PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

LARBI TEBESSI UNIVERSITY

FACULTY OF EXACT SCIENCES AND SCIENCES OF NATURE AND LIFE

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE

Takalem : A deep learning approach for Algerian sign language recognition

Submitted by
Fathi Abdelmalek

Supervised by
Pr. Mohamed Amroun
Dr. Issam Bendib

May 16, 2023

Dedication

To my dear mother
To my dear father
To my dear sisters
To my dear brothers
To my adorable grandmother
To all my uncles, aunts and cousins
To all my friends and colleagues
To every one who have supported me

Fathi Abdelmalek

Acronyms

AJSL Algerian jewish sign language

ANN Artificial Neural Network

AR Augmented Reality

ASP Algerian sign language

CNN Convolutional Neural Network

CRF conditional random field

DL Deep Learning

DNN Deep Neural Network

FSL French sign language

HMM Hidden Markov Model

IMU Inertial Measurement Unit

LSTM Long-Short Term Memory

ML Machine Learning

RNN Recurrent Neural Network

SL Sign language

SLR Sign language recognition

SVM Support Vector Machines

VR Virtual Reality

Abstract

Sign language recognition plays a crucial role in facilitating communication between individuals with hearing impairments and non-signers. However, the lack of an Algerian Sign Language (ASP) dataset and limited time constraints pose challenges in developing a customized dataset. In this study, an existing American Sign Language (ASL) dataset is utilized, and a Long Short-Term Memory (LSTM) model is employed for sign classification. The LSTM model achieves an accuracy of 86.16% for character-level classification and an impressive 98.05% for word-level classification. Remarkably, the model exhibits minimal errors in word classification and infrequent errors in character classification, primarily due to the selection of the most frequently predicted sign from a pool of 150 predictions per gesture. While these results demonstrate the effectiveness of the model.

***Keywords*—** Sign Language Recognition, Machine Learning, Deep Learning, LSTM

Contents

I State of the art **1**

1 Sign languages and sign language recognition **2**

1 Introduction 2
2 Sign languages 2
2.1 Sign languages in Algeria 3
2.2 Algerian Sign Language 3
3 Sign language recognition 3
3.1 Definition 3
3.2 Sign language recognition (SLR) applications and use . . 4
3.2.1 Sign language (SL) translation 4
3.2.2 Virtual and augmented reality 4
3.2.3 Research and linguistics 5
4 Deep Learning 5
5 Conclusion 6

2 Related works **7**

1 Introduction 7
1.1 Sign Language Recognition: An Overview 7
1.2 Motivation for Exploring Related Works 7
1.3 Objectives of the Chapter 8
1.3.1 Review and Analyze Existing Methodologies . . 8
1.3.2 Explore Datasets for Training and Evaluation . 8
1.3.3 Evaluate Performance Metrics 8
1.3.4 Identify Advancements and Future Directions . . 8
2 Methods of sign language recognition 9
2.1 Vision-Based Sign Language Recognition Methods 9
2.2 Data Glove-Based Sign Language Recognition Methods . 9
2.3 Hybrid Approaches in Sign Language Recognition 9
3 Datasets for Training and Evaluation 10
3.1 Text-to-Sign Synthesis Methods 10
3.2 Speech-to-Sign Synthesis Methods 10
3.3 Challenges and Future Directions in Synthesis 11
4 Evaluation Metrics and Performance Analysis 11
5 Synthesis 12
6 Conclusion 12

II Contribution **13**

1 Conception of Takalem gloves **14**

1 Introduction 14
2 Hardware architecture and configuration 14
3 Dataset collection and preprocessing 14
4 Proposed deep learning architecture 14
5 Conclusion 14

2 Results and discussion **15**

1 Introduction 15
2 Evaluation criteria 15
3 Experiments 15
4 Discussion 15
5 Conclusion 15

List of Figures

List of Tables

Part I

State of the art

Chapter 1

Sign languages and sign language recognition

1 Introduction

Deaf, mute, and hard-of-hearing people, have their language system, its Sign language (SL), there are over 300 different SL used in the world, and each one has a unique grammar and vocabulary, one of them is Algerian sign language (ASP), is used by more than 240,000 individuals around Algeria. Sign language recognition (SLR) technology aims to provide a means of automatically interpreting SL and translating it into text or speech, to facilitate the communication process between normal and deaf/mute people.

