Sign Language Recognition System

A Project Report submitted by

Pranali Kanampalliwar

in partial fulfillment of the requirements for the award of the degree of

MTech in Data and Computational Science

॥ त्वं ज्ञानमयो विज्ञानमयोऽसि ॥

Indian Institute of Technology Jodhpur
School of Artificial Intelligence and Data Science

January 2022

Declaration

I hereby declare that the work presented in this Project Report titled Sign Language Recognition System submitted to the Indian Institute of Technology Jodhpur in partial fulfilment of the requirements for the award of the degree of Masters in Technologies - Data and Computational Science, is a bonafide record of the research work carried out under the supervision of Dr Sumit Kalra The contents of this Project Report in parts, have not been submitted to, and will not be submitted by me to, any other Institute or University in India or abroad for the award of any degree or diploma.

Pranali Kanampalliwar
M21AI566

Certificate

This is to certify that the Project Report titled Sign Language Recognition System, submitted by Pranali Kanampalliwar(M21AI566) to the Indian Institute of Technology Jodhpur for the award of the degree of Masters in Technologies - Data and Computational Science,is a bonafide record of the research work done by her under my supervision. To the best of my knowledge,the contents of this report, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr Sumit Kalra

Abstract

Sign language recognition system or gesture recognition system converts the sign gesture into text caption or audio signals. This facilitates for an easier communication between people with hearing listening problem with normal people. This model can be a great social impact but still not used in real time due major variation in sign languages across different region and complex nature. Existing system focuses on sign gestures performed for letters and number and building classification model to for those gestures. However, these models are not able to design reliable features and for variety of sign gestures they are not able to adapt as expected. The proposed system is for Sign gestures recognition system is based on neural network processing which uses both three-dimensional convolutional neural network and long short memory techniques for analysis. This method will identify sign gestures performed signer includes plurality of frames associated with these gestures to first three-dimensional convolution neural networks and second Convolution neural network. These frames are taken from video datasets in which signer has performed the different gestures. The first 3DCNN produces motion information whereas the other one produces pose and color information. The first network will be implementing optical flow algorithm to detect gestures and then fusing with motion information. Other network will be using pose and color information for identifying the performed gestures and determining whether the identification corresponds to singular gestures across plurality of frames using recurrent neural network that comprises one or more long short-term memory unit.

Contents

1. Introduction

Communicating through gestures is most natural way of expressing ourselves. People with hearing or listening problems which we refer as deaf and dumb people mostly communicate via these gestures. Around 400 million people suffers from this listening and speaking problems, some are suffering since they have born, some are accidently become impaired while rest are facing these problems at their old ages. Communication via gestures can also been seen other living beings like birds or animals as they don’t speak like humans. Sign languages are being used since many years by these people for their easy communication and help them to talk to others. There are various sign languages being used worldwide. These languages differ the way they are communicated, their letters, numbers, words, etc. Some of the most famous sign languages are American Sign Language, British Sign Language, Chinese Sign Language, Korean Sign Language, French Sign Language, German Sign Language, Portuguese Sign language, Egyptian Sign language, Arabic Sign language and many more. Every sign language has its own grammar and set vocabulary. But only problem is these sign languages can only be understood by people having good knowledge of sign languages or have learnt sign languages. Hence an impaired person can only communicate with person who is expert in this and cannot talk to normal people. Since last few years researchers have shown interest in making sign language system. Due emerging advance technologies like Artificial intelligence, Human computer Interaction, Computer Vision, etc., many techniques have been tried and tested. The idea of these techniques is mostly to track the hand movements and then translating the gestures performed. Different SLR techniques have their own journey over the period. Initially sensors have been used to track these movements which were good start, different gloves techniques were tried to track the gesture and performance of these wearable techniques been also evaluated. Besides, there are various handcrafted feature models and deep learning-based models being proposed. Among all the techniques, deep learning-based models have shown better results in recognition and classification problem more accurately, yet these are not implemented in real time situation due to limitations and still been researched. Proposed deep learning techniques includes Deep Belief Networks, Connectionist Temporal Classification, Support vector machine, Principal Component analysis, Hidden Markov, Convolution Neural Networks, Recurrent Neural Network, etc.

