HUMAN ACTIVITY RECOGNITION BASED ON ELM USING DEPTH IMAGES

AHMED KAWTHER HUSSEIN

UNIVERSITI MALAYSIA PERLIS

2017

HUMAN ACTIVITY RECOGNITION BASED ON ELM USING DEPTH IMAGES

By

AHMED KAWTHER HUSSEIN

(1440211436)

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

School Of Computer and Communication Engineering
UNIVERSITI MALAYSIA PERLIS

2017

UNIVERSITI MALAYSIA PERLIS

DECLARATION OF THESIS

Author's full name : AHMED KAWTHER HUSSEIN

Date of birth : 13 AUGUST 1979

Title : HUMAN ACTIVITY RECOGNITION BASED ON ELM USING DEPTH IMAGES

Academic Session : 2016/2017

I hereby declare that the thesis becomes the property of Universiti Malaysia Perlis (UniMAP) and to be placed at the library of UniMAP. This thesis is classified as :

☐	**CONFIDENTIAL**	(Contains confidential information under the Official Secret Act 1972)*
☐	**RESTRICTED**	(Contains restricted information as specified by the organization where research was done)*
☑	**OPEN ACCESS**	I agree that my thesis is to be made immediately available as hard copy or on-line open access (full text)

I, the author, give permission to the UniMAP to reproduce this thesis in whole or in part for the purpose of research or academic exchange only (except during a period of _____ years, if so requested above).

Certified by:

SIGNATURE **SIGNATURE OF SUPERVISOR**

(NEW IC NO. / PASSPORT NO.) **NAME OF SUPERVISOR**

Date :_________________ Date : _________________

NOTES : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentially or restriction.

ACKNOWLEDGEMENT

First and foremost, I am most thankful to Allah S.W.T for giving me the strength, health and time to complete my thesis dissertation. I would like to express my special acknowledgement and deepest appreciation to my supervisors, Prof. Dr. Puteh binti Saad and Dr. Ruzelita binti Ngadiran for their kindness, encouragement and support towards the completion of this PhD thesis.

Special thanks to School of computer and Communication Engineering. I also want to thank to all lecturers in, School of computer and Communication Engineering, Universiti Malaysia Perlis for all the knowledge they shared, gave to me, the guidance and the friendly attitude to me when I need them.

I am extremely grateful to my parents and my sister for their love, prayers, caring and sacrifices for educating and preparing me for my future. My Special thanks goes to my friends those who were directly or indirectly supported me throughout the whole duration of my Project dissertation.

Last but not least, with deepest love, I dedicate my appreciation to my Parents, who has always been there with love, support, understanding and encouragement for all my endeavours. Although far away from them, they always concentrate on my life here and gave me great support and encouragement. I can feel their selfless love all the time. Your love and encouragement has been the most valuable thing in my life!

Thanks a lot.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACO	Ant colony optimization
ANN	Artificial Neural Networks
ASL	American Sign Language
BCA	Bee Colony Algorithm
BPNN	Back-propagation Neural Network
CCD	Comparative Coding Descriptor
CoG	Center Of Gravity
DAG	Directed Acyclic Graph
DBN	Dynamic Bayesian Network
DCSF	Depth Cuboid Similarity Feature
DFT	Discrete Fourier Transform
DMHIs	Two Depth Change Induced Motion History Images
DTW	Dynamics Time Wrapping
ELM	Extreme Machine Learning
FFT	Fast Fourier transform
FNN	Feed Forward Neural Network
GA	Genetic Algorithm
HAR	Human Activity recognition
HMCR	Harmony Search Considering Rate
HMM	Hidden Markov Model

HOG	Histogram Oriented Gradient
HOGPH	Histogram of Oriented Gradient Pattern History
HOJ3D	Histograms Of 3D Joint Locations
HOR	Histogram-of-Oriented-Rectangles
HSO	Harmony search optimization
K-NN	K-Nearest Neighbour
MHI	Motion History Image
PAR	Pitch Adjusting Rate
PMT	Projected Motion Template
ROP	Random Occupancy Pattern
SA	Simulated Annealing
SAX	Symbolic Aggregate approXimation
SIFT	Scale-Invariant-Feature-Transform
SLFNs	Single Hidden Layer Feed Forward Neural Networks
STF	Spatio-Temporal Feature
STIPs	Spatio-Temporal Interest Points
SVMs	Support Vector Machines
TS	Tabu Search
VMT	Volume Motion Template
3D-MHI	Dimensional Motion History Image

