

People's Democratic Republic of Algeria.
Ministry of Higher Education and Scientific Research



Machine learning for sign language recognition: Application on smart buildings

Submitted September 2020, in partial fulfillment of
the conditions for the award of the degree **Master Software engineering**.

Mohamed Mehdi BOURAHLA

ID: M201532027293

Supervised by Roufaida LAIDI and Abdallah KAMECH

Computer Science department
Saad Dahleb Blida University

Date: 23 / 09 / 2020

In front of the jury:

Hafida ABED

Fatima Zahra ZAHRA

Promotion: 2019/2020

Abstract

Recent advances in mobile computing, wireless sensing and communication technologies, consumer electronics have modernized our cities and living environments. Buildings, roads, and vehicles are now empowered with a variety of smart sensors and objects that are interconnected via machine-to-machine communication protocols, accessible via the Internet, to form what is known as the Internet of Things (IoT). The power of IoT expands when coupled with Machine Learning, since the later offer techniques that allow analyzing the vast amount of data generated by sensors and actuators. Smart buildings are an appealing example of IoT and machine learning applications offering higher energy saving and occupants satisfaction through dynamic control.

Vocal virtual assistants (e.g., Amazon Alexa, Google Home) are now a central component of the smart house. However, they are not adapted to deaf and mute people who communicate using sign language. Efficient alternative communication means inside the house are required to assist the interaction of deaf and hearing-impaired people.

The main goal of this thesis is to conceive and realize a solution based on machine learning for sign language recognition that allows the control of a smart home environment through gestures.

Keywords: Smart buildings, Machine learning, Sign language, Disabled people, Human-Computer interaction.

Résumé

Les progrès récents dans l'informatique mobile, la détection sans fil et les technologies de communication, l'électronique grand public ont modernisé nos villes et nos milieux de vie. Les bâtiments, les routes et les véhicules sont désormais dotés de divers capteurs et objets intelligents interconnectés par des protocoles de communication machine à machine, accessibles via Internet, pour former ce qu'on appelle l'Internet des objets (IoT). La puissance de l'IoT se développe lorsqu'il est couplé avec le Machine Learning, puisque ce dernier offre des techniques qui permettent d'analyser la grande quantité de données générées par les capteurs et les actionneurs. Les bâtiments intelligents sont un exemple attrayant d'applications d'IoT et d'apprentissage automatique offrant des économies d'énergie plus élevées et la satisfaction des occupants grâce au contrôle dynamique.

Les assistants vocaux virtuels (par exemple, Amazon Alexa, Google Home) sont désormais une composante centrale de la maison intelligente. Cependant, ils ne sont pas adaptés aux personnes sourdes et muettes qui communiquent en langage gestuel. Des moyens de communication alternatifs efficaces à l'intérieur de la maison sont nécessaires pour faciliter l'interaction des personnes sourdes et malentendantes.

L'objectif principal de cette thèse est de concevoir et de réaliser une solution basée sur l'apprentissage automatique pour la reconnaissance du langage des signes qui permet le contrôle d'un environnement domestique intelligent à travers des gestes.

Mots clés: Bâtiments intelligents, apprentissage automatique, langue des signes, personnes handicapées, interaction homme-machine.

ملخص

لقد حدث التطور الأخير في مجال الحوسبة المتنقلة والاستشعار اللاسلكي وتقنيات الاتصال، والإلكترونيات الاستهلاكية مدننا وبيئتنا المعيشية. أصبحت المباني والطرق والمركبات الآن مزودة بمجموعة متنوعة من المستشعرات والأجسام الذكية المتصلة ببعضها عبر بروتوكولات الاتصال من جهاز إلى جهاز، والتي يمكن الوصول إليها عبر الإنترنت، لتكوين ما يُعرف بإنترنت الأشياء. تتوسع قوة إنترنت الأشياء عند اقترانها بتعلم الآلة، حيث إن أحدث تقنيات تتيح تحليل كمية كبيرة من البيانات التي تولدها أجهزة الاستشعار والمشغلات. تُعد المباني الذكية مثلاً جذاباً على تطبيقات إنترنت الأشياء وتعلم الآلة التي توفر طاقة أعلى ورضا راكبي السيارة من خلال التحكم الديناميكي.

