>
<td>103</td>
</tr>
<tr>
<td>5.8.2 Preliminary Result on IXMAS</td>
<td>104</td>
</tr>
<tr>
<td>5.8.3 Final Result on IXMAS</td>
<td>107</td>
</tr>
<tr>
<td>5.9 Complicated Results of IXMAS, From Single-View to Multi-View</td>
<td>110</td>
</tr>
<tr>
<td>5.10 Comparison with State-of-the-Art Methods</td>
<td>113</td>
</tr>
<tr>
<td>5.11 Summary</td>
<td>114</td>
</tr>
</tbody>
</table>

6 CONCLUSION AND FUTURE WORKS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Conclusion</td>
<td>116</td>
</tr>
<tr>
<td>6.2 Future Works</td>
<td>117</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Some applications of HAR [1]</td>
</tr>
<tr>
<td>2.1</td>
<td>The categories and sub-categories of methods on HAR [27]</td>
</tr>
<tr>
<td>2.2</td>
<td>Overview of action modelling categorization [25]</td>
</tr>
<tr>
<td>2.3</td>
<td>The primary stages in human action recognition</td>
</tr>
<tr>
<td>2.4</td>
<td>Sample images of IXMAS dataset</td>
</tr>
<tr>
<td>2.5</td>
<td>Sample of IXMAS images with different view angle</td>
</tr>
<tr>
<td>2.6</td>
<td>Sample images of MuHAVi dataset</td>
</tr>
<tr>
<td>2.7</td>
<td>Sample images of and extracted silhouettes from Weizmann</td>
</tr>
<tr>
<td>3.1</td>
<td>The general steps of research methodology</td>
</tr>
<tr>
<td>3.2</td>
<td>The stages of KWCC</td>
</tr>
<tr>
<td>3.3</td>
<td>The stages of DSWHC</td>
</tr>
<tr>
<td>3.4</td>
<td>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.</td>
</tr>
<tr>
<td>3.5</td>
<td>(a) GDA Algorithm to separate 12 actions into two-category, (b) The structure of the proposed classifier</td>
</tr>
<tr>
<td>3.6</td>
<td>The stages of PPHS</td>
</tr>
<tr>
<td>3.7</td>
<td>Proposed hierarchical classifier</td>
</tr>
<tr>
<td>3.8</td>
<td>Proposed MHARSD approach</td>
</tr>
<tr>
<td>3.9</td>
<td>Framework of the MHARSD</td>
</tr>
<tr>
<td>3.10</td>
<td>Overall techniques used in the MHARSD</td>
</tr>
<tr>
<td>3.11</td>
<td>Example of RGB images, silhouettes and ROI from different views</td>
</tr>
<tr>
<td>3.12</td>
<td>Example of image, depth information, and its silhouette</td>
</tr>
<tr>
<td>3.13</td>
<td>Structure of shape representation and sample on features [143]</td>
</tr>
<tr>
<td>3.14</td>
<td>Feature extraction phases using GSSR</td>
</tr>
<tr>
<td>3.15</td>
<td>Example of a centroid point on a contour and silhouette</td>
</tr>
<tr>
<td>3.16</td>
<td>Example of a bounding box on a silhouette and contour points</td>
</tr>
<tr>
<td>3.17</td>
<td>Example of a main ellipse, major and minor axes</td>
</tr>
<tr>
<td>3.18</td>
<td>Intervals of angular variation</td>
</tr>
<tr>
<td>3.19</td>
<td>Axes and orientation of an ellipse</td>
</tr>
<tr>
<td>3.20</td>
<td>Silhouette and smallest convex hull (red)</td>
</tr>
<tr>
<td>3.21</td>
<td>Silhouette, equivalent circle, and their diameters</td>
</tr>
<tr>
<td>3.22</td>
<td>Example of feature fusion steps</td>
</tr>
<tr>
<td>3.23</td>
<td>Flowchart of MDFV</td>
</tr>
<tr>
<td>3.24</td>
<td>MDFV algorithm</td>
</tr>
<tr>
<td>3.25</td>
<td>Sample of MDFV-selected keyframes at different thresholds</td>
</tr>
<tr>
<td>3.26</td>
<td>BPSO-based feature subset selection</td>
</tr>
<tr>
<td>3.27</td>
<td>Fitness function for BPSO-based feature subset selection</td>
</tr>
<tr>
<td>3.28</td>
<td>Extracted features from sequence 1 and 2</td>
</tr>
<tr>
<td>3.29</td>
<td>Comparison of and distance between sequences 1 and 2</td>
</tr>
<tr>
<td>3.30</td>
<td>Matching the frames of sequence1 and 2</td>
</tr>
<tr>
<td>3.31</td>
<td>Proposed DMRRP classifier</td>
</tr>
<tr>
<td>3.32</td>
<td>Flowchart of DMRRP</td>
</tr>
<tr>
<td>5.1</td>
<td>Comparison of keyframe selection methods on MuHAVi-8</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------</td>
</tr>
<tr>
<td>5.2</td>
<td>Comparison of keyframe selection methods on MuHAVi-14</td>
</tr>
<tr>
<td>5.3</td>
<td>Accuracy variation based on the keyframe numbers on MuHAVi-8</td>
</tr>
<tr>
<td>5.4</td>
<td>Preliminary confusion matrix (heatmap) on MuHAVi-8</td>
</tr>
<tr>
<td>5.5</td>
<td>Confusion matrix (heatmap) on MuHAVi-8</td>
</tr>
<tr>
<td>5.6</td>
<td>Accuracy variations according to keyframe number on MuHAVi-8</td>
</tr>
<tr>
<td>5.7</td>
<td>Preliminary confusion matrix (heatmap) on MuHAVi-14</td>
</tr>
<tr>
<td>5.8</td>
<td>Confusion matrix (heat map) on MuHAVi-14 (multi-view and camera 4 using LOSO-CV and LOAO-CV; camera 3 using LOSO-CV)</td>
</tr>
<tr>
<td>5.9</td>
<td>Confusion matrix (heat map) on MuHAVi-14 (camera 3 LOAO-CV)</td>
</tr>
<tr>
<td>5.10</td>
<td>Accuracy variations according to keyframe number on IXMAS</td>
</tr>
<tr>
<td>5.11</td>
<td>Confusion matrix of the preliminary results on IXMAS</td>
</tr>
<tr>
<td>5.12</td>
<td>Confusion matrix of the final results on IXMAS</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Introduction 135</td>
</tr>
<tr>
<td>B</td>
<td>Result of BPSO on MuHAVi-8 135</td>
</tr>
<tr>
<td>C</td>
<td>Result of BPSO on MuHAVi-14 141</td>
</tr>
<tr>
<td>D</td>
<td>Result of BPSO on IXMAS 149</td>
</tr>
</tbody>
</table>
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 intention 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.

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