2. Problem Statements

A significant problem or barrier of in-effective communication between, impaired people, who use only sign language and other community member have no prior understanding of sign language. These communication barriers occur in public services worker who are not familiar with sign languages. This also limits speech impaired person working at different places communicating with people who are unfamiliar with sign languages. Efforts have been made to solve these challenges like interpreters, but these are not very efficient. Advance technology has been proposed like machine learning techniques. This uses special algorithms that translate sign gestures to text or audio output that are universally understandable. Hence research is still in progress, as it has not been established which model can solve problems and used in real time. There is need of an efficient sign language recognition system. The objective SLR system, it is highly independent of the input data type. The input to system should be of any type, either it is video or live communication through camera. The signer must follow no special instructions like, distance from camera, be in a certain frame, with background, any wearable devices while performing gestures, particular camera features, any specific types of cloths, limited to specific sign language, i.e., they are free to perform any sign gestures, etc. For real time conversation the SLR system will be able to communicate with good response time.

3. Literature Survey

In recent years many significant changes have been found in the field of sign language recognition field. Mostly sign languages recognition techniques can either be done using sensors and tracking the movement of the signer or tracking the movement by using camera and using emerging computer vision techniques. Sensor based technique uses motion sensors and electronic gloves. Initially SL approached using two bend sensors having push button for tracking motion which failed to track signs for signer having small hands. Both sensor based and hand gloves-based system dataset has been evaluated and gloves system had better accuracy than sensor-based system. Also, data fusion had higher accuracy at classification level as compared to data fusion at feature level. But the sensor-based approach was good in tracking of hands and handling environmental constraints, but this was compulsory to wear gloves during the signing but was difficult for real-time system. This sensor based could not capture the non- manual gestures that are basic component of sign language.

Table. Studying different approaches being performed by researchers

Title of the Study	Sign Language Used	Features and Classifiers
Sign Language Fingerspelling Recognition Using Depth Information and Deep Belief Networks	American Sign Language (ASL)	Deep-Belief Network (DBN)
DeepASL: Enabling Ubiquitous and Non-Intrusive Word and Sentence-Level Sign Language Translation	American Sign Language (ASL)	Hierarchical bidirectional deep recurrent neural network (HB- RNN) + Connectionist Temporal Classification (CTC)
Deep learning-based Sign language recognition system for static signs	Indian Sign Language (ISL)	Convolutional Neural Networks (CNN)
Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM	Arabic Sign Language (ArSL)	Convolutional Neural Networks (CNN) + Bidirectional LSTM
MyoSign: enabling end-to-end sign language recognition with wearables	American Sign Language (ASL)	Convolutional Neural Networks (CNN) + Bidirectional Long Short Term Memory Layers (BLSTM)
SonicASL: An Acoustic-based Sign Language Gesture Recognizer Using Earphones	American Sign Language (ASL)	SubNet + Connectionist Temporal Classification (CTC)
DeSIRe: Deep Signer-Invariant Representations for Sign Language Recognition	Portuguese Sign Language (PSL) + American Sign Language (ASL)	Convolutional Neural Networks (CNN)
Exploration of cChinese sign language recognition using wearable sensors based on deep belief net	Chinese Sign Language (CSL)	Deep-Belief Net (DBN)
Benchmarking deep neural network approaches for Indian Sign Language recognition	Indian Sign Language (ISL)	Deep Convolutional Neural Network (DCNN)
Connectionist Temporal Fusion for Sign Language Translation	German Sign Language (DGS) + Chinese Sign Language (CSL)	Connectionist-Temporal Classification (CTC)
SubUNets: End-to-End Hand Shape and Continuous Sign Language Recognition	German Sign Language (DGS)	Convolutional Neural Networks (CNN) + CTC
Two Dimensional Convolutional Neural Network Approach for Real-Time Bangla Sign Language Characters Recognition and Translation	Bangla Sign Language (BdSL)	Convolutional Neural Network (CNN)

Vision-based hand gesture recognition for indian sign language using convolution neural network	Indian Sign Language (ISL)	Convolutional Neural Network (CNN)
Deep Learning-Based Sign Language Digits Recognition	American Sign Language (ASL)	Convolutional Neural Network (CNN)
Neural sign language translation by learning tokenization	German Sign Language (DGS)	3-dimensional convolutional neural network (3DCNN) + 2- dimensional convolutional neural network (2DCNN)

3.1. RGB based Sign language recognition systems

Hence recent sign language recognition is a computer vision-based approach. The vision based SLR system has single or multiple cameras for sign capturing. New cameras can provide in-depth information which helps us to find more information about performed signs. While most of the research focusses on hand gestures, there is also some research focusing on linguistic information conveyed by the face and head of person performing gestures. As cost-effective consumer depth cameras have become available in recent years, it has become practical to capture RGB videos and depth maps as well as track a set of skeleton joints in real time. If we compare 2D RGB images and RGB-D images, RGB-D images provide both photometric and geometric information. Therefore, recent research work has been motivated to investigate SL recognition using both RGB and depth information.

