Ramaiah Institute of Technology

(An Autonomous Institute, Affiliated to VTU)

MSR Nagar, MSRIT post, Bangalore - 54

A Dissertation PBL Report on

Sign Language Recognition

Submitted by

M Sneha	1MS14CS058
Mipsa Patel	1MS14CS148
Tilak S Naik	1MS14CS134
Vibha Karanth	1MS14CS136

Bachelor of Engineering in Computer Science & Engineering

Under the guidance of

Pramod Sunagar
Assistant Professor
Department of Computer Science
Ramaiah Institute of Technology

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

M. S. RAMAIAH INSTITUTE OF TECHNOLOGY
(Autonomous Institute, Affiliated to VTU)
BANGALORE - 560054
www.msrit.edu, 2017

DECLARATION

I Student of seventh semester BE, Dept of Computer Science and Engineering, Ramaiah Institute of Technology, Bangalore, hereby declare that the project entitled "Sign Language Recognition", thesis completed and written by me under the guidance of Pramod Sunagar, Dept of CSE, Bangalore for the partial fulfillment of the requirements for the award of the degree of Bachelor of Engineering.

Place: Bangalore
Date: 28$^{\text{th}}$ October, 2017

(1MS14CS058 M Sneha)
(1MS14CS148 Mipsa Patel)
(1MS14CS134 Tilak S Naik)
(1MS14CS136 Vibha Karanth)

ACKNOWLEDGEMENT

Abstract

Sign language is the language used by the hearing and speech impaired to communicate among themselves and others. Unless the concerned people know the sign language properly, there is a communication barrier between them. In most cases, an interpreter is required to carry out such a conversation. To reduce the dependency of the hearing and speech impaired people on interpreters, we develop a real time sign language detection system that converts the gestures from images and videos into English sentences, using various deep learning techniques.

Contents

List of Figures

1 Introduction

1.1 General Introduction

Sign Language is a vision based language that uses the movements of the hands and facial expressions for communication. Subtle differences among different hand gestures can have a huge impact on the meaning that they convey. Along with this, it is important to prevent the background noise of the image from affecting the gestures being communicated. We consider such problems and challenges and develop an effective sign language recognition system.

1.2 Statement of the Problem

There exist models that convert American Sign Language to sentences. However, most of these techniques are not based on deep learning. These models also cannot handle the nuances of the Indian Sign Language. We use the Indian sign Language and deep neural networks to develop a model that can convert gestures of ISL into English text.

1.3 Objectives of the project

- To develop an Indian Sign Language recognition system.

- To design a deep neural network which can give a better performance and robustness compared to the existing image processing and machine learning models.

1.4 Project deliverables

- An application that obtains frames from sign language communication videos, processes them, and obtains the meaning being conveyed by them in the form of English text.

- Comparison of performance of our model with existing algorithms used for sign language recognition.

1.5 Current Scope

Image and video processing is one of the important fields in Machine Learning and Deep Learning right now. The huge number of images and videos on social media and otherwise has led to an increased interest in the field. The algorithms and techniques developed as a result of this can be used in more impactful fields like helping the speech and hearing impaired lead more convenient lives.

The hand signs along with the emotions displayed by the individual provide the content and tone of the sentence. A tool to convert their words into a widely used language enhances their ability to communicate.

1.6 Future Scope

Since such applications will always be necessary, research and development of new algorithms to improve the efficiency of interpretation of sign language will be endless. Deep learning has a lot more to offer than what has already been explored. With such possibilities, there will always be scope to develop newer models that can be used in sign language recognition. The models used in this can also be used in other applications such as gesture controlled systems, tracking unusual activity, etc. with slight tweaks. With slight modification, it can also be used for recognition of other sign languages.

Moreover, the current work can be further extended to output speech instead of text. A similar generative model can be used to map speech or text to actions by stitching together pre-recorded videos or actually generate an interface with a skeletal structure. This model can be seamlessly integrated with various applications on different platforms to support high productivity.

2 Literature Survey

2.1 Introduction

There have been a lot of different approaches used to solve the problem of sign language recognition. However, most of these use image processing in MATLAB, upon which Machine Learning algorithms are applied. There has also been some work done using artificial neural networks.

2.2 Related work with the citation of references

[1] deals with the double handed Indian Sign Language. It is captured as a series of images, processed with the help of MATLAB and then converted to speech and text. However this approach uses image processing techniques which are highly sensitive to lighting conditions.

[2] proposes an image processing, computer vision and neural network based approach to identify the characteristics of the hand in images taken from a video through webcam. Identification of hand shapes from continuous frames is done by using series of image processing operations. Interpretation of signs and corresponding meaning is identified by using Haar Cascade Classifier. Finally displayed text is converted into speech using speech synthesizer.

