Arabic Sign Language Recognition System on Smartphone

THESIS SUBMITTED IN PARTIAL FULFILMENT OF REQUIREMENT
FOR THE AWARD OF THE DEGREE OF

Master of Technology
in
Software Engineering

Under the guidance of
Prof. Rajni Jindal
(H.O.D, Department of Computer Science and Engineering)
Delhi Technological University

Submitted By -
Abbas Muhammad Zakariya
(Roll No. 2K17/SWE/21)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahabad Daulatpur, Main Bawana Road, Delhi-110042

June 2019

CERTIFICATE

This is to certify that report entitled "**Arabic Sign Language Recognition on Smartphone**" submitted by **Abbas Muhammad Zakariya** (Roll No. 2K17/SWE/21) in partial fulfilment of the **MASTER OF TECHNOLOGY** degree in Software Engineering at DELHI TECHNOLOGICAL UNIVERSITY is a record of the original work carried by him under my supervision.

Place: Delhi

Date:

SUPERVISOR

PROF. RAJNI JINDAL

Head of Department

Department of Computer Science and Engineering

Delhi Technological University

Bawana Road, Delhi -110042

DECLARATION

I hereby declare that the thesis work entitled **"Arabic Sign Language Recognition on Smartphone"** which is being submitted to Delhi Technological University, in partial fulfilment of requirements for the award of degree of Master of Technology (Software Engineering) is a bonafide report of Major Project-II carried out by me. The material contained in the report has not been submitted to any university or institution for the award of any degree.

Place: Delhi | ABBAS MUHAMMAD ZAKARIYA

Date: | Roll No. 2K17/SWE/21

ACKNOWLEDGEMENT

I am very thankful to **Professor Rajni Jindal** (Head of department, Computer Science Engineering Department) for guiding me throughout the process of carrying out this project and also the faculty members Computer Science Engineering Department of Delhi Technological University for the full support they gave us at all time throughout the duration of our studies and more especially on this project, all my gratitude to the university for providing all the necessary facilities needed infrastructure, laboratories and comfortable working environment to work on with no difficulties faced. I also would like to express my appreciation to all support provided by the laboratory assistants regarding various areas and at any time requested.

Last but not the least, I would like to express my sincere gratitude to my parent and friends for constantly encouraging me throughout the duration of my study.

ABBAS MUHAMMAD ZAKARIYA

Roll No. 2K17/SWE/21

M. Tech. (Computer Science and Engineering)

Delhi Technological University

ABSTRACT

Deaf And other verbally challenged people face challenges most of the time communicating with the society, sign language is what they commonly use between them to represent what they want to say to each other for example numbers, words or phrase. To bridge this communication Barrier between them and the society an automated system to stand as a translator between them and the society is needed, which will translate the sign language into text or speech so that the communication would be easier. Recently many researches have been done in such area, but most of the developed Systems are only executable on computers, which are difficult and impractical to take around.

Research on sign language recognition for Arabic language is relevantly few compared to other languages, we are proposing in our study the use of smartphone as a platform for Arabic Sign Language recognition system, because of its portability and availability in the society, as previews studies shows the power and computational constraints of smartphones, we propose a system where most processing task is taken off the smartphone, a client server application system is to be implemented where the client would be a smartphone application that will Capture an image of the Sign to be recognized and sends it over to the server and in turn the server returns the predicted sign. On the server application where most of the sign recognition task takes place, sign image background is detected under HSV color space and set to black, the sign gesture is segmented by detecting the largest connected component in the frame, features extracted from the frame are the binary pixels, Support Vector Machine is used to classify our sign images, we are able to classify 10 Arabic Sign Language with an experimental accuracy result of 92.5%.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1. SVM: Support Vector Machine
2. ASL: American Sign Language
3. ArSL: Arabic Sign Language
4. ANN: Artificial Neural Network
5. DT: Decision Tree
6. HSV Hue Saturation Value
7. BSL British Sign Language
8. LSF French Sign Language
9. LSE Spanish Sign Language
10. HCI Human Computer Interaction
11. RGB Red Green and Blue
12. SURF Speeded Up Robust Features

CHAPTER 1

INTRODUCTION

1.1 Introduction

Sign language is the medium of communication between verbally challenged people, hand gestures is what they commonly use between them to represent what they want to each other for example numbers, words or phrase. They as well use face and body gestures, the majority of the people in the society don't understand sign language, so that brings the communication gap, which gets the interest of researchers to try as much as they can to minimize that communication gap.