Recent advancements in wearable sensor technology have made it possible to develop SLR systems using sensor gloves. These gloves contain multiple sensors that measure various parameters related to hand movement and position. Machine Learning (ML) and Deep Learning (DL) algorithms can then be used to analyze the sensor data and recognize the corresponding SL gestures.

In this chapter, we will provide an overview of sign languages and SLR technology. We will discuss the challenges associated with SLR and the potential benefits of developing such technology.

2 Sign languages

SL is one of the methods of communication, which is defined as a set of visual symbols or gestures that are used in a very systematic way for words, concepts, or ideas of a language.[1]

Despite the complexity of sign languages, they can be broken down into smaller units, such as signs, hand shapes, and movements. Sign languages typically use a combination of these units to form words and sentences. For example, ASP uses hand shapes, movements, and facial expressions to convey meaning.

Sign languages are not just visual representations of spoken languages, they are unique and independent languages with their syntax, grammar, and vocabulary. Recognizing and understanding them is therefore crucial for effective communication between hearing and deaf communities. In recent years, there has been increasing interest in developing technology to aid sign language recognition and translation.

2.1 Sign languages in Algeria

There is no official ASP document or reference except for a dictionary published recently by the Algerian government. It contains some signs used by the deaf and other signs borrowed from the Old French sign language (FSL).[2]

There is also Algerian jewish sign language (AJSL), an old SL which developed in several Jewish communities in the region of M'zab, Algeria, which is located in the northern part of the Sahara desert.[3]

2.2 Algerian Sign Language

ASP is an SL derived from FSL, used by the deaf/mute community of Algeria. It was officially recognized by the Algerian law as official SL in Algeria in May 2002[4]. Technically, it is a visual-gestural language that uses hand shapes, movements, and facial expressions to convey meaning. Therefore, this community is often excluded from basic communication, this has caused many deaf Algerians to go without access to education, employment opportunities, and other basic rights. For that, The government of Algeria opened many deaf schools around the country to teach them the language itself, and basic education like any normal person.

Even with the existence of deaf schools, the teachers themselves are not qualified neither master ASP, all of them are hearing individuals who hold different degrees which are not related to deaf education or SL, and they have never been trained to use the language before they get hired, some teachers attend training courses to learn alphabet only. And this makes it difficult for those pupils to get basic education and go even to middle or high school and college, most of them are marginalized and can only be manual workers, they are denied access to a high-quality education that meets their special needs to improve their lives and live as equal to their peers.[2]

3 Sign language recognition

3.1 Definition

SLR is the process of interpreting and translating the gestures, movements, and facial expressions of SL into written or spoken language. It involves capturing, processing, and analyzing data from various sensors and devices such as gloves and cameras.

The task of SLR is a challenging one due to the complexity and variability of sign languages. They are rich and expressive, and there are many different sign languages used around the world, each with their own unique vocabulary, grammar, and syntax. Moreover, sign languages are not universal, meaning that a sign used in one language may have a completely different meaning in another language.

Despite these challenges, significant progress has been made in the field of SLR in recent years, thanks to advances in sensor technology, computer vision, ML and DL. Researchers have proposed a wide range of approaches to tackle the problem of SLR, including rule-based systems, template matching, Hidden Markov Model (HMM), Artificial Neural Network (ANN), and DL methods.

In recent years, DL-based methods, particularly Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), have shown promising results in SLR, achieving state-of-the-art performance on several benchmark datasets. These methods can learn meaningful representations of the data directly from raw input, which makes them well-suited to complex and dynamic data like SL.

3.2 SLR applications and use

SLR technology has the potential to empower deaf and hard-of-hearing individuals by enabling them to communicate more effectively with the wider community. Moreover, it can be used in other areas alongside basic use, here are the main areas of use of SLR technology.

