



Scientific Computing Department
Faculty of Computer and Information Sciences
Ain Shams University

A Proposed Vision Purposive Architecture for Arabic Sign Language Recognition using Deep Learning Paradigm

Thesis submitted as a partial fulfillment of the requirements for the degree of
Master of Science in Computer and Information Sciences

By

Menna Tu-Allah Ahmed ElBadawy

Teaching Assistant at Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Under Supervision of

Prof. Dr. Mohamed Fahmy Tolba

Emeritus Professor in Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Prof. Dr. Howida AbdelFattah Shedeed

Professor Head of Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Dr. Ahmed Samir Elons

Doctor in Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

13 January 2018
Cairo

Acknowledgment

First of all, I would like to thank GOD for his endless blessings, for giving me the power and strength to complete this work and for giving me, supportive people.

Second, I would like to express my sincere gratitude to my supervisors; Prof. Dr. Mohamed Fahmy Tolba for his support, patience, guidance, and the special supervision experience he gave me. I am deeply thankful.

Also, I would like to thank Prof. Dr. Howida Shedeed for her encouragement and Dr. Ahmed Samir Elons for his scientific vision.

Third, I would like to thank my family for being available all the time and for the love they gave me through the years especially my lovely husband for his support and time were given to me to achieve my goals. Thank you for accepting me through the tough times and for always believing in me.

My dear friends who have helped me through the last time and kept on encouraging me to get this work done; Nareman Mahmoud, Alaa Salah, Naira Mousa, Mohga Mohamed, Pansy Shokry, Nermeen Mohamed and Ghada Hamed without you it would have been much harder.

Last but not least, I would like to thank all my professors, colleagues and students who kept on encouraging me. Thank you for being in my life.

Menna

Abstract

Arabic Sign Language ArSL is widely used in the Arabian countries due to its facilities to communicate with Hearing Impaired HI individuals. ArSL Recognition becomes vital for communication among HI persons and many technologies were employed to serve the recognition purpose and a lot of researches had been conducted to study either static or dynamic gestures.

Sign Language Recognition breaks the barrier between deaf and normal people. As the only way to communicate with HI persons is acting the meaning of words by hands and body, this way will deliver the meaning for either the normal or HI people. After increasing of Sign Language usage and importance, translation system for such languages become an essential need and also the requirement for a standard dictionary for ArSL.

The main purpose of the thesis is to develop an Arabic Sign Language Recognition ArSLR system which translates ArSL to Arabic words using deep techniques. The dataset which is used in our system is taken from the standard Arabic Sign Language dictionary published in 2005 [1]. The words are represented in 40 postures those were captured and collected from different signers and in different environments. The dataset was recorded with a digital camera from different angles.

The machine learning hand gesturing model is developed to recognize ArSL using two integrated deep models; Convolutional Neural Network CNN and Deep Belief Network DBN, as deep techniques have recently shown a significant gain for building hierarchical architecture for unlabeled data to learn, make usage of the self-learning attitude, and modeling the feature information learned from each network.

The system was tested with 25 words that are represented in 40 postures. The data is represented by images that are taken by a digital camera with resolution 1280x720. The system achieves an average accuracy 90% on the observed trained data and 86% on the unseen postures. It also achieves an average accuracy of 77% for gestures that are represented as sequences of trained postures. However, misclassification of postures in the sequence has been observed due to images closeness in multiple words' sequence which is lead to confusion between more than one gesture that happens to result in lower accuracy rate. The system is tested on hardware with Intel Core i5-2520M CPU, 2.5 GHz, 3.78 GB memory RAM, Intel HD Graphics 3000, and Windows8 64 bit operating system.

After the new generation of input sensors, ArSL has to make use of the sensors' advantages. New modern sensors are used and tested for the recognition purpose. Also, their effects on the accuracy rate are observed to achieve the best model that results in high accuracy rate with low computational power. A system based on Leap Motion data input is developed to show the impact of the new data format. Also, the integration between Leap Motion sensors with number of other input sensors are fed to a system to maximize each sensor's advantage and increase the recognition accuracy. And one additional system is developed to include a new features set other than hand gestures to enlarge the features space used and so the accuracy obtained.

List of Publications

1. Menna Ahmed, A. S. Elons, and Howida A. Shedeed. "Facial expressions recognition for Arabic Sign Language translation". 9th International Conference on Computer Engineering Systems (ICCES), pp 330-335, IEEE, December 2014.
2. Menna Ahmed, A. S. Elons, and Howida A. Shedeed. "Arabic sign language recognition using leap motion sensor". 9th International Conference on Computer Engineering Systems (ICCES), pp 368-373, IEEE, December 2014.
3. Menna ElBadawy, A. S. Elons, Howida A. Shedeed, and M. F. Tolba. "A Proposed Hybrid Sensor Architecture for Arabic Sign Language Recognition". Advances in Intelligent Systems and Computing, vol 323, pp 721-730. Springer, Cham, 2015.
4. Menna ElBadawy, A. S. Elons, Howida A. Shedeed, and M. F. Tolba. "Arabic Sign Language Recognition with 3D Convolutional Neural Networks". 8th International Conference on Intelligent Computing and Information Systems (ICICIS), vol 2, pp 66-71, IEEE, December 2017.
5. Menna ElBadawy, A. S. Elons, Howida A. Shedeed, and Mohamed F. Tolba. "Arabic Sign Language Recognition (ArSLR) Using Deep Techniques". Neural Computing and Applications International Journal, 2017 (Submitted).