Many researches have been recently made on sign language recognition system using aiding devices like gloves that are having inbuild sensors [1]. Similarly, other more complex which uses camera and Kinect devices to aide in capturing acceleration movement [2]. But majority of those research done are using computers as their platform, which are difficult and impractical take around. The effective device to address this problem is the smartphone, it can and with very good manner bridge the communication gab which is to be addressed between the people and deaf-mute. In this major project I, we are proposing in this major project I an Arabic Sign Language (ArSL)recognition system on smartphone platform.

As in natural language, gesture-based communication isn't all inclusive; it changes as per the nation, or notwithstanding as per the locales. Communication via gestures in the Arab World has as of late been perceived and recorded. Numerous endeavors have been made to set up the gesture-based communication utilized in individual nations, including Jordan, Egypt, and the Gulf States, by attempting to institutionalize the sign language and spread it among individuals from the hard of hearing network and those concerned. Such endeavors delivered many gesture-based communications, nearly the same number as Arabic-talking nations, yet with a similar sign letter [3].

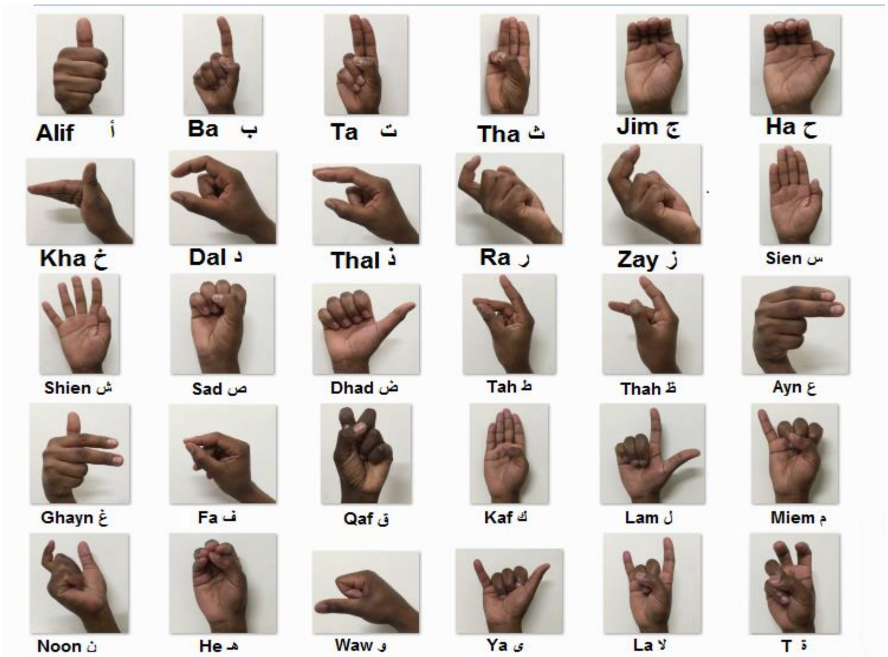


Figure 1 Arabic sign language alphabets

Figure 1 above depicts Arabic Sign Language Alphabets. Communication via gestures recognition applications in a perfect world includes the accompanying stages segmentation, feature extraction and classification. The principle goal of the segmentation stage is to expel the background and noise, leaving just the Region of Interest (ROI), which is the main helpful data in the picture. In the feature extraction stage, the unmistakable highlights of the ROI will be extracted. These highlights can be the flows, edges, shapes, corners, minutes, surfaces, colors or others. With regards to communication through sign language, these highlights are basically practically equivalent to the personality of each communication via gestures motion. Next, the feature removed will experience classification whereby the features of each sign will be assembled in like manner, and this will be utilized as a database to predict new sign and decide to which group of gestures they belong [5].