3.2.1 SL translation

SL translation is one of the primary applications of SLR technology. By capturing and analyzing SL gestures, SLR systems can convert them into written or spoken language, enabling effective communication between deaf and hearing individuals. This technology can be utilized in various contexts, such as educational settings, customer service centers, healthcare facilities, and public institutions. For example, in education, SLR translation can support deaf students by providing real-time SL interpretation during lectures, ensuring they have access to the same educational content as their hearing peers. In customer service, SLR translation can facilitate communication between deaf customers and service representatives, improving accessibility and customer satisfaction. Additionally, SLR translation can be integrated into translation apps or devices, allowing deaf individuals to communicate with individuals who do not understand SL, bridging the communication gap.

3.2.2 Virtual and augmented reality

SLR technology can enhance the immersive experience of virtual and augmented reality environments by incorporating SLR capabilities. In VR/AR applications, SLR enables users to interact with the virtual world using SL gestures, making the experience more intuitive and inclusive for deaf users. For instance, in a Virtual Reality (VR) game, SLR can recognize and interpret

SL commands as input, allowing players to control their characters or perform actions using SL gestures. In Augmented Reality (AR), SLR can be used to overlay real-time SL translations onto the user's field of view, enabling seamless communication between deaf and hearing individuals in AR scenarios. This integration of SLR in VR/AR not only enhances entertainment experiences but also opens up new possibilities in training simulations, remote collaboration, and interactive storytelling.

3.2.3 Research and linguistics

SLR technology plays a crucial role in linguistic research and the study of sign languages. By capturing and analyzing SL data, SLR systems provide valuable insights into the structure, grammar, and syntax of sign languages. Researchers can use SLR to examine the linguistic patterns and variations within SL communities, contributing to the documentation and preservation of sign languages. SLR can also assist in studying the cognitive aspects of SL processing and acquisition. Furthermore, SLR can be used to create SL corpora and databases, which serve as valuable resources for linguistic analysis and comparison across different SL systems. Overall, SLR technology empowers researchers and linguists to delve deeper into the intricate nature of sign languages, leading to a better understanding and appreciation of deaf culture and communication.

4 Deep Learning

DL, a subfield of ML, has revolutionized various domains by enabling the automatic learning of intricate patterns and representations from raw data. In SLR, DL proves to be a promising approach due to its ability to process and analyze complex visual and temporal information.

ANN serve as the foundation of DL models. These computational structures are inspired by the biological neurons in the human brain and consist of interconnected layers of artificial neurons. CNN are widely employed in image recognition tasks, including the analysis of hand gestures in sign language. CNNs leverage convolutional layers to automatically learn spatial features from input images, enabling the model to discern intricate patterns and variations.

In addition to CNNs, RNN play a vital role in sign language recognition. RNNs are particularly effective in handling sequential data, which is inherent in SL gestures. RNNs can capture the temporal dependencies in a sequence of gestures and retain information over time. Long-Short Term Memory (LSTM) networks, a type of RNN, excel in modeling long-range dependencies and have demonstrated promising results in SLR tasks.

Training and optimization are essential steps in the deep learning workflow. During the training phase, a neural network is exposed to labeled training data, and its internal parameters are adjusted to minimize the discrepancy between predicted and actual outputs. Backpropagation, a gradient-based optimization algorithm, is commonly used to update the model's parameters. The training process is iterative, and the model undergoes multiple epochs to improve its performance gradually.

Transfer learning, a technique widely employed in DL, has shown promise in SLR. It involves leveraging pre-trained models trained on large-scale datasets and fine-tuning them on smaller SL datasets. Transfer learning allows the model to benefit from the learned representations in the pre-trained models, leading to improved performance and reducing the need for extensive training data.

Despite the remarkable potential of DL in SLR, it comes with its own set of challenges and limitations. One major challenge is the scarcity of annotated SL datasets, which hinders the training process and necessitates domain-specific data collection efforts. Overfitting, a phenomenon where the model performs well on training data but fails to generalize to unseen data, is another challenge that requires careful regularization techniques.

Moreover, DL models often demand significant computational resources, including high-performance computing units and memory capacity. The interpretability of DL models is another concern, as they are often seen as "black boxes" that make it challenging to understand the internal decision-making process.