LIST OF SYMBOLS

zStart	Initial position of COG on z-axis
zMin	Minimum position of COG on z-axis.
zEnd	Final position of COG on z-axis
zHip	The position of hip center on z-axis
zLeftHand	The position of left hand on z-axis
zRightHand	The position of right hand on z-axis
Tr	The distance of the right hand from the hip
Tl	The distance of the left hand from the hip
Tm	The scaling factor that is used to limit the values of the two hands distance in the range [0, 1].
Trs	Transitional distance between right hand and hip after scaling
T_{ls}	Transitional distance between lef hand and hip after scaling
a_u	Number of actions
$m-1$	Number features that are used to identify
e	Number of experiments
s	Number of subjects
d	The total amount of data
n	Rows
m	Columns
η	The learning rate of the algorithm

Pengecaman Aktiviti Manusia Berdasarkan Imej-Imej Kedalaman

ABSTRAK

Aktiviti Pengiktirafan Manusia (HAR) telah mendapat faedah penyelidikan yang agak besar dalam beberapa dekad kebelakangan ini kerana aplikasi yang luas terutamanya dalam bidang perubatan, pengawasan, interaksi manusia-mesin, permainan dan hiburan. pengekstrakan ciri adalah satu langkah penting dalam algoritma HAR. Walau bagaimanapun, pada masa ini kebanyakan penyelidikan memberi tumpuan kepada ciri-ciri biasa seperti spatial domain dan domain frekuensi ciri-ciri. Ciri-ciri tersebut tidak mempunyai konteks dan tidak menyeluruh dalam alam semula jadi. Malangnya, membina ruang ciri yang menyeluruh daripada aktiviti manusia adalah sukar kerana keluasan dan jenis uncountable tindakan manusia. Ini membawa kepada masalah yang mencabar mereka bentuk sistem HAR yang menggunakan berasaskan konteks ciri pengekstrakan tindakan manusia. Dalam karya ini ruang ciri kontekstual komprehensif untuk pengecaman aktiviti manusia dibentangkan menggunakan imej kedalaman, jumlah fratures adalah 11. dalam aspek klasifikasi, mesin pembelajaran extrem hanya menggunakan lelaran tunggal dalam peringkat latihan untuk menentukan berat output. mesin pembelajaran extrem amat berkesan kerana ia cenderung untuk mencapai optimum global berbanding dengan kaedah pembelajaran FNN tradisional yang mungkin akan terperangkap dalam optimum tempatan. Kelemahan algoritma ELM memegang nombor terhingga darjah kebebasan untuk meniru satu set data yang diberikan. Ini darjah terhingga kebebasan adalah akibat daripada sifat rawak juga wajaran yang ditetapkan antara input dan lapisan tersembunyi. A potensi kemajuan yang mungkin dalam prestasi dalam kajian ini boleh dicapai dengan memberikan wajaran berdasarkan pengoptimuman functionan Objektif (ELM) menggunakan meta-heuristik itu. Harmony Cari Algoritma yang merupakan sebahagian daripada fungsi pengaktifan meta-heustric dan Tansig yang mengeluarkan neuron tersembunyi un diperlukan juga turut dipersembahkan dalam kerja ini. pendekatan yang dikemukakan oleh itu menyelesaikan masalah ijazah yang tidak terhingga kebebasan berat input serta menghadkan bilangan neuron dalam lapisan tersembunyi, sekali gus meningkatkan prestasi algoritma ELM itu. algoritma ELM dioptimumkan kemudiannya digunakan untuk melaksanakan pengelasan konteks yang dibangunkan berdasarkan ruang ciri. Ketepatan dicapai adalah 100% dalam latihan dan 94.95% semasa ujian dengan tindakan membuang dan 100% dalam latihan dan 100% dalam ujian tanpa tindakan membuang. pengoptimuman Gready MPE dengan HSO telah acehived ketepatan 94.95%. Selain itu, 60% daripada ciri-ciri yang telah mencapai ketepatan lebih 100%. Oleh itu, pendekatan yang boleh digunakan untuk melaksanakan pengecaman aktiviti manusia untuk pelbagai tujuan.