In spite of numerous works have been done of sign language system with computer platform, next to no has been done on cell phone platforms. Past research on sign language recognition cell phone have demonstrated that the major disadvantage of phones are the computational and resources limitations and constrains [6]. Be that as it may, the advantages of utilizing a cell phone platform over computer platform are its portability, accessibility, and usability to mention but few. [5].

CHAPTER 2

Background and Motivation

In this chapter the following are discussed: About Sign Language, Sign Language Recognition process which includes Skin segmentation, Feature extraction and classification,

2.1 Sign Language

2.1.1 History

Sign languages are been used throughout history by deaf communities. In the fifth century BC found one of the earliest sign language records from plato's Cratylus in which Socrate says: "If we hadn't a voice or a tongue, and wanted to express things to one another, wouldn't we try to make signs by moving our hands, head, and the rest of our body, just as dumb people do at present?".[14] Fingerspelling systems are among the only most known about historical sign language till the 19th century, which were used to translate words from spoken language to sign language, the first manual alphabet for sign language was developed by León (1520–1584).[15]

In the 1720, the present British fingerspelling systems was formed and it is used by the deaf communities or we can say in the class rooms, and it is as well used in British colonies like South Africa, Uganda, India, Australia and New Zealand, and also Norway, Germany, the United States Grand Cayman an Island in the Caribbean, Indonesia, republics and provinces of the former Yugoslavia. In the 18th century manual alphabet was published by the French man called Charles-Michel de l'Épée which till date stands unchanged in North America and France.

Meanwhile in the 1755, Abbé de l'Épée has founded a school specially for deaf children in France precisely in Paris, Gallaudet University which is the only deaf people's liberal arts University in the whole world, it was founded by Edward Miner Gallaudet in the 1857 Washington, D.C., and in the 1864 it has become the National Deaf-Mute College. Thomas Hopkins Gallaudet who is Edward Miner Gallaudet has come along with Laurent Clerc from Paris to the United States where they founded the American School for the Deaf in the 1817 at Hartford, Connecticut, Laurent Clerc was one of the most famous graduate of the Deaf School for children in Paris which was founded by Abbé de l'Épée.[16]

Ideally most people might think that there is a linguistic relation between Sign languages and Spoken languages, but Sign language has generally no such relation to its corresponding spoken language but rather there is a complex correlation between them and it varies much to the country in which the sign language is used than even the spoken language, if we could take some examples on such scenarios we could observe that American Sign Language(ASL) is basically has been driven from the French Sign Language but it is used in America and English-speaking Canada, like whiles Australia, New Zealand and the United Kingdom use varieties of Australian, New Zealand and British Sign language, which are not similar to the American Sign Language, the national language for Mexico and Spain is Spanish but the Sign language used in both the countries are totally different, Similarly the Sign language used in Bolivia is based on American Sign Language instead of any other sign language that is used in a Spanish country, and sometimes sign languages in a region could be correlated to the location of Deaf schools in the geographical area and not necessarily corresponding to the different numbers of dialects used in the country.[17][18]

International organization and the World Federation of Deaf gave a term “International Sign” which is a sign language that is used mostly in an international events for Deaf like the Deaflympics and also is used at the World Federation of Deaf meetings, International Sign is also sometimes referred to as Gestuno, while some studies showed that it is a pidgin and it is complex for a typical pidgin but is more like a couplet Sign language standing on its own.[19]

2.1.2 Relationships with spoken languages

That misconception of sign language is somehow related to spoken language is always there, where people think sign language are expressions of spoken language in signs and also the misconception that sign languages are actually invented by people which have no hearing disability.[20] Deaf school teachers with no hearing disabilities like de l’Épée, Charles-Michel and even Thomas Hopkins Gallaudet are incorrectly referred to as the inventors of sign languages which is not correct because sign languages are basically invented by the deaf people themselves who don’t or have a very less knowledge of spoken languages.