Table of Contents

Acknowledgment	II
Abstract	III
List of Publications	V
Table of Contents	VI
List of Figures	VIII
List of Tables	IX
List of Algorithms	X
List of Abbreviations	XI
Chapter 1. Introduction	2
1.1 Motivation	2
1.2 Sign Language History	3
1.3 Data Description	3
1.4 Thesis Organization	4
Chapter 2. Related Work	7
2.1 Pre-processing based researches	7
2.2 Utilized inputs' based researches	8
2.3 Discussion	13
Chapter 3. Scientific Background	15
3.1 Convolutional Neural Network (CNN)	16
3.2 Deep Belief Network (DBN)	17
Chapter 4. Purposive Architecture	24
4.1 3D CNN	24
4.1.1 Data Specification	24
4.1.2 Preprocessing	25
4.1.3 Training	28
4.1.4 Experimental Results	29
4.2 Integrated Deep Models	31
4.2.1 Recognition System Phases	32
4.2.1.1 Preprocessing Step	32
4.2.1.2 Best Frame	32
4.2.1.3 Feature Extraction Step	33
4.2.1.4 Training Step	34
4.2.2 Experimental Results and Discussion	35
4.3 Discussion	40
Chapter 5. Facial Expressions Recognition	45

5.1	Recognition System	46
5.1.1	Facial Expressions Recognition.....	48
5.1.2	Recursive Principle Component Analysis (RPCA)	49
5.2	Data and Results.....	52
5.3	Conclusion.....	55
Chapter 6.	ArSLR Using Modern Sensors	57
6.1	Sensors	57
6.1.1	2D Sensors	58
6.1.2	3D Sensors	59
6.1.3	Kinect and Depth Cameras Module.....	59
6.1.4	Leap Motion Sensor	60
6.2	ArSLR Using Leap Motion Sensor	61
6.2.1	Learning Methodology	62
6.2.3	Experimental Results	65
6.3	Proposed Hybrid Sensors	67
6.3.1	Data Specification.....	69
6.3.2	Experimental Results	70
6.4	Conclusion.....	72
Chapter 7.	Conclusion and Future Work	75
7.1	Conclusion.....	75
7.2	Future Work	76
References	79
Appendix A.	Sample images from the database used	88

List of Figures

Figure 3-1 The proposed deep architecture	15
Figure 3-2 Input Data of size 5x5x3	16
Figure 3-3 DBN Feed Forward Phase	18
Figure 3-4 Reconstruction data process using the same weight matrix.	19
Figure 4-1 ArSLR System with 3D CNN.....	24
Figure 4-2 Sample images from the Dataset used.	25
Figure 4-3 The steps of Scoring Algorithm.....	28
Figure 4-4 Recognition system using Deep Architecture.....	32
Figure 4-5 Sample images after body tracking and resizing	33
Figure 4-6 DBN Architecture	34
Figure 4-7 Sample images from used dataset.	37
Figure 4-8 Two similar images for two different classes.	39
Figure 5-1 Four expressions of a person with a chain.....	47
Figure 5-2 Integration of facial expression module with ArSLR System. ...	47
Figure 5-3 The Face Tracking output.	48
Figure 5-4 The Facial Expression Recognition Module.....	49
Figure 5-5 Facial Expressions Recognition System Accuracy.....	53
Figure 6-1 a-Virtual Technologies Cybergloves b-5th wireless dataglc.	59
Figure 6-2 Kinect images in sign language recognition [57].	60
Figure 6-3 Leap Motion Device [58].....	61
Figure 6-4 The leap motion interaction and its resulting data.....	62
Figure 6-5 Multilayer Perceptron Structure.....	64
Figure 6-6 Recognition module using Leap Motion data.....	65
Figure 6-7 Capturing lab.....	67
Figure 6-8 Hybrid Sensors Recognition System.	68

List of Tables

Table 2-1 Sign language related work	10
Table 4-1 Configuration table for the input depth	28
Table 4-2 The Classification results with the CNN configuration.	30
Table 4-3 Dataset used in the system	36
Table 4-4 Sensitivity analysis for CNN and DBN	41
Table 4-5 Comparative results	42
Table 5-1 Face tracking results for 5 different persons.	53
Table 5-2 Recognition Accuracy against the number of neurons in a hidden layer.....	54
Table 6-1 The recognition rate against the number of words.....	66
Table 6-2 Dataset used in the system.	69