In conclusion, DL techniques offer immense potential for SLR, enabling the automatic extraction of meaningful representations from visual and temporal data. This section has provided a comprehensive overview of the key concepts, methodologies, and challenges associated with DL in the context of SLR. By leveraging DL, we can unlock new possibilities for accurate and real-time interpretation of SL gestures.

5 Conclusion

In this chapter, we have introduced sign languages and the importance of SLR systems in facilitating communication between deaf or hard-of-hearing individuals and the hearing world. We have also presented a review of the literature on SLR systems, including the techniques and methodologies that have been used to develop these systems.

From our review, it is clear that SLR is a challenging task that requires the use of sophisticated techniques such as ML, computer vision, and signal processing. While progress has been made in this area, there are still many open challenges that need to be addressed, such as improving the accuracy and robustness of recognition systems, developing systems that can recognize different sign languages, and addressing the issue of data sparsity.

In summary, the field of SLR is a promising area of research with many potential applications. We hope that this chapter has provided the reader with a good understanding of the current state of the art in SLR and the challenges that lie ahead.

Chapter 2

Related works

1 Introduction

1.1 Sign Language Recognition: An Overview

Sign language recognition refers to the process of interpreting and understanding sign language gestures to facilitate communication between individuals who are deaf or hard of hearing and those who are not proficient in sign language. Sign languages are complex visual languages with distinct grammar, syntax, and cultural variations. Recognizing and translating sign language gestures in real-time present significant challenges due to the intricacies and nuances involved.

The development of robust and accurate sign language recognition systems has gained increasing attention in recent years due to the potential impact on inclusive education, healthcare accessibility, smart home integration, and emergency communication. By bridging the communication gap between signers and non-signers, these systems contribute to creating an inclusive environment for individuals with hearing impairments.

1.2 Motivation for Exploring Related Works

Exploring related works in sign language recognition is essential for several reasons. Firstly, understanding the existing literature and research provides valuable insights into the state of the field, including advancements, methodologies, and limitations. This knowledge enables researchers to build upon previous findings and avoid duplicating efforts.

Secondly, analyzing related works helps identify gaps and challenges in current approaches to sign language recognition. By understanding the limitations of existing methods, researchers can propose innovative solutions to address these limitations and enhance the overall accuracy, speed, and usability of recognition systems.

Furthermore, advancements in sign language recognition technology have the potential to transform various sectors. In education, sign language recognition systems can assist in teaching sign language to non-signers, facilitating inclusive classrooms and promoting sign language literacy. In healthcare settings, these systems enable effective communication between healthcare providers and patients with hearing impairments, enhancing the quality of care and patient outcomes.

Moreover, integrating sign language recognition into smart homes and devices enhances accessibility and convenience for individuals who rely on sign language as their primary mode of communication. Additionally, in emergency situations where verbal communication may be challenging or impossible, sign language recognition systems can play a crucial role in ensuring effective communication and timely assistance.

1.3 Objectives of the Chapter

This chapter aims to provide a comprehensive overview of sign language recognition by exploring related works in the field. The specific objectives are as follows:

1.3.1 Review and Analyze Existing Methodologies

The chapter will examine various methodologies and techniques used in sign language recognition, including vision-based methods, data glove-based methods, and hybrid approaches. By analyzing these methodologies, their strengths, limitations, and performance, researchers can gain insights into the different approaches used in the field.

1.3.2 Explore Datasets for Training and Evaluation

The chapter will investigate existing sign language datasets utilized for training and evaluating sign language recognition systems. It will discuss the characteristics of representative datasets, such as size, diversity, annotation methods, and challenges associated with dataset collection and annotation.

1.3.3 Evaluate Performance Metrics

The chapter will delve into the evaluation metrics commonly used to assess the performance of sign language recognition systems. It will discuss metrics such as accuracy, precision, recall, and F1 score, highlighting their relevance and interpretation in the context of sign language recognition.

1.3.4 Identify Advancements and Future Directions

By analyzing related works, the chapter aims to identify advancements, trends, and potential research directions in sign language recognition. This includes identifying areas of improvement, emerging technologies, and challenges that require further exploration.