Some early work based on RGB-D cameras only focused on very small numbers of signs from static images. While this system only used static RGB and depth images, some studies employed RGB-D videos for ASL recognition. A hidden Markov model was developed to recognize 19 ASL signs collected by Kinect and the performance was compared colored-glove and accelerometer sensors. For this same Kinect data, they also compared signer to be standing and sitting and concluded that they get higher accuracy when the users were standing. Later a hierarchical conditional field method was developed to recognize 24 manual signs (seven one handed and seventeen two handed) from handshape and motion in RGB-D videos. For ISLR system, support Vector machine (SVM) classifier recognized 37 signs from Indian Sign Language based on 3D skeleton points, 34 signs of Chinese sign language based on color images and skeleton joints and 34 signs of Brazilian Sign Language using handshape, movement and position captured by Kinect camera.

SL consists of hand gestures, facial expression, and body poses. However most existing work has focused only on hand gestures without combining facial expressions and body poses. While a few attempted to combine hand and face, they only used RGB videos. Hence there is need of model that combines multi- channel RGB-D videos (RGB and depth) with fusion of multi-modality features (hand, face, and body).

3.2. Convolution Neural Network based SLR

Different areas of Computer vision like image classification, description of image, object detection and many more have been experimented with deep neural networks. Many efforts have been made to extend CNNs from image to video domain, which is more challenging since video data is much larger than images. Hence it is very difficult to handle video data in a limited memory space. To extend image-based CNN structures to the video domain we need to perform classification and fine-tuning on each independent frame, and then later conduct fusion like average scoring for predicting the action class of the video. A Two stream framework was introduced for incorporating temporal information from the videos. One stream was based on RGB images, on the other hand stacked optical flows. It has a very innovative process to learn this temporal information from CNN structure, this was still image-based, after the first convolution layer the third dimension of stacked optical flows immediately collapsed. A model with sequential information of extracted features from different segments of videos were given as input to Recurrent Neural Network (RNN) structures, showed good results for action recognition. They emphasized pooling strategies and how different features to be fused, while the rest focused on training DNN structures from end-to-end which also integrates CNNs and RNNs. These networks mainly use CNN to extract spatial features, then RNN is applied to extract the temporal information of the spatial features.

3.3. Three-Dimensional Convolutional Neural Network based SLR

3DCNN approach was recently proposed to learn the spatial-temporal features from 3D convolution operations and has been substantially completed in video evaluation duties alongside element video caption and movement detection. Mostly 3DCNN is trained with fixed-length clips and later fusion is performed to obtain the final category of the entire video. A 3D-ResNet was proposed by replacing all the 2D kernels in 2D-ResNet with 3D convolution operations, with its advantage of adverting gradient vanishing and explosion,3D-ResNet outperforms many complex networks. 3DCNN has proposed for ASL recognition, it shares properties with video action recognition, therefore, many networks for video action have been applied to this task. Temporal residual networks were proposed for gesture and sign language recognition and a temporal convolution on top on top of the features extracted by 2DCNN for gesture recognition, later a Hierarchical Attention Network with Latent Space (LS-HAN) which eliminates the pre-processing of the temporal segmentation. *Pu* proposed to employ 3D residual convolutional network (3D- ResNet) to extract visual features which are then fed to a stacked dilated convolution network with connectionist temporal classification to map the visible capabilities into textual context sentence. Camgoz attempted to generate spoken language translations from sign language video, and proposed SubUNets for simultaneous hand shape and continuous sign language recognition.

A weakly supervised framework to teach the network from videos with ordered gloss labels but no actual temporal locations for non-prevent sign language recognition. In prior work, our research team proposed a 3D-FCRNN for ASL recognition by combining the 3DCNN and a fully connected RNN. Below diagram shows the entire flow and development for 3DCNN ASLR system.