Human Activity Recognition Based On ELM Using Depth Images

ABSTRACT

Human Activity Recognition (HAR) has gained considerable research interest in recent decades due to its vast applications especially in the fields of medicine, surveillance, human-machine interaction, gaming and entertainment. Feature extraction is a key step in HAR algorithms. However, at present most research is focused on common features such as spatial domain and frequency domain features. Such features lack context and are not comprehensive in nature. Unfortunately, building a comprehensive feature space of human activities is difficult due to the vastness and uncountable nature of human actions. This leads to the challenging problem of designing a HAR system that uses context-based feature extraction of human actions. In this work a comprehensive contextual feature space for human activity recognition is presented using depth image,the total number of fratures is 11. in classification aspect, extrem learning machine uses only a single iteration in the training stage to determine the output weights. extrem learning machine is extremely effective as it tends to achieve the global optimum compared to the traditional FNN learning methods which might get trapped in a local optimum. The drawback of ELM algorithm holds an infinite number of degrees of freedom for approximating a given data set. These infinite degrees of freedom are a consequence of the random nature of the weights assigned between the input and the hidden layer. A possible potential improvement in performance in this research can be achieved by assigning the weights based on an objective functionan optimization of the (ELM) using the meta-heuristic. Harmony Search Algorithm which is a part of meta-heustric and Tansig activation function which remove un needed hidden neuron are also presented in this work. The presented approach hence solves the problem of the infinite degree of freedom of the input weights as well as restricting the number of neurons in hidden layer, thus increasing the performance of the ELM algorithm. The optimized ELM algorithm is then used to perform the classification of the developed context based on feature space. The accuracy achieved was 100% during training and 94.95% during testing with throw action and 100% during training and 100% during testing without throw action. Gready optimization of the ELM with HSO has acehived an accuracy of 94.95%. Moreover, 60% of the features have achieved an accuracy of over 100%. Thus, the approach can be utilized to perform the human activity recognition for various purposes.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Human Activity Recognition (HAR) is defined as a process of identifying human actions based on a set of observational data. HAR has gained considerable research interest in recent decades due to its vast applications especially in the field of medicine, surveillance, human machine interaction, gaming, and entrainment. A robust HAR system would change the way humans interact with technology. In the field of medicine, HAR systems can be used to assist rehabilitation processes and also used as a supervisory monitoring system for patients. In addition, HAR systems provide a huge impact on gaming and entertainment industries as it makes the media consumption a much more interactive and immersive experience for users.

Various HAR systems have been developed using different types of sensors. Much work has been done in vision-based HAR systems as in the case of Rougier, Meunier, St-Arnaud, and Rousseau (2011) and Raptis, M., & Sigal, L (2013). Vision-based systems usually suffer from certain limitations such as light sensitivity, background distinction, and environmental noise. Hence, other sensing methodologies such as RGB-D based sensing and data-based sensing have attracted researchers as in the case of Chattopadhyay and Maurya (2014), Ni et al. (2013), and Aggarwal and Ryoo (2011). Depth information in such systems serves as complimentary information and can improve the robustness of the HAR system.

Furthermore, depth cameras are insensitive to illumination changes and can extract skeleton features easier than vision-based systems. Wearable sensors are another way to extract data for HAR systems. Lara and Labrador (2013) presented a survey on the wearable sensor based HAR systems.

1.2 Problem Statement

HAR systems are, in essence, pattern recognition systems and hence have four main components i.e. sensing, preprocessing, feature extraction, and classification (Jain, Duin, & Mao, 2000). Even though HAR systems have a rich research base, there are certain issues lacking in the literature. The work presented hereby will focus on two issues in the existing HAR systems namely, feature extraction techniques and classification.

Feature extraction is a key step in HAR algorithms. In the available literature, most research is focused on common features such as spatial domain and frequency domain features. Such features however, are lack of context and not comprehensive in nature. Unfortunately, building a comprehensive feature space of human activities is difficult due to the vastness and uncountable nature of human actions. This leads to a challenging problem in designing a robust HAR system that uses context based feature extraction of human actions. This work therefore aims to develop a comprehensive contextual feature space for human activity recognition such that any similar actions having similar features performed by different individuals should be identified by the system.

Several classification algorithms such as Support Vector Machine (Khemchandani & Sharma, 2016), *K*-Nearest Neighbor (Chua, Leman, & Pham, 2011), Neural Network (NN) based algorithms, and etc. have been proposed in the literature. A common problem in those classifiers is the degree of freedom is insufficient to optimize the results of classification. This work will focus on a neural network based algorithm known as Extreme Learning Machine (ELM) (Huang et al., 2006).