أصبح المساعدون الافتراضيون الصاحب (على سبيل المثال، Amazon Alexa و Google Home) الآن مكوناً أساسياً في المنزل الذكي. ومع ذلك، فإنها غير مكيفة مع الصم وكم الأشخاص الذين يتواصلون باستخدام لغة الإشارة. وسائل الاتصال البديلة الفعالة داخل المنزل مطلوبة للمساعدة في تفاعل الصم وضعاف السمع.

والهدف الرئيسي لهذه الفرضية هو تصور وتحقيق حل يستند إلى تعلم الآلة للتعرف على لغة الإشارات التي تسمح بالتحكم في بيئة منزلية ذكية من خلال الإيماءات.

الكلمات المفتاحية: المباني الذكية ، التعلم الآلي ، لغة الإشارة ، المعاقون ، التفاعل بين الإنسان والحاسوب.

Dedications

I dedicate this thesis to my parents for whom a whole book of dedications would not be enough to express my gratitude to them. For all of their support, encouragement, love and sacrifices that have brought me to this point in my life and continue to keep me moving forward. May Allah preserve them as well as all moms and dads in the world.

To my brothers and beloved sister who always believed in me and saw me as the little computer kid in the family.

My dear friends Abdelouahab, Chakib, Chouaib, Hadjer, Mohamed, Oussama, Rym for their love and support, may they find here the expression of my gratitude.

All my CSC club family who over the past five years have taught me to overcome challenges and make a difference and have opened the door to many opportunities for me.

Acknowledgements

First of all, I would like to thank ALLAH the greatest for all the patience, health and strength he has given me to do this modest work. Every letter typed on this thesis and every information withheld is by his grace glory to him.

I would like to thank Ms. Roufaida LAIDI my supervisor at CERIST for her help, her infinite patience, her availability and her golden advice which fueled my reflection, and especially for her efforts and all her contributions and corrections along the way.

I would like to express my gratitude to my supervisor Mr. Abdallah KAMECHE at Saad Dahleb university for having mentored me for this project as well as for all the knowledge I have acquired from him.

I also thank all the professors of the computer science department of Saad Dahleb University and any contributor to the knowledge I have acquired in recent years without whom I will not be where I am.

My gratitude to our delegate Mr. Imade ANANE and to Ms. Rym BOUCHETERA for their support to all our promotion.

All my thanks to Mr. Chakib KERRI for supporting me and helping me prepare the presentation of this project. Thanks to Mrs. Imene HENNI MENSOUR for inspiring me and opening the door to many opportunities.

Contents

Abstract	i
Acknowledgements	iv
GENERAL INTRODUCTION	1
1 Background and Context	3
1.1 Introduction	3
1.2 Sign Language	3
1.3 Smart Homes	4
1.4 Machine Learning	5
1.4.1 Data Collection	6
1.4.2 Data Preparation	6
1.4.3 Model Selection	7
1.4.4 Model Evaluation	8
1.4.5 Hyperparameter Tuning	9
1.5 Deep Learning	9
1.5.1 Neurons	10
1.5.2 Activation Functions	10
1.5.3 Artificial neural networks	11
1.5.4 Convolutional neural network	12

1.5.5	Recurrent neural network	13
1.6	Context	17
2	State of the art	18
2.1	Introduction	18
2.2	Vision-based gesture recognition	19
2.2.1	Data acquisition	20
2.2.2	Image processing	20
2.2.3	Segmentation	20
2.2.4	Tracking	21
2.2.5	Dimensionality reduction	22
2.2.6	Classification	22
2.2.7	Results	23
2.2.8	Discussion	25
2.2.9	Conclusion	25
2.3	Sensor based gesture recognition	26
2.3.1	Data acquisition	26
2.3.2	Signal preprocessing	26
2.3.3	Data segmentation	27
2.3.4	Feature extraction	28
2.3.5	Classification	30
2.3.6	Results	30
2.3.7	Discussion	31
2.3.8	Conclusion	32
2.4	Wireless based gesture recognition	32
2.4.1	Types of wireless transmission used in SLR	32
2.4.2	Signal preprocessing	34