3.4. Recent American SLR using 3DCNN

The multi-channel multi-modality 3DCNN framework for ASL recognition. The multiple channel channels encompass RGB, Depth, and Optical float on same time because of the reality the multiple modalities embody hand gestures, facial expression, and body poses. While the vast picture is used to symbolize frame pose, to higher version hand gesture and facial features change, the areas of arms and face is received from RGB picture primarily based totally at area guided with aid of skeleton

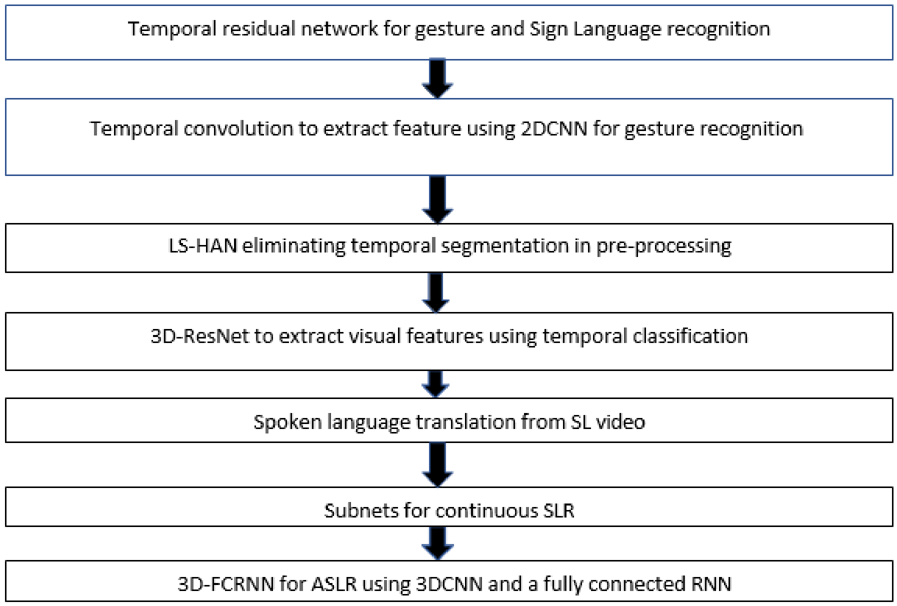

Fig. ASLR proposed work in recent years

joints. The entire framework includes important components: proxy video technology and 3DCNN whole framework consists of two main components: proxy video generation and 3DCNN modeling. Firstly, proxy videos are generated for every ASL sign via way of means of choosing a subnet of frames, spanning the complete video clip of every ASL sign, to symbolize the general temporal dynamics. Then the generated proxy frames of depth, optical flow, RGB hands and RGB face are fed into the multi-circulated 3DCNN component. The predictions of these networks are weighted to benefit the final effect of ASL recognition.

ASL experts to accumulate an ASL dataset of hundred manual signs including both hand gestures and facial expressions with complete annotation at phrase label and temporal boundaries for beginning and finishing points. By fusing multiple channels with this framework, the accuracy of recognizing ASL signs can be improved. This technology for identifying the appearance of precise ASL terms has valuable applications for generations that could benefit people.

4. Limitations to existing research

While attempting to translate sign gestures, researchers have tried and tested many techniques. There are multiple sign languages being used in different parts of the world. All sign languages have their own sets of letters and hence different gestures are used to perform those letters. Every sign language has their own vocabulary and grammar which are different from other languages. Thus, researchers have mostly focused on their regional languages and proposed solution according to them. There either a need of a universal sign language, which is ideally not possible, or a system is required to able to understand any sign language and give output irrespective of sign language being performed which are actually very complex task. While training the model datasets like video or image are given, which are highly dependable on the background environment or hardware requirement. Hence after training, when models are tested in real time it doesn't efficiently and accurately performs. Different models have different drawbacks like, Device based models suffers high cost, PCA based models are dependent on background and lighting, Fuzzy & Markov's models suffers with high computations, Hidden Markov's model worked only at word level but failed in sentence level, Support vector machines failed to differentiate between similar gestures or that have similar tensions on same finger, etc.

