Sign Language Recognition Using Convolutional Neural Networks

N.Priyadharsini
dept of Computer Science and Engineering
Sri Venkateswara College Of Engineering
chennai
pdpriyadharsini1@gmail.com

N.Rajeswari
dept of Computer Science and Engineering
Sri Venkateswara College Of Engineering
chennai
raji@svce.ac.in

Abstract— Abstract-Sign language is a lingua among the speech and the hearing impaired community. It is hard for most people who are not familiar with sign language to communicate without an interpreter. Sign language recognition appertains to track and recognize the meaningful emotion of human made with fingers, hands, head, arms, face etc. The technique that has been proposed in this work, transcribes the gestures from a sign language to a spoken language which is easily understood by the hearing. The gestures that have been translated include alphabets, words from static images. This becomes more important for the people who completely rely on the gestural sign language for communication tries to communicate with a person who does not understand the sign language. We aim at representing features which will be learned by a technique known as convolutional neural networks (CNN), contains four types of layers: convolution layers, pooling/subsampling layers, non-linear layers, and fully connected layers. The new representation is expected to capture various image features and complex non-linear feature interactions. A softmax layer will be used to recognize signs.

Keywords-*Convolutional Neural Networks, Softmax (key words)*

*****

I. INTRODUCTION

Introduction to sign language

SIGN language is the most expressive way for communication between hearing impaired people, where information are majorly conveyed through the hand/arm gestures(signing [1], [2],[3]) . SLR plays an important predominant role in developing the gesture-based human–computer interaction systems [1]–[5] .The recent years of statistics have witnessed an increased research interest in the interaction and intelligent computing. In the present day framework, computers have become a key element of our society for interaction. As the hand signs constitutes a powerful interaction human communication modality, they can be considered as an intuitive and convenient mode for the communication between normal human and human with hearing problems.

Introduction to Image Recognition

A sign may also be recognized by the environment as a compression technique for the information transmission subsequently reconstructed by the receiver. The signs are majorly divided into two classes: Static signs and Dynamic signs. The dynamic signs often include movement of body parts. It may also include emotions depending on the meaning that gesture conveys. Depending on the context, the gesture may be widely classified as:

- Arm gestures
- Facial / Head gestures
- Body gestures

Static signs includes only poses and configurations whereas dynamic gestures include strokes, postures and phases which is referred here.

Image Recognition System

The first step of sign language recognition system is to acquire the sign data. There can be various ways to get the data.

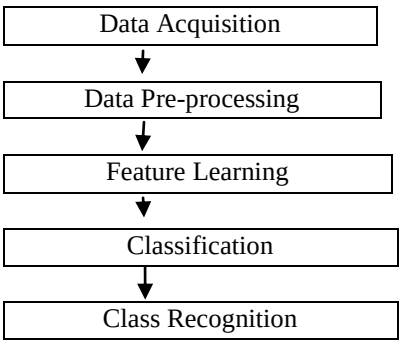

Fig 1 : Generalized block diagram of Image recognition system

Data Acquisition: The process of capturing the photographic images, such as of a physical scene.

Data Pre Processing: Goal of the pre-processing is an improvement of the input image data that suppresses unwanted noise or enhances image features important for further processing.

Feature Learning: The Feature learning starts from the initial measured data to builds its derived features intendeds to be informative and facilitating the subsequent learning

and generalization steps, and in some cases leading to better human interpretations.

Classification: Classification is the process related to categorization, the process in which ideas and objects are differentiated from others.

Recognition: Focuses on the recognition of patterns and regularities in data. Pattern recognition is the process of classifying input data into objects or classes based on key features.

Applications of Recognition System

Sign language recognition has a broad range of applications[6] such as the following:

- developing aids for hearing impaired people;
- Loud venue;
- Class rooms;
- Recording studio;
- Essential when person wants to communicate after his/her throat injury;
- communicating in the video conferencing;

II. SIGN LANGUAGE

A Sign Language (SL) is the natural way of communication of deaf community. About 360 million people worldwide having hearing problem (about 5%). In recent survey of Indian Pediatrics 4 out of 1000 children born in India have hearing problem. The well-known sign languages are namely [8]:

- American Sign Language (ASL)
- Israeli Sign Language
- Indian Sign Language (ISL)
- Pakistani Sign language
- South Korean Sign Language
- Taiwan Sign Language
- Arabic Sign Language and so on

More than 1 million of adults and 0.5 million children in India make use of Indian Sign language[7].

Image Expression Database

Dataset is made of 26 alphabets along with two words of our formal language with an average length. The system was designed to recognize 26 sign language hand signs of 10 for each alphabet and words.280 images are used as data set. The dataset constitutes of 252 images for the training segment and 28 images for testing. With the 252 training data, which is again splitted into 201 for training and 51 for validating the training.