It is clear that spoken languages developed by borrowing some elements from other different spoken languages they are in contact with, like whiles sign language does borrows from spoken language but they borrow elements from spoken languages in different ways unlike spoken language does, in most sign languages proper names of peoples and places are spelled out by using Fingerspelling or also is referred to as manual alphabet, so sign language does borrows word from spoken language by spelling it using manual alphabets and as well there are so phrases or concept which at a moment that they have no particular sign to represent them, in such case as well they are spelled by manual alphabet, sometimes in Fingerspelling are used as the source or base for a brand new sign, some of such are called initialized signs where the beginning letter of a spoken word will be represented by a handshape where it carries same meaning of the spoken word.

2.1.3 Acquisition

Children with no hearing disabilities as we all know acquire their spoken natural native language automatically as they grow, similarly deaf children acquire sign language as they grow up with no problem, due to the plasticity of the brain of children either of spoken and sign languages are much easier to acquire at that young age than at adult age as suggested in languages by the Critical Period hypothesis.

The study from University of McGill done regarding who natively acquired sing language and those learned it shows that for the American Sign Language (ASL) users that natively acquired the language (by birth) have much better performance than those who did learned the language when a task was given to them to copy videos showing sentence on American Sign Language(ASL), as well the study found that there is difference between the two groups in their ASL sentences grammatical morphology , so the study concluded that there is a very special and critical period on learning or acquiring sign language. [21]

2.1.4 Interpretation

Sign language interpreters are those who translate sign language to spoken language, sign language interpreters are used to facilitate the communication of the deaf people and the hearing people, sign languages have their syntax not similar to spoken language so that will require a considerably higher amount of effort from the interpreter’s part.

The flow of interpretation is most of the time between sign language and spoken language which are been used in a particular same region or country, for example spoken English the United Kingdom and British Sign Language(BSL), the France spoken French and French Sign Language(LSF), The United States spoken English and the majorities of the anglophone Canada and the American Sign Language(ASL), and the Spain spoken Spanish with the Spanish Sign Language(LSE), etc. although its less frequent, but there are interpreters who can also translate between non paired sign language and spoken language for example French Sign Language (LSF) and American Sign Language(ASL).

Recently there have been a development in computer science and artificial intelligence (AI) precisely Computer Vision, algorithms developed using deep learning can have the ability to do a machine translation which automatically translate into written text a short video of sign language which might contain some few sentences, mostly very simple clause or we can sentence.[22]

Remote interpreting:

Ideally an interpreter is supposed be physically available or present to perform the interpretation between the two group or parties, but from the 2000s there has been a revolutionary development the technological sector, so it is from then become easily and reliably possible to provide an interpreter remotely, where the interpreter would be in a different geographical location from the sign gesturer and also the people to be translated to and all the three parties might be in a different geographical location, technically there is this terminology called the Video Remote translation(VRI), here we have client number one which is the signer (user of Sign language) and client number two who is or are the hearing people who would like to communicate with client number one, Both of the client will be in the same place while the interpreter is in another remote location, a video communication media will be provided between the interpreter and the sign language user, an audio communication link is provided between the interpreter and the hearing people, VRI mostly is used when there is no available interpreter on site.

There is as well a situation where VRI must be used, in a scenario where the two clients are not in same location but in a remotely different locations and will like to communicate between each other, so in such a case all the three parties are in different location the interpreter is used to allow the clients communicate to each other using VRI.

Interpretation on television:

In Television programs at times sign language is simultaneously provided, at the corner bottom of the television screen you will notice the sign language user is there, while the television program occupies the full television screen, the Mayor of New York City in his press conference you will notice a sign language translator is provided on the stage just beside the public official, allowing the both of them appearing on a single frame exactly at the same time.

providing sign translation to television programs tend to cause some distraction to some of the people watching the program, for that reason some programs are repeated twice one with no sign translation another with a sign language translation, for BBC you might notice the majority of their program in the early morning and the late nights are provided by sign language translation, there are some television technologies that may provide the option of hiding or showing sign language translation or even hide or show a subtitle for the television programs.[23]

2.1.5 Telecommunications

When we go back a couple of decades to see how telecommunication started to help sign language, we can see on the 1964 New York World's Fair the first communication remotely between two deaf was made each on different city, on the World's Fair the company AT&T introduced their videophone which was given a trademark of "Picturephone". However, due to the limitation of bandwidth by then video communication was not made available until the 2000s when sufficient bandwidth was available.