[3] proposes a system that consists of three phases, a training phase, a testing phase and a recognition phase. In the training phase, each class is trained with a multiclass support vector machine (MSVM). Hu invariant moment and structural shape descriptors are combined to make a combinational feature vector that are to be extracted from the input image in the testing phase after applying

preprocessing. In the recognition phase, different classes are used for testing an input gesture.

2.3 Conclusion of Survey

Deep neural networks have revolutionized how different Machine Learning problems are approached. Using them to solve this problem can give significant results, and that is precisely what this project explores.

3 Project Management Plan

3.1 Schedule of the project

Figure 1 shows the project schedule.

4 Software Requirement Specification

4.1 Project Overview

Images of numbers and letters, and videos of words and sentences are fed as input to the model and interpreted text is received as output using various deep learning techniques.

4.2 Hardware

We use a system with NVIDIA® GeForce® 940MX (4 GB DDR3) GPU and 8 GB RAM to perform our experiment. A camera is required to record hand gestures and facial expressions.

4.3 Software Requirements

- Python

- OpenCV

- CUDA

- PyTorch for training and testing the deep neural network.

4.4 Functional Requirements

- Take image and video inputs of sign language conversation.

- Extract frames from videos for training.

- A training algorithm to reduce train-test error and generalize the model to make it invariant to lighting conditions.
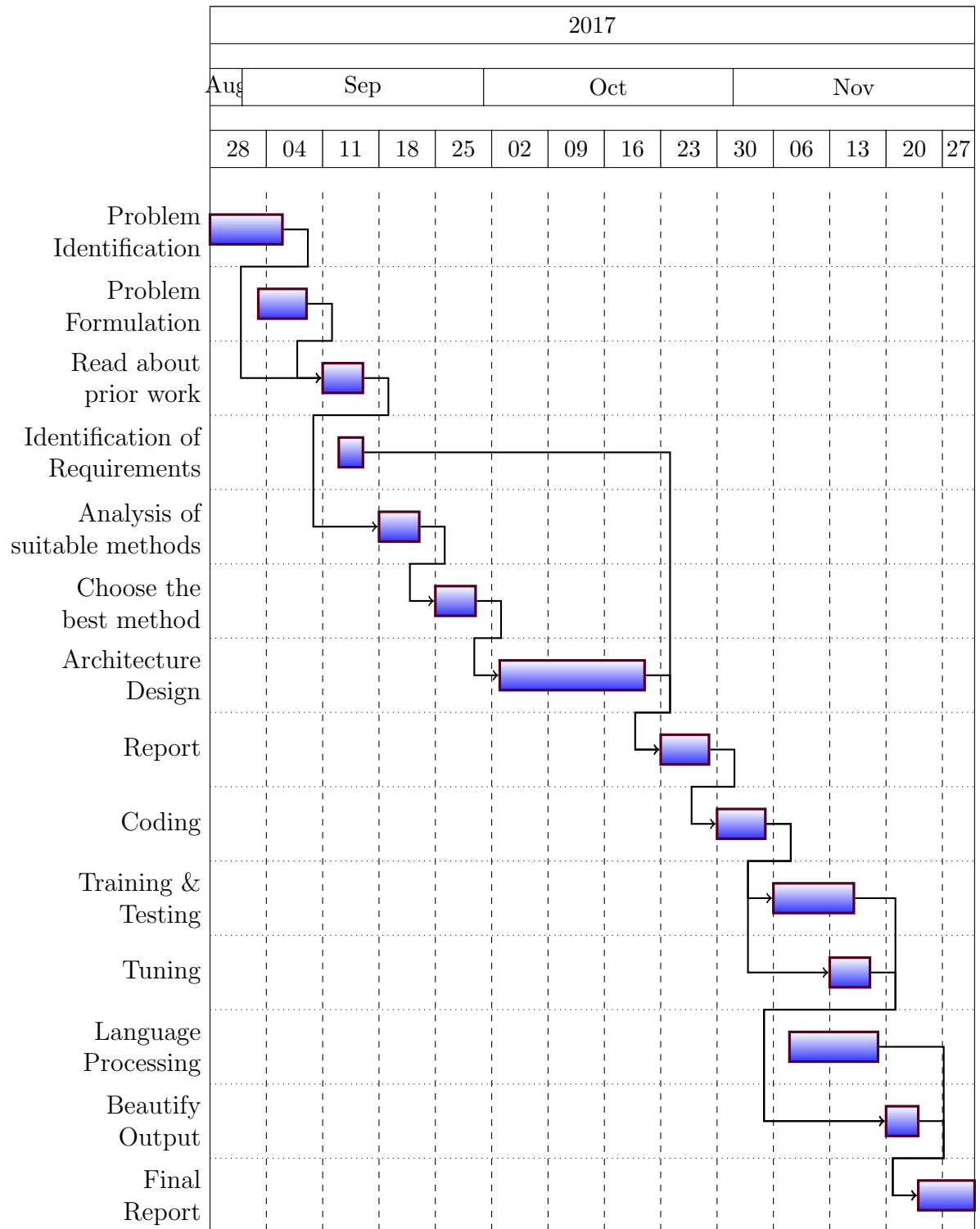
Figure 1: Project Timeline

- Convert the received output classes to letters, words or sentences.