2 Datasets for Training and Evaluation

In the context of SLR, synthesis refers to the process of generating natural and expressive SL gestures from textual or spoken input. Synthesis plays a vital role in bridging the communication gap between SL users and non-signers by providing a means for converting spoken or written language into SL.

2.1 Text-to-Sign Synthesis Methods

Text-to-sign synthesis methods focus on converting written text into SL gestures. These methods involve linguistic analysis of the input text to determine the appropriate sign vocabulary and grammatical structure. Various techniques, such as rule-based systems, statistical models, and ML algorithms, have been employed in text-to-sign synthesis.

Rule-based systems utilize predefined rules and linguistic knowledge to map words or phrases to corresponding signs. These systems typically rely on manually crafted linguistic resources, such as sign dictionaries and grammatical rules, to generate SL sequences.

Statistical models, such as HMM and conditional random field (CRF), have been used to learn the statistical relationships between textual input and SL output. These models are trained on annotated corpora of text-sign pairs, allowing them to capture the patterns and dependencies between words and signs.

ML algorithms, including ANN and DL models, have shown promising results in text-to-sign synthesis. These models can learn the mapping between text and SL directly from data, leveraging large-scale annotated datasets to improve the quality and naturalness of generated SL gestures.

2.2 Speech-to-Sign Synthesis Methods

Speech-to-sign synthesis methods aim to convert spoken language into SL gestures. These methods involve audio analysis and processing techniques to extract relevant features from the input speech signals. The extracted features are then mapped to corresponding SL gestures using statistical models or ML algorithms.

Speech recognition algorithms, such as hidden HMMs and Deep Neural Network (DNN)s, are often used to transcribe spoken language into textual representations. The resulting text can then be processed using text-to-sign synthesis methods to generate SL gestures.

Multimodal approaches, combining audio and visual information, have also been explored in speech-to-sign synthesis. These approaches leverage audiovisual recordings of SL performances to learn the mapping between speech and SL gestures. By incorporating both acoustic and visual features, these methods can capture the nuances of spoken language and produce more accurate and natural SL outputs.

2.3 Challenges and Future Directions in Synthesis

Despite significant advancements in text-to-sign and speech-to-sign synthesis, several challenges persist in achieving high-quality and expressive SL generation. Some of these challenges include the variability in SL dialects, the lack of standardized linguistic resources, and the need for large-scale annotated datasets for training ML models.

Future research directions in SL synthesis include improving the naturalness and expressiveness of generated SL gestures, developing domain-specific synthesis models for different SLs, and integrating real-time feedback mechanisms to enable interactive and adaptive synthesis.

3 Evaluation Metrics and Performance Analysis

Evaluation metrics play a crucial role in measuring the performance of SLR systems. There are various evaluation metrics used for this purpose, including accuracy, precision, recall, F1-score, and confusion matrix.

Accuracy is the most commonly used metric that measures the percentage of correctly classified signs by the model. Precision measures how many of all predicted signs were correct while Recall measures how many actual signs were recognized correctly.

F1-score evaluates both precision and recall simultaneously to provide a more comprehensive understanding of model performance. It considers false positives and false negatives as well.

The Confusion Matrix is another effective way to evaluate SLR models. It provides information about true positive (correctly recognized) and false negative (incorrectly not-recognized) rates.

It's important to note that selecting appropriate evaluation metrics depends on the specific requirements of each application or dataset being used for training/testing purposes. Therefore choosing an optimal metric should be done with caution based on its suitability for application-specific goals

4 Methods of sign language recognition

SLR methods encompass a range of techniques and approaches used to interpret and understand SL gestures. These methods can be categorized into several broad categories, including vision-based methods, data glove-based methods, and hybrid approaches that combine multiple modalities for improved accuracy and robustness.

4.1 Vision-Based Sign Language Recognition Methods

Vision-based SLR methods rely on computer vision techniques to analyze and interpret SL gestures captured by video cameras or depth sensors. These methods typically involve extracting relevant features from the visual data and mapping them to corresponding signs in a predefined SL vocabulary.