III. FEATURES FOR SIGN LANGUAGE

A. Recognition

An important problem while designing the sign language recognition system is the extraction of suitable distinct features that efficiently characterize the variations in signs. Since the pattern recognition techniques are rarely independent of the problem domain, it is believed that a proper choosing of the features significantly affects the performance of classification.Three issues must be considered in feature extraction. The first issue is analyzing the region used for feature extraction. While some authors follow the ordinary framework of dividing the images into small intervals, called pixels, from each which a local feature vector is extracted, other researchers prefer to extract global statics from the whole speech utterance. Another important question is what are the best feature types for this task. Finally, what is the effect of ordinary image processing such as noise removal on the overall performance of the classifier?

B. Sign Language Recognition Using CNN

Four layers are there in a CNN, for the classification problems. The layers are: convolution layers, pooling/sub-sampling layers, non-linear layers, and fully connected layers.

Convolutional Layer

The convolution is a special operation that extracts different features of the input. The first it extracts low-level features like edges and corners. Then higher-level layers extract higher-level features. For the process of 3D convolution in CNNs. The input is of size N x N x D and is convolved with the H kernels, each of them sized to k x k x D separately. Convolution of one input with one kernel produces one output feature, and with H kernels independently produces H features respectively. Starts from top-left corner of the input, each kernel is moved from left to right. Once the top-right corner reached, kernel is moved one element downward, and once again the kernel is moved from left to right, one element at a time. Process is done continuously until the kernel reaches the bottom-right corner.

Pooling/Sub-Sampling Layers

The pooling layer reduces the resolution of features. The features are robust against noise and distortion. There exists two ways to do pooling: 1.max pooling and 2.average pooling. For both cases, the input is divided into non-overlapping dimensional spaces. For average pooling, the average of the given values in the region are calculated. For max pooling, the maximum value of the given values is selected.

Non-Linear Layers

CNNs in particular rely on a non-linear “trigger” function to signal distinct finding of likely features on each

 ISSN: 2321-8169
625 – 628

and every hidden layers. CNNs may use the variety of specific functions like ReLUs (rectified linear units) and continuous trigger functions to efficiently implement this non-linear triggering.

Fully Connected Layer

Fully connected layers are always used as the final layer. These layers are mathematically sums the weigh of previous layer of features, indicating the precise mix of "ingredients" to determined specific target result. All the elements of all the features of the previous layer get used in the mathematical calculation of each element of the each output feature.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

C. Training Phase

In training sequence, the workflow of proposed system is as follows:

- The reference image is extracted and the individual image is pre-processed. In this preprocessing, filters are applied so as to enhance the useful content of the frame information and to reduce the unwanted information as much as possible.
- All the features of processed image are then extracted using CNN and stored in database associated with that sign.
- The same procedure is followed for all the sign to be included in the system. And the complete reference database is prepared.

.

D. Test Phase

The workflow of testing sequence can be outlined as below:

- The input image is extracted.
- The images are pre-processed in order to get the Processed image as similar to that in training sequence.
- After obtaining the Processed image in testing, the images are matched with the previously maintained database.
- The difference between them is measured depending feature. On finding the nearest\ match, the image is recognized as that particular sign and corresponding output is flashed on the screen.

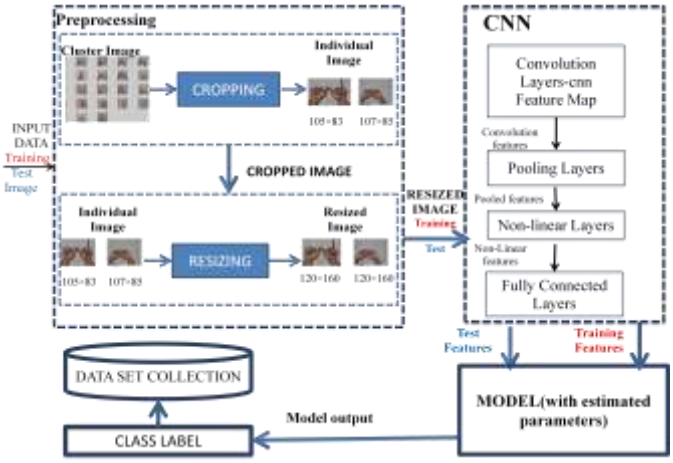

Fig 2: Sign Language Recognition Using CNN outline

IV. IMPLEMENTATION

Image Acquisition Model

Image acquisition is the process of creating the photographic images, such as the interior structure of an object. The term is often assumed to include the compression, storage, printing, and display of such images.

Preprocessing Model

Main aim of pre-processing is an improvement of the image data that reduce unwanted deviation or enhances image features for further processing. Preprocessing is also referred as an attempt to capture the important pattern which express the uniqueness in data without noise or unwanted data which includes cropping, resizing and gray scaling.

Cropping

Cropping refers to the removal of the unwanted parts of an image to improve framing, accentuate subject matter or change aspect ratio.

Resizing

Images are resized to suit the space allocated or available. Resizing image are tips for keeping quality of original image. Changing the physical size affects the physical size but not the resolution.