The Internet now provides wide range and a variety of video communication platforms or services which could be used by deaf people to communicate with each other, some platforms are specially equipped for deaf users in which they have features that might help the deaf, for example the ability to zoom in to the video to concentrate on the sign gestures.[24]

2.2 Skin segmentation

Skin recognition is the way toward discovering skin-colored pixels and locales in a picture or a video. This procedure is commonly utilized as a preprocessing venture to discover districts that conceivably have human appearances and appendages in pictures. Skin picture acknowledgment is utilized in an extensive variety of picture handling applications like face recognition, skin disease recognition, motion following and HCI (Human Computer Interaction). The essential key for skin recognition from a picture is the skin coloring. Yet, color can't be the main integral factor because of the variety in skin tone as indicated by various races. Different factors, for example, the light conditions likewise influence the outcomes. Subsequently, the skin tone is regularly joined with different signs like surface and edge highlights. This is accomplished by separating the picture into individual pixels and ordering them into skin colored and noncolored. One basic technique is to check if each skin pixel falls into a characterized color extent or qualities in a few directions of a color space. skin color spaces include RGB, HSV, YCbCr, YIQ, YUV, etc. and they are all might be used in skin color segmenting [12].

2.2 The Color space

This is a model that can be used to represent color data into four or three different color components. Different color models are used for different applications such as computer graphics, image processing, TV broadcasting, and computer vision. Different color space is available for the skin detection. They are: RGB based color space (RGB, normalized RGB), Hue Based color space (HSI, HSV, and HSL), Luminance based color space (YCBCr, YIQ, and YUV)

These models are clarified in this way in next areas. color space determination is the essential procedure in skin color demonstrating and further for arrangement. At least one-color spaces can give an ideal limit an incentive for discovery of pixels of skin in a given picture. The decision of suitable shading space is regularly dictated by the skin location system and the application [12].

2.2.1 Red, Green, and Blue (RGB) Color Model

RGB color space is generally utilized and is regularly the default color space for putting away and speaking to advanced pictures. We can get some other color space from a straight or non-direct change of RGB. The RGB color space is the color space utilized by PCs, designs cards and screens or LCDs. As appeared in figure 2 underneath from [12] it comprises of three parts, red, green and blue, the essential hues. Any color can be acquired by blending the three base hues. Contingent upon what amount is taken from each base color, any color can be made. Switching this procedure, an explicit color can be separated into its red, blue and green segments. These qualities can be utilized to discover comparative color pixels from the picture. Standardized RGB is a portrayal that is effectively acquired from the RGB esteems by a straightforward standardization strategy [12].

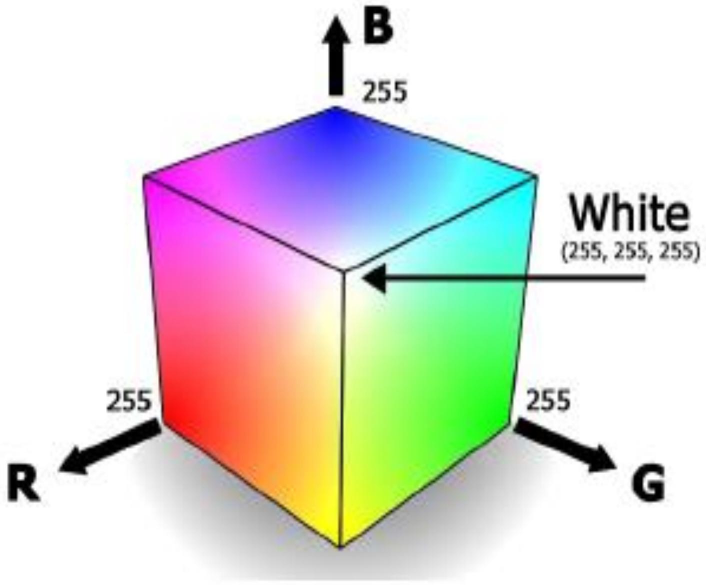


Figure 2 RGB Color Model

2.2.2 YCbCr (Luminance, Chrominance) Color Model

YCbCr is an encoded non-linear RGB signal, regularly utilized by European TV studios and for picture pressure work. As appeared in figure 2 below from [12], color is spoken to by luma (which is luminance found from nonlinear RGB) developed as a weighted total of RGB values[12]. YCbCr is a normally utilized color space in advanced video area.