5. Methodology

5.1. Overview of existing Sign Language Recognition Techniques

There are many SLR techniques been researched and have given prominent results but still faces many challenges in many areas. Most of the work have been done in isolated sign language recognition. We can find very limited researcher who have focused on continuous sign language recognition due to its complex nature. Sign language recognition research over the years can be classified in broadly two categories sign language appearance-based approaches and pose based approaches. In appearance-based approaches will extract representations which are unitary in nature from input frames which are later used for recognizing purpose. Initially appearance-based approaches had used shallow statistical modeling and handicraft features data. Whereas in pose-based approaches signer pose representation are given as input like pose representations of body, face, and hand landmarks estimations. If we try to focus on continuous SLR like using Principal component analysis techniques, which are fast and efficient. These takes 3 frames per second from video and static gestures are analyzed. These techniques were highly dependent on background lighting condition. Later Fuzzy logic and Markov method came into picture. In this meaningless gesture motion such as preparatory motion and useless movement between sign words were rejected using state automata and fuzzy partitioning. For hand motion they concentrated mostly on speed and velocity. They do sentence-based recognition without any pause between two sign words, but this system faced high computational burden. In device-based approaches gloves-based approaches are research most. For this a large vocabulary of language interpreter was created. Here the problem of key frame extraction was using time-varying parameter detection was solved. In these approach discontinuities in frames. Four features position, posture, orientation, and motion were analyzed statistically. For gestures recognition HMM model was used. But the only limitation for this was that it is person dependent and using gloves is expensive and it needed physical connection between user and computer. Histogram is proven to be good solution for drawbacks like person dependency, orientation, position dependency, scalability. Our focus is to resolve major drawbacks from previous research and make use of techniques with better performance. Major drawbacks like key frame extraction, dependency on orientation and real time recognition.

5.2. Motivation for 3D CNN technique

Deep Neural Networks has been utilized to solve the complexity problems for speech recognition, computer vision, action recognition and natural language processing. RNN and CNN are special scaled editions of DLN which have attracted researcher's attention. RNN is DLN sequence based for processing temporal sequences i.e., natural language. It acts as internal memory for learning temporal dynamics in sequential type of data. For continuous gestures RNN outperforms non-recurrent models predicting both ends of gestures in a sequence having higher accuracy and learning motion sequence hierarchy. A 3DCNN has capability to extract arranged spatial-temporal features from multiple frames in absence of temporal models. Hence temporal CNN with bidirectional LTSM and detection network through three stage optimization process has affected performance positively by tuning feature extraction and sequence learning components. This method avoids overfitting problem as well. 3DCNN has correctly classified gesture from multi-model frames without any sequence being spotted. This can also be applicable for on-line recognition.

5.3. *Proposed Solution*

5.3.1. *Architecture*

Below diagram shows the real time gesture recognition system, it includes model creation and recognition. The video clips are used for training and may be generated by training framework or provided by user. The training video clips covers the gesture to be recognized from multiple distances and angles. Having a diverse set of visual characteristics in training video clips will enable high accuracy recognition. Each frame of the video is processed and pose estimation is applied to pixels for the body, fingers, and face. This will result in training video clips with overlaid pose estimation pixels. Optical flow is extracted from frames with overlaid pose estimation pixels. After feature extraction line, corner, shape, and edge rendering is performed to allow borders of the shapes in training media to be accurate and enable differentiation of one part from another. This results in very precise feature identification, advantageously enabling far more accurate recognition of movement or flow of objects that occur across time. The extracted and processed features are provided for training a 3DCNN model, as the first bit stream and second bi t stream for a second 3DCNN includes spatial and color information i.e., RGB information. The output layers of 3DCNN s are then fused, thereby enabling the convolution to run across both 3DCNNs, so flow and RGB or spatial information can be processed together as part of the same convolutional kernel.