2.4.3	Segmentation	34
2.4.4	Feature extraction	35
2.4.5	Classification	35
2.4.6	Results	36
2.4.7	Discussion	37
2.4.8	Conclusion	38
2.5	Summary and Discussion	38
2.5.1	Used approaches in existing works	38
2.5.2	Comparison criteria	39
2.5.3	Discussion	40
2.5.4	Adopted approach	41
2.6	Conclusion	41
3	System Overview	42
3.1	Introduction	42
3.2	System overview	42
3.2.1	Word-based architecture	42
3.2.2	Sequence-based architecture	50
3.3	Conclusion	52
4	Implementation & Evaluation	53
4.1	Introduction	53
4.2	IMPLEMENTATION	53
4.2.1	SignFi dataset	53
4.2.2	Sentences dataset	55
4.2.3	Tools & libraries	56
4.2.4	Experimentation	61
4.3	Evaluation	69

4.3.1	Recognition Accuracy	69
4.3.2	Parametric study	75
4.3.3	Comparison between the two methods	79
4.3.4	Comparison with other works	80
4.4	Deployment	82
4.5	Conclusion	84
	GENERAL CONCLUSION	85
	Bibliography	86

List of Tables

1.1	The commonly used activation functions	11
2.1	Results obtained by various vision-based gestures recognition	24
2.2	Taxonomy of extracted features in three domains	29
2.3	Results obtained by various sensor-based gestures recognition	31
2.4	Results obtained by various wireless-based gestures recognition	37
2.5	Summary of the currently used approaches in hand gesture recognition . . .	39
2.6	Comparison between the existing approaches in hand gesture recognition .	40
4.1	SignFi dataset summary	54
4.2	The selected ASL words	56
4.3	Comparison of word-level SLR systems	81
4.4	Comparison of Sentence-level SLR systems	82

List of Figures

1.1	Machine learning: a new programming paradigm [7]	6
1.2	Machine learning algorithm process	6
1.3	Artificial neuron	10
1.4	Convolutional neural network architecture	12
1.5	Recurrent neural network architecture	14
2.1	Vision based hand gesture recognition process	19
2.2	Sensor-based gesture recognition process	26
3.1	Word-based architecture overview	43
3.2	CSI representation of MIMO	44
3.3	Wavelet analysis results of air-drawn gestures [54]	45
3.4	Raw CSI measurements do not capture how CSI phases change over sub-carriers and sampling time [55]	46
3.5	Architecture of the CNN model	48
3.6	Sentence structures	49
3.7	Architecture of the RNN model	50
3.8	Sequence-based architecture overview	50
3.9	Architecture of the CRNN model	52
4.1	Floor plan and measurement settings of the lab and home environments from SignFi [55]	55

4.2	Pytorch vs TensorFlow: Number of Unique Mentions	61
4.3	The model's Keras implementation	63
4.4	Architecture and parameter settings of the CNN model	64
4.5	Architecture and parameter settings of the LSTM model	67
4.6	ConvLSTM structure with hyperparameters	68
4.7	Laboratory environment performance evaluation	70
4.8	Home environment performance evaluation	71
4.9	Home and lab combination performance evaluation	72
4.10	Multiple users scenario performance evaluation	73
4.11	Sentence-level performance evaluation	74
4.12	ConvLSTM model evaluation	75
4.13	Batch normalization impact on the accuracy	76
4.14	Dropout impact on the FC layer	77
4.15	Dropout impact on the convolutional unit	77
4.16	Average vs. Max pooling layer	78
4.17	Training time and average accuracy RNN vs. LSTM vs. GRU	79
4.18	Architectures' black boxes representation	79
4.19	Web application architecture	83
4.20	Screenshots from the web application	84