Since the portrayal makes it simple to dispose of some excess color data, it is utilized in picture and video pressure guidelines like JPEG, MPEG1, MPEG2 and MPEG4.The change effortlessness and unequivocal partition of luminance and chrominance segments makes YCbCr color space. In this configuration, luminance data is put away as a solitary segment (Y), and chrominance data is put away as two color distinction parts (Cb and Cr). Cb speaks to the contrast between the blue part and reference esteem. Cr speaks to the contrast between the red part and a reference esteem.

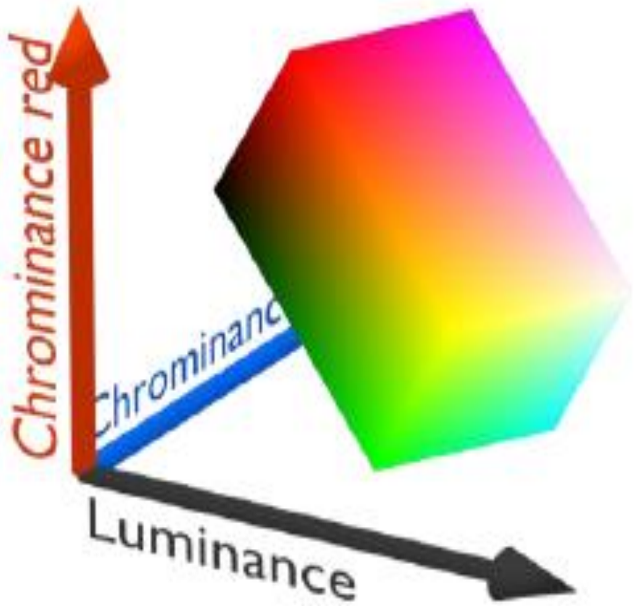


Figure 3 YCbCr Color Model

2.2.3 Hue Saturation Value (HSV) Color Model

The HSV shading space is more instinctive to how individuals encounter color than the RGB color space. As tint (H) shifts from 0 to 1.0, the relating hues differ from red, through yellow, green, cyan, blue, and fuchsia, back to red. As saturation(S) differs from 0 to 1.0, the comparing hues (tones) shift from unsaturated (shades of dim) to completely immersed (no white segment). As esteem (V), or splendor, fluctuates from 0 to 1.0, the relating hues turn out to be progressively more brilliant. The color part in HSV is in the range 0° to 360° point all lying around a hexagon as demonstrated figure 4 [12]. With RGB the color will have values like (0.5, 0.5, 0.25), while for HSV it will be (30°, √3/4, 0.5) [12].

HSV is best utilized when a client is choosing a color intuitively It is generally a lot simpler for a client to get to a coveted color when contrasted with utilizing RGB.

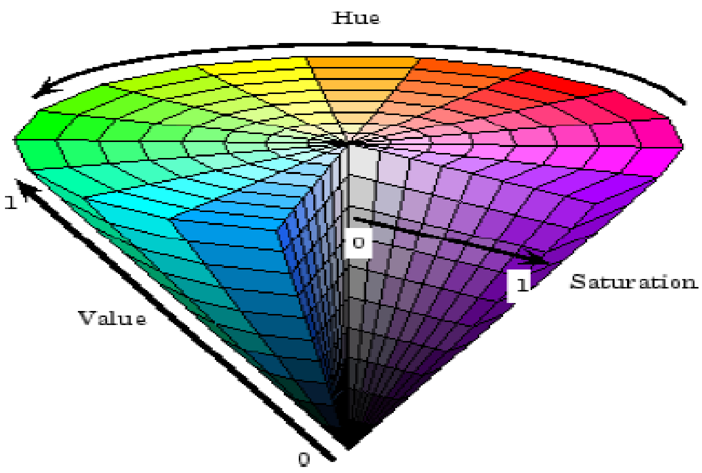


Figure 4 HSV Color Model

2.3 Feature Extraction

There are many feature extraction mechanisms, one of the most popular and widely used is the histogram of oriented gradients, we are going to talk about that in detail.