Extreme Learning Machine (ELM) algorithm was adopted for human activity recognition system. Even though ELM algorithm has better performance than the conventional NN, there are two issues that ELM suffers from. First, the input weights of the ELM algorithm are randomly assigned. This leads to an infinite number of degree of freedom for approximating a data set. Second, the number of neurons in the hidden layer of ELM is assigned arbitrarily. Both of these issues limit the optimization performance of ELM algorithm. Niu, Ma, Li, Yan, and Li (2016) identify the same problems with the ELM algorithm and proposed a self-adjusting ELM (SA-ELM) based on the idea of the improved the teaching learning based optimization, the input-weights and the bias of hidden layer of extreme learning machine are adjusted with "teaching phase" and "learning phase" to minimize the objective function values. This work hence aims at optimizing the ELM algorithm in the two identified areas.

1.3 Motivation

The motivation for developing a robust HAR system arises from the many fold applications that such systems have. As mentioned, HAR systems provide huge benefit in the field of medicine and medical care. Loblaw, Nielsen, Okoniews, and Lakhani (2013) presented a system for respiratory sensing using an infrared camera. A similar sleeping respiration measurement system was also introduced by Kuo, Lee, and Chung (2010). Furthermore, a nursing care monitoring system for understanding human behavior using HAR was discussed in Liu, Chung, Chung, and Thonnat (2007). In addition, HAR systems are also applied in video surveillance. Vishwakarma and Agrawal (2013) provided a framework for application of human action recognition in video surveillance. Human Machine Interaction is yet another area of interest in HAR systems. Chaudhary, Raheja, Das, and Raheja (2011) presented a survey of human machine interaction using hand gestures.

HAR is needed for wide range of applications to provide the perception functionalities. For example: with HAR robots can deal with human in more effective way. Moreover, different services with HAR can be automated without the need to human being.

Due to the vast area of application of HAR systems, it is incentivizing to develop a robust and comprehensive HAR system that caters to such applications.

1.4 Research Questions

The research questions of this work are as follows:

i. How to develop a computational scheme for feature development that is scalable and comprehensive in nature.

ii. How to develop and optimize the ELM algorithm in order to get an improved performance.

1.5 Objectives

This research aims at achieving certain objectives which are as follows:

i. To develop a scalable scheme for feature development in HAR system.

ii. To develop an optimization scheme of Extreme Learning Machine (ELM) algorithm for classification purposes.

iii. To evaluate the developed schemes and algorithms using standard benchmark data.

1.6 Scope

The research presented in this work operates under a set of scope. First, an RGB-D camera was used for the detection of human activity. The Kinect sensor contains many advanced sensing hardware such as RGB camera and a depth sensor that provide full-body 3D motion capture. Furthermore, the Kinect is known for its capability of human skeletal tracking making it a better choice over a regular RGB camera. The research used only one stationary camera to collect the human activity data. There are no other sensors such as inertial measurement sensor, wearable sensors, and etc. that are used to collect the data. There are 10 action types: walk, sit down, stand up, pick up, carry, throw, push, pull, wave hands, clap hands. There are 10 subjects; each subject performs each action twice (as the design of the UTkincet data set). Lastly, the data collected is divided into training data and testing data such that a part of the data is used for training purposes and the remaining part is used for testing purposes.

1.7 Contribution

The contributions of the research presented are as follows:

i. The research presented a computational scheme for development of features in HAR system.

ii. A Harmony Search Optimization (HSO) was introduced to assign input weights for the ELM algorithm. The proposed (HSO) introduces roulette probability distribution to select the harmonics. The roulette probability distribution helps in selecting harmonics from the memory according to their fitness. Furthermore, a Tansig activation function was used to limit the number of hidden layer neurons in order to improve the performance of the ELM.

1.8 Thesis Outline

This thesis consists of five chapters namely, Introduction, Literature review, methodology, Experimental results, and Conclusion.

In chapter one, the overview of human activity recognition and their applications is described. The problem statements, objectives, and research scope are also presented.

Chapter two reviews the literature of features representation in HAR such as Space-time Features, Frequency Features, Local Descriptor, Optical Flow Based, and Skeleton Joints. In addition, the role of classifier in the HAR system is also described.

Chapter three presents the design of skeleton based features for human activity recognition. The chapter presents the process of joint reduction and defines different features used for HAR. The chapter discusses the action provided by UTkinect dataset and