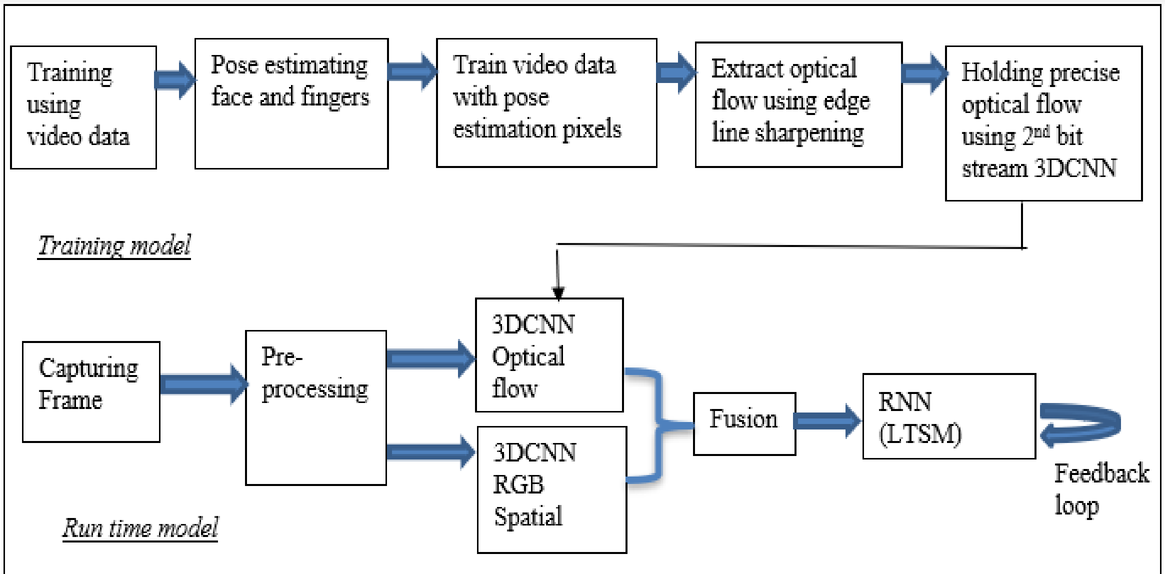

Fig. Block diagram of SLR framework

Similarly, recognition process starts from capturing frames from devices with multiple apertures or webcam and other sensors. The frame captured is implemented using its own thread and another different thread is used for recognition system that are ready to accept a frame. The captured frame is preprocessed with pose estimation for the body, face and fingers and resulting pose estimations laid on top existing frame pixels using a transparent layer. The resulting frames are provided to recognition process, and both the 3DCNN begins the recognition process i.e., one from motion or optical flow perspective and the other from a RGB or spatial information perspective. The two 3DCNN are fused together to enable their output layers to be processed jointly and using both their data streams. The recognition results for each frame pixels are provided to RNN which uses Long Short- term Memory i.e., LTSM to track the recognition process temporally. The RNN with LTSM uses in own feedback loop to track state across more than a single round of recognition.

5.3.2. *Facial Feature Extraction Process*

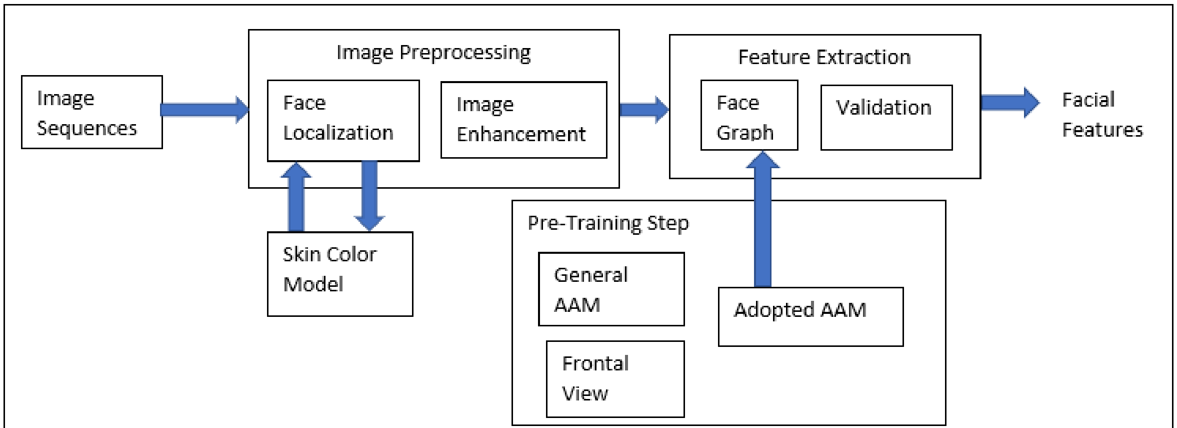

Fig. Processing chain for facial extraction

Facial feature extraction is major part of proposed SLR system. As shown in below figure the process can be divided into an image preprocessing stage and feature extraction stage. Since the input image sequences mostly covers the signing space entirely, the signer face region should be initialized in each image. After this region is cropped and upscaled for further processing. To localize areas of interest like eyes, mouth a face graph is matched iteratively to facial region using user adapted active appearance model. Later numerical descriptions of facial expressions, head pose, line of sight and lip outline, etc. are all computed. For every image sequence, all extracted features are merged into feature vector and later used for classification.