2.3.1 Histogram of Oriented Gradients

The histogram of the dark dimensions of an image or of a segment of an image is frequently utilized a) for edge detection, b) for extraction of features, and c) as a manual for transforming the dimensions higher than ever to encourage the showcase of the picture. The last change results in another histogram a histogram of the changed dim dimensions that may enhance the PC's capacity to do edge or limit location and extraction of textural features.

Let g denote an arbitrary gray level, and let b denote an arbitrary interval width. Usually b is restricted to small values with respect to g. Let N denote the number of pixels whose gray values lie in the interval (g - (b/2), g + (b/2)). Since most medical images have large low-frequency Fourier harmonics, the gray level at any pixel is in most instances close to the gray levels at neighboring pixels. Consequently, when b is small, the N pixels whose gray levels lie in (g - (b/2), g + (b/2)) form just a few connected sets of points or "components" in the x-plane [13].

The segments of p(x) determined by these components constitute an N-pixel slice of p(x). Consider a set of N-pixel slices of p(x). In some textural pictures (i.e., pictures consisting of large regions of nearly uniform textures) the sizes, shapes, and distributions of the components among these slices are similar. Under this condition a measure of the visibility of details among these slices is the variation of gray level-or local contrast-within each slice. This variation is approximately proportional to the interval width b. Let h(g) denote the histogram of g, i.e., the frequency of occurrence of g in the picture.

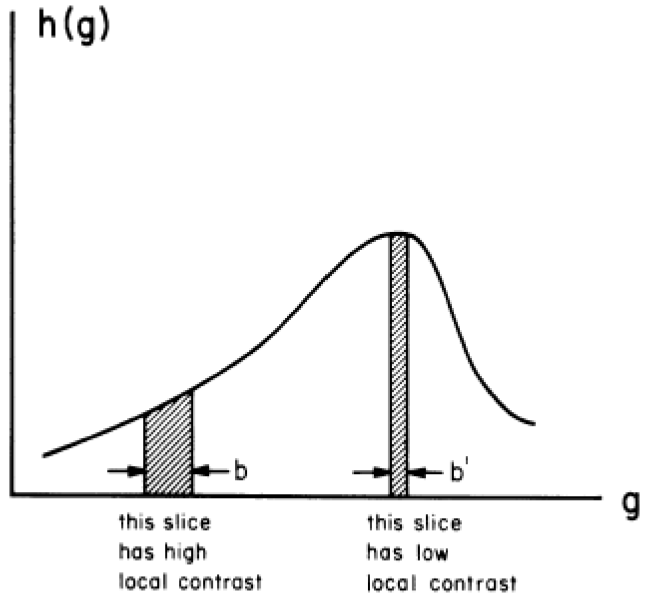


Figure 5 Relation of local contrast to h(g).

When b is small, and the histogram h(g) can be approximated as a continuous function; b and h(g) are related as follows:

$$N = \int_{g-(b/2)}^{g+(b/2)} h(y)\, dy \cong bh(g).$$

Thus, under the assumption of low-frequency dominance, the relative visibility of details at gray level g is approximately inversely proportional to h(g). This relation of the local contrast to h(g) is illustrated in Fig. 5.[13].