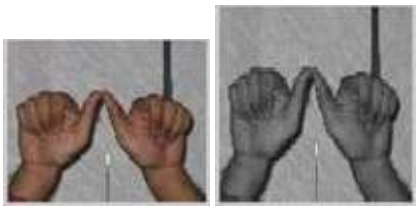

Fig 3: Output of Pre-processing

Feature Learning

Comprised of one or more convolutional layers and followed by one or more fully connected layers as in a standard multilayer neural network.It implicitly extract relevant features.A Feed-forward network that can extract topological properties from an image.Like almost every other

neural networks CNNs are trained with a version ofthe back-propagation algorithm.

Convolutional layer: Core building block of a CNN. Layer's parameters consist of a set of learnable filters (or kernels). Filter is convolved across the width and height of the input volume. Computing the dot product between the entries of the filter

Pooling Layer: Reduce the spatial size of the representation to reduce the amount of parameters. Independently operates on every depth slice of the input. Most common form is a pooling layer with filters of size 2x2 applied with a stride of 2 down samples every depth slice in the input by 2 along with both the width and the height, discarding 75% of the activations spatially, using the MAX operation

ReLU layer: ReLU is the abbreviation of Rectified Linear Units which increases the nonlinear properties

Fully connected layer: Neurons in a fully connected layer have full connections to all activations in the previous layer**.** The activations are computed with the matrix multiplicaton.

V. RESULTS

The system was tested on 10 different images for each sign. The accuracy was checked as per correctness of every gesture made i.e. Alphabets and Words. Fig.4 shows the accuracy of SL recognition system performance in case of each and every alphabetic sign. The maximum accuracy is 100% for all the alphabets. This implies that the system works efficiently for most of the alphabetic character recognitions.

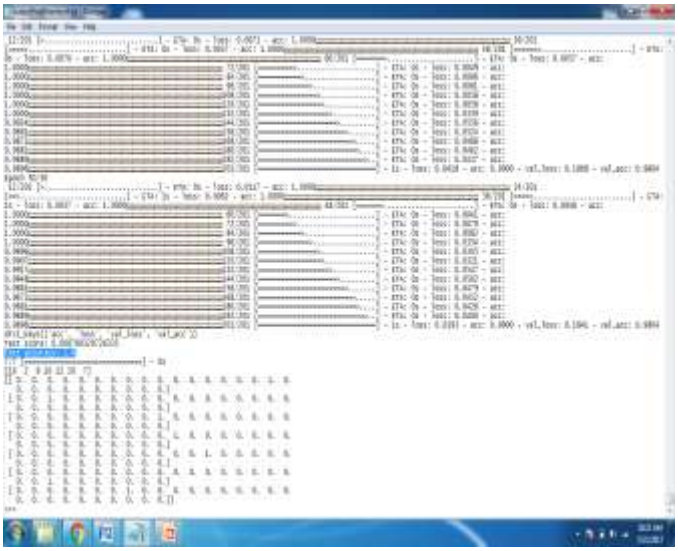

Fig 4: Accuracy based on Individual Alphabet

VI. CONCLUSION

This proposal will help to achieve high performance in recognizing the sign language, which is the main communication bridge between the deaf and dumb people and the normal people. It is hard for most of the people who is not familiar with the sign language to communicate without an interpreter. In this proposal, we have created an idea of translating the static image of sign language to the spoken language of hearing. The static image includes alphabet and some words, used in both training and testing of data. Feature representation will be learned by a technique known as convolutional neural networks. The new representation is expected to capture various image features and complex non-linear feature interactions. A softmax layer will be used to recognize signs.

REFERENCES

[1] S. C. W. Ong and S. Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 873–891, Jun. 2005.

[2] L. Ding and A. M. Martinez, "Modelling and recognition of the linguistic components in American sign language," *Image Vis. Comput.*, vol. 27, no. 12, pp. 1826–1844, Nov. 2009.

[3] D. Kelly, R. Delannoy, J. Mc Donald, and C. Markham, "A framework for continuous multimodal sign language recognition," in *Proc. Int. Conf. Multimodal Interfaces*, Cambridge, MA, 2009, pp. 351–358.

[4] G. Fang, W. Gao, and D. Zhao, "Large vocabulary sign language recognition based on fuzzy decision trees," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 34, no. 3, pp. 305–314, May 2004.

[5] Y. Li, X. Chen, J. Tian, X. Zhang, K. Wang, and J. Yang, "Automatic recognition of sign language subwords based on portable accelerometer and EMG sensors," in *Proc. Int. Conf. Multimodal Interfaces—Workshop Mach. Learn. Multimodal Interaction*, Beijing, China, 2010, pp. 1–7.

[6] C. L. Lisetti and D. J. Schiano, "Automatic classification of single facial images," Pragmatics Cogn., vol. 8, pp. 185–235, 2000.

[7] N. Purva, K. Vaishali, *"Indian Sign language Recognition: A Review",* IEEE proceedings on International Conference on Electronics and Communication Systems, pp. 452-456, 2014.

[8] F. Pravin, D. Rajiv, *""HASTA MUDRA" An Interpretation of Indian Sign Hand Gestures",* 3rd International conference on Electronics Computer technology, vol. 2, pp.377-380, 2011