List of Algorithms

Algorithm 3-1 Train RBM Algorithm	20
Algorithm 3-2 Pre-Train DBN Algorithm.....	21
Algorithm 3-3 Train DBN Algorithm.....	22
Algorithm 4-1 Scoring Algorithm	26
Algorithm 4-2 Canny Edge Detection Algorithm [25].....	27
Algorithm 5-1 RPCA Algorithm [44].....	51

List of Abbreviations

<u>Abbreviation</u>	<u>Stands for</u>
ANN	Artificial Neural Network
ArSL	Arabic Sign Language
ArSLR	Arabic Sign Language Recognition
CD	Contrastive Divergence
CNN	Convolutional Neural Network
DBN	Deep Belief Network
HI	Hearing Impaired
HMM	Hidden Markov Models
ITIDA	Information Technology Industry Development Agency
LMC	Leap Motion Controller
MCMC	Gibbs Monte-Carlo Markov Chain
MDC	Minimum Distance Classifier
MKNN	Modified K-Nearest Neighbors
MLP	Multilayer Perception
PCA	Principle Components Analysis
RBM	Restricted Boltzmann Machines
RPCA	Recursive Principle Components Analysis
SIFT	Scale Invariant Feature Transform
SL	Sign Language
SLR	Sign Language Recognition
3D CNN	3 Dimension CNN

Chapter 1

Introduction

Chapter 1. Introduction

1.1 Motivation

Hearing Impaired (HI) individuals usually acquire the same level of mental capability as the normal hearing persons in terms of studying. The term 'deaf and dumb' is not practical its usage since HI individuals are only lacking in their hearing capability, not their intelligence level [2].

People with disabilities meet barriers of all types. However, technological capabilities reduce many simple barriers. By using computing technology for tasks such as reading and writing documents, communicating with others, and searching for information on the Internet, students and employees with disabilities are capable of handling a wider range of activities independently.

A gesture is a form of non-verbal communication made with part of the body and used instead of verbal communication (or in combination with it). Most people use gestures and body language in addition to words when they speak. A sign language is a language which uses gestures instead of sound to convey meaning combining hand shapes, orientation and movement of the hands, arms or body, facial expressions and lip-patterns.

For many HI individuals, SL is the principal mean of communication. The main problem is that few people who are not themselves HI ever learn to sign. Another problem is that many of them (HI people) are also not able to read or write a spoken language. These problems increase the isolation of HI people; they may be confined in many of their interactions to communicate only with other HI people. In Egypt, the number of HI people according to the last study done by the "Central Agency for Public Mobilization and Statistics" in 2010 is around 4 million [3]. 70% of this number can't read or write the

Arabic language. Due to the need of both normal and impaired hearing people, researchers started to explore the automation for the translation process between natural language and SL.

1.2 Sign Language History

Researchers all over the world started to employ the technology to break the barriers for HI individuals; they tried to build translation systems from SL to spoken languages and vice versa. In our research, the focus was on translating the ArSL signs into Arabic text.

Contrary to public common sense, SL is not universal. Where people speak a different phonetic language there is also a different SL. Besides the locality nature of sign languages, ArSL is not standard. Recently in 2005, the standard ArSL Dictionary was accredited and publically published to the Arabian deaf community and then updated in 2008 [1]. Although this dictionary is not widely used in various countries, the Arabian Gulf countries encouraged their deaf community to learn and use this dictionary.

1.3 Data Description

In our thesis, a lot of techniques were conducted to achieve best accuracy and performance. In order to measure each technique performance, conclude results, and further work, a dataset of ArSL is used. The dataset contains RGB images taken from a single digital camera with size of 1280x720 and are saved in a bitmap format.

Each posture was trained with a number of samples (2, 3, 5 samples) to learn the general features and accept the variety of the postures of different people, as two people were used in this database with two different

environments. One signer used in a light and simple background with low noise level, and the other one used in a noisy background filled with objects.

1.4 Thesis Organization

This thesis is organized into seven chapters including this one. Their contents are described briefly as follows:

Chapter 2 (*Titled: Related Work*) Describes the different techniques developed in SLR field.

Chapter 3 (*Titled: Scientific Background*) Describes the techniques used in the purposive system developed in this thesis.

Chapter 4 (*Titled: Purposive Architecture*) Introduces the developed purposive deep architecture and measure its importance and results. In this chapter, new techniques used to prove its benefits in the ArSLR.

Chapter 5 (*Titled: Facial Expressions Recognition*) Proposes facial expression recognition algorithm that utilized and employed in the ArSLR process. In this chapter, integrated recognition system used to show the essence of the facial expressions in the recognition process.

Chapter 6 (*Titled: ArSLR Using Modern Sensors*) Divided into two parts. The first part contains an ArSLR system that utilizes the use of leap motion sensor to input the data that feed the system and shows the impact of new data format on the overall recognition rate. And Second part contains an ArSLR system that utilizes the use of different sensors to