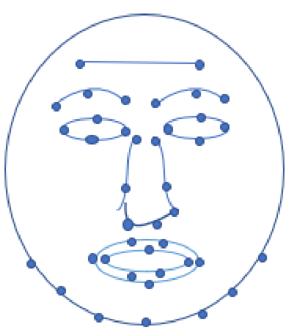

Fig. Face graph with landmarks points

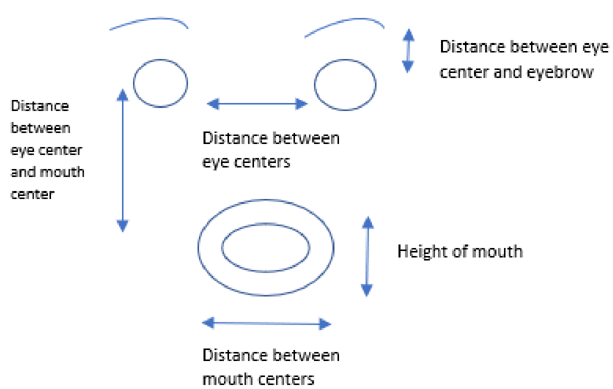

Fig. Facial parameters representations

5.3.3. *Pose feature extraction*

Pose detection that are carried out as a part of filtering operations. An input frame includes the subject with background and has been processed using pose detection algorithm that superimposes a skeleton on the subject. A subsequent step uses the skeleton to eliminate the background information that is not relevant for generating the output frame. If the subject is referencing any external object, the pose detection algorithm can recognize that skeleton of the subject, for example, pointing to an object that can include the external object in the output frame for processing by the neural network.

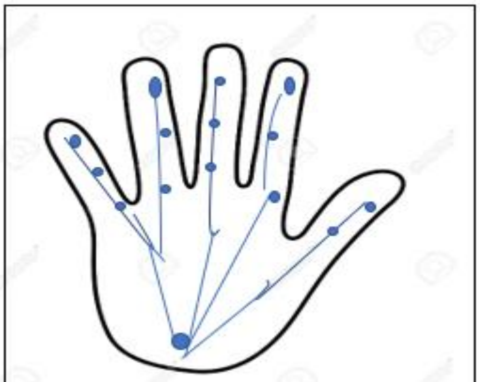
Fig. Point for hand


Fig. Skeleton points for human body

5.3.4. *Translation System*

SLR translation system includes components like Facial emotional recognition (FER), an RNN and 3DCNN. This whole process should have some pre-rules and post-rules. The operation will start by receiving frame at RNN whose first pre-rules checks whether the frame includes pixels. Upon confirming whether the frame contains pixels RNN registers the frame as the first frame in sign language movement it is trying to recognize. If it doesn’t contain the pixels then this process terminates, which cause the processing engine to wait additional data, may continue the recognition process based on its own internal logic and processing, or may exit all together. Post registering the frame the 3DCNN process the frame and checks whether the frame matches the frame of any action or motion it has trained to recognize. The output of the recognition is sent to 3DCNN to verify. If the results are not positive, RNN receives negative indication from 3DCNN and sends indication actions cannot be recognized, thus terminates the current process. If the results are positive, 3DCNN send an indication to RNN i.e., frame have matched. The RNN receives this information and before accepting it, executes its pre rule to determine whether 3DCNN was successful, and if it was, sends frame to FER module collaborator. The operations at FER module can include using the pre-rule to check whether the part of frame SL it can recognize. Upon confirming that frame corresponds to at least one candidate SL movement, FER performs facial and emotion detection and passes the results to RNN. FER modules rely on training the images that includes relevant emotions that are commonly used as a part of sign languages. Pose estimation results are available on subset of training images and during execution phase of neural network, these training images may be used to recognize emotions in captured videos in real time. RNN executes next pre rule which checks if FER has detected emotion, if yes RNN executes next pre-rule that checks whether the emotion is compatible with the sign language movement. If RNN determines that the 3DCNN recognized movement and the FER module recognized emotions are compatible, the process moves to operation identified, wherein RNN persists both 3DCNN and FER module data. If emotions detected by FER and sign language movement by 3DCNN are not compatible, RNN persists the frame but not emotions. This is followed by RNN continuing to identify the sign language movement based on subsequent frames, or existing the current process if there are no subsequent frames or information