Various techniques have been employed in vision-based approaches, including hand shape analysis, motion analysis, and spatiotemporal modeling. Hand shape analysis focuses on extracting information related to hand shape and configuration, such as fingertip positions, hand contours, and hand landmarks. Motion analysis techniques capture the dynamics of hand movements, including trajectory, speed, and acceleration. Spatiotemporal modeling methods aim to capture the spatial and temporal relationships between different hand movements and gestures.

4.2 Data Glove-Based Sign Language Recognition Methods

Data glove-based SLR methods involve the use of sensor-equipped gloves to capture and analyze hand movements during SL production. These gloves are equipped with sensors, such as flex sensors or Inertial Measurement Unit (IMU), which measure the bending angles of fingers or capture hand orientation and motion.

Flex sensor-based methods utilize the bending angles of individual fingers to recognize SL gestures. These sensors provide information about the flexion and extension of each finger, which can be used to distinguish different signs based on finger configurations.

IMU-based methods utilize inertial sensors, such as accelerometers and gyroscopes, to capture hand motion and orientation. These sensors provide data on the acceleration, angular velocity, and orientation of the hand, which can be used to infer SL gestures.

4.3 Hybrid Approaches in Sign Language Recognition

Hybrid approaches combine multiple modalities, such as vision and data glove sensors, to enhance the accuracy and robustness of SLR systems. By integrating visual information with data from sensors, these methods can capture both fine-grained hand movements and global hand positions, leading to more comprehensive representations of SL gestures.

Hybrid approaches often involve fusing the data from different modalities at different levels, such as feature fusion, decision-level fusion, or early fusion. Feature fusion combines the extracted features from vision and sensor data to create a unified feature representation. Decision-level fusion combines the decisions made by individual classifiers trained on different modalities to make a final decision. Early fusion combines the raw data from different modalities at the input level to create a joint representation for further processing.

5 Synthesis

Synthesis in SLR refers to the process of combining different approaches and techniques to achieve better results. With the increasing interest in DL, researchers have incorporated ANN models such as LSTM into SLR systems.

In addition to using LSTM, researchers also employ other methods such as HMM, Support Vector Machines (SVM), and CNN. By combining these various techniques, it is possible to improve accuracy rates and reduce error rates in sign language recognition.

Moreover, datasets play a crucial role in synthesis since they provide a means for training and testing models. Researchers use publicly available datasets or create their own annotated corpus of data for specific purposes.

Evaluation metrics are used to measure the performance of synthesized systems. Metrics like precision, recall and F1-score help researchers assess the effectiveness of their approach.

The development of synthesized systems has great potential for improving communication between hearing-impaired individuals and those who do not know SL. Future research studies on this topic will continue exploring new ways of integrating different techniques that enable accurate recognition algorithms with faster processing times than current state-of-the-art solutions can offer.

6 Conclusion

SLR is a complex task that involves the use of advanced algorithms and techniques in computer vision, ML, and DL. The development of SLR systems has made it possible for the deaf community to communicate more effectively with people who do not understand SL.

We have discussed the different methods used in SLR such as ANN and LSTM. We have also talked about some datasets commonly used in SLR research, as well as evaluation metrics that are frequently employed to assess the performance of these systems.

There have been significant advancements in DL-based approaches for SLR over recent years. However, much work still needs to be done to improve accuracy rates further. More comprehensive data sets need to be created with more diverse sets of gestures so that models can learn better.

All things considered; we believe that continued research into new algorithms and techniques will pave the way for even more sophisticated SLR systems capable of discerning subtle nuances between signs accurately.

Part II

Contribution

Chapter 1

Conception of Takalem gloves

1 Introduction

2 Hardware architecture and configuration

3 Dataset collection and preprocessing

4 Proposed deep learning architecture

5 Conclusion

Chapter 2

Results and discussion

1 Introduction

2 Evaluation criteria

3 Experiments

4 Discussion

5 Conclusion

References

[1] H. Abdelouafi and M. Omari, “Teaching sign language to the deaf children in adrar, algeria.”

[2] H. Abdelouafi, “Challenges of deaf education in algeria.”

[3] S. Lanesman, *Algerian Jewish Sign Language: its emergence and survival.*

[4] “Algerian sign language - wikipedia.”