5 Design

5.1 Introduction

This project has been designed such that it incorporates some of the major applications of machine learning and deep learning. Various concepts such as image and video processing, different neural network architectures, etc. are used. Simplicity and effectiveness are the key factors considered while designing and developing this application.
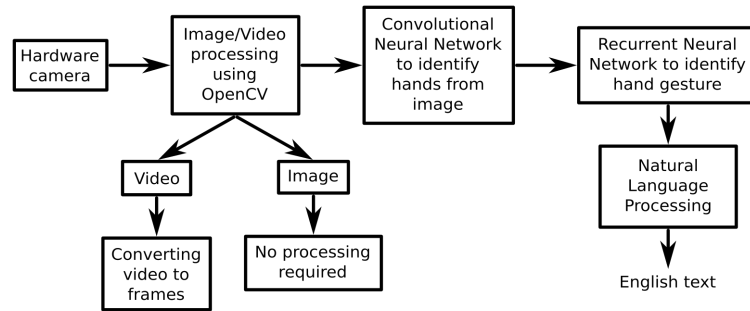
5.2 Architecture Design

Figure 2: Architecture Diagram

Figure 2 gives a brief overview of the architecture that the Sign Language Recognition application follows.

Input module
A camera is required to record the entire conversation. Letters and digits can be considered as single frames, whereas words and sentences are recorded as videos.

Image and Video Processing
Image and video processing is done using Python and OpenCV. Videos are converted into a series of frames to simplify the process of recognizing gestures.

Convolutional Neural Network
These networks are a class of deep feedforward neural networks that are based on weight sharing. They are commonly used in image and video processing to identify manifolds. In this project, they are used to identify hands from the given frames.

Recurrent Neural Network
Recurrent Neural Networks are used whenever memory of previous computations is required. It is used to identify the hand gesture by considering the gestures denoted by the current and the previous frames.

Natural Language Processing
Natural Language Processing is used to convert the output of the Recurrent Neural Network into proper English text.

5.3   Data Flow Diagram
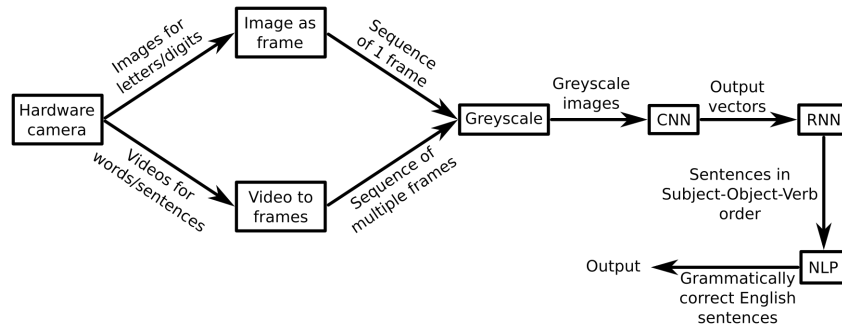

Figure 3: Data Flow Diagram

Figure 3 gives the interaction between the different modules of the application and how data flows between them.

Input
The input to the application is taken in the form of either images or videos. Images are considered as sequences of a single frame, whereas videos are converted into a series of frames using OpenCV.
These frames are converted into grayscale images to ensure uniformity in the training and test data in terms of the background, etc.

Convolutional Neural Network
This neural network is given grayscale images as input. It processes them and converts them to vectors that denote the positions of the hands in the image.

Recurrent Neural Network
This network takes in the input vectors and outputs letters/digits/words in the order of $subject - object - verb$.

Natural Language Processing module
NLP is required to convert this sentence in the form of $subject - object - verb$ into a proper English sentence.

6

5.4 Conclusion

It is a challenge to convert hand gestures to text due to wide variety of signs and their associated meanings. It is important that a sign language recognition system is as accurate as possible, due to the number of people that could use it and benefit from it. In such a case, even if such applications already exist, there is always a possibility of developing more effective or advanced systems using newer technology, and the work presented here tests one such approach.

References

[1] Kusumika Krori Dutta, Satheesh Kumar Raju K, Anil Kumar G S, and Sunny Arokia Swamy B. Double handed indian sign language to speech and text. In *Proceedings of the 2015 Third International Conference on Image Information Processing (ICIIP)*, ICIIP '15, pages 374–377, Washington, DC, USA, 2015. IEEE Computer Society.

[2] K. Dabre and S. Dholay. Machine learning model for sign language interpretation using webcam images. In *2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, pages 317–321, April 2014.

[3] K. Dixit and A. S. Jalal. Automatic indian sign language recognition system. In *2013 3rd IEEE International Advance Computing Conference (IACC)*, pages 883–887, Feb 2013.