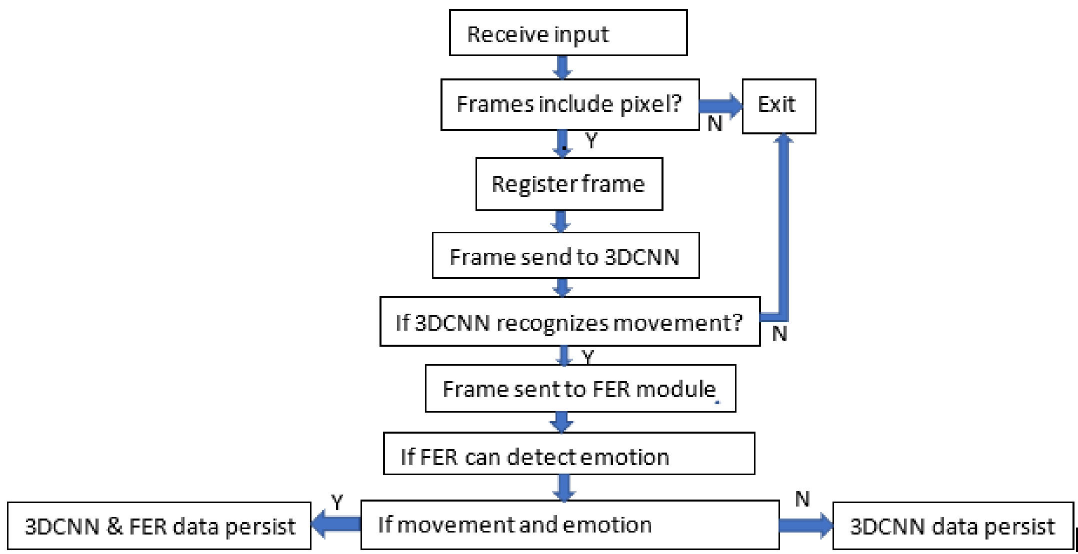

Fig. Flowchart for the operation of Sign language translation system

6. Future Scope and Social Impact

Efficient Sign gesture recognition system will help to have easier communication between impaired and people with no knowledge of sign languages. These will help the deaf and dumb community to communicate easier in interviews, educational institutes, malls, airport, and other public places. This will have great social impact, as deaf and dumb people will participate in many social activities and many new opportunities will be opened for them. Apart from impaired community, developer can advance this model for communicating with devices.

7. Relevant References

1. Kumud Tripathi∗, Neha Baranwal and G. C. Nandi, Continuous Indian Sign Language Gesture Recognition and Sentence Formation, Procedia Computer Science 54 (2015) 523 – 531
2. Suharjitoa*, Ricky Anderson, Fanny Wiryanab, Meita Chandra Ariestab, Gede Putra, Kusumaa, Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output, 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, 13-14 October 2017, Bali, Indonesia
3. HAMZAH LUQMAN, An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion, SDAIA-KFUPM Joint Research Center for Artificial Intelligence, KFUPM, Dhahran 31261, Saudi Arabi

4. B. P. Pradeep Kumar1* and M. B. Manjunatha2, Performance Analysis of KNN, SVM and ANN Techniques for Gesture Recognition System, Indian Journal of Science and Technology, Vol 9(S1), DOI: 10.17485/ijst/2017/v9iS1/111145 December 2016
5. Pavlo Molchanov, Shalini Gupta, Kihwan Kim, and Jan Kautz, Hand Gesture Recognition with 3D Convolutional Neural Networks, NVIDIA, Santa Clara, California, USA
6. MUNEER AL-HAMMADI, (Member, IEEE), GHULAM MUHAMMAD, (Senior Member, IEEE), WADOOD ABDUL, (Member, IEEE), MANSOUR ALSULAIMAN, MOHAMED A. BENCHERIF, AND MOHAMED AMINE MEKHTICHE, Hand Gesture Recognition for Sign Language Using 3DCNN, Deanship of Scientific Research (DSR) with King Saud University, Riyadh, Saudi Arabia,
7. Shogun Katoch a, Varsha Singh b,*Uma Shanker Tiwary bIndian Sign Language recognition system using SURF with SVM and CNN,
8. Ulrich von Agri's Æ Jo¨rg Zieren Æ Ulrich Canzler Æ Britta Bauer Æ Karl-Friedrich Kraiss, Recent developments in visual sign language recognition
9. Thad Starner and Alex Pentland, Real-Time American Sign Language Recognition from Video Using Hidden Markov Models, Massachusetts Institute of Technology
10. Longlong Jing∗Elahe Vahdani∗Matt Huenerfauth · Yingli Tian†, Recognizing American Sign Language Manual Signs from RGB-D Videos
11. Marek Hrúz 1,*,† , Ivan Gruber 1,*,† , Jakub Kanis 1, Matyáš Bohá ˇcek 1,2 , Miroslav Hlavá ˇc 1and Zden ˇek Kr ˇnoul 1, One Model is Not Enough: Ensembles for Isolated Sign Language Recognition
12. Jie Huang, Wengang Zhou, Houqiang Li, and Weiping Li, SIGN LANGUAGE RECOGNITION USING 3D CONVOLUTIONAL NEURAL NETWORKS, University of Science and Technology of China, Hefei, China