2.4 Classification

Several factors could affect the classification of image, it is a very complex process, for every a good and successful classification sufficient data or training samples are a very necessary, selection of the right classification method is not an easy task at all because there are many factors to be considered, for different classification there is different advantages over others, for there will be different classification results on each classification method selected and used, talking of classification methods many classification approaches are used some include ANN, fuzzy and expert system.[25]

The following list categories of classification approaches with their examples:

1. **Supervised classification approaches:** Decision Tree classifier, ANN, Minimum Distance and Maximum Likelihood.
2. **Unsupervised classification approaches:** K-means Clustering and ISODATA
3. **Per-pixel Classifier:** SVM, ANN, Minimum Distance and Maximum Likelihood, Decision Tree.
4. **Subpixel Classifiers:** Spectral mixture analysis, Fuzzy-set and subpixel classifier
5. **Parametric Classifiers:** Linear discriminant analysis and Maximum Likelihood.
6. **Nonparametric Classifiers:** Evidence reasoning, ANN, SVM, Expert System and Decision tree.
7. **Contextual Classifiers:** Iterated condition modes, frequency based contextual classifiers and point-to-point contextual correction
8. **Spectral-Contextual Classifiers:** contextual algorithms and ECHO.
9. **Object-oriented classifiers:** eCognition**.**
10. **Per-field classifiers:** GIS-based classification approaches.
11. **Hard Classification:** SVM, ANN, Minimum Distance and Maximum Likelihood, Decision Tree.
12. **Soft Classification:** Spectral mixture analysis, Fuzzy-set and subpixel classifier.
13. **Spectral Classifiers:** ANN, Minimum Distance and Maximum Likelihood,

CHAPTER 3

Related Work

3.1 Related research reviewed

Sign language has dependably been a piece of human medium of communication. The utilization of motions or sign isn't fixed to ethnicity, age, or sexual orientation [3]. As of late, many researches have proposed plenty of approaches for the recognition of sign language.

Sakshi Lahoti et al [8] presented an android based approach to recognize American Sign Language, they used YCbCr systems to segment the skin from an image captured by an android smartphone and later HOG is used to extract features from the image then finally Support Vector Machine (SVM) is used Classification the system has obtained an accuracy of 89.54%.

Setiawardhana et al [7] developed an android sign language translation, OpenCV is used for hand detection and K-NN is used for classification, their system is able to detect palm of hand for gesture recognition up to 50 cm away and optimally provided the hand is placed in a 0° position upright.

S. M. Halawani [3], similarly has proposed an approach to Arabic Sign Language Translation Systems (ArSL-TS), he introduced a model which runs on cell phones to build up an animated Arabic sign gesture from Arabic text given as an input to the model which enables user to make an interpretation of Arabic text into Arabic Sign Language for the people with hearing problems on cell phones.

M. Huenerfauth [4] has developed a Machine Translation prototype of American Sign Language classifier predicate to translate from English to American Sign Language which will help some of the individuals within the deaf community with reasonable English writing and reading level to gain information and interact with other part of the society, he also discussed some of the significant previews Sign Language Machine Translation Researches.

Cheok Ming Jin et al [5] proposed a mobile approach for American Sign Language recognition, Canny Edge detection with seeded region growing are used for hand gesture segmentation from the hand gesture image, Speeded Up Robust Features (SURF) algorithm is used for the extraction of feature points which are later on classified using Support Vector Machine (SVM), their system is able to predict 16 Classes of American Sign Language at an accuracy of 97.13%.

Ruslan Kurdyumov et al [10] have developed a system to give immediate feedback for who are learning American Sign Language, users just need to use their device webcams to practice the American Sign Language and the system shows them how good they practice the gesture and how they will improve it, it serves as just an instructor. For feature extraction they normalized and scaled their gesture images to 20 x 20px and use the pixels as their features, they have used K-NN and SVM for classification and found out that SVM has more classification accuracy than K-NN where SVM recorded an accuracy of around 93%, it has outperformed K-NN by 10%.

Sign language has dependably been a piece of human medium of communication. The utilization of motions or sign isn't fixed to ethnicity, age, or sexual orientation [3]. As of late, a few researches extend in sign language have been discussed in [4]. In [3], an Arabic Sign Language Translation Systems (ArSL-TS) is presented. That introduce model which runs on cell phones to build up a symbol-based gesture interpretation to avatar demonstration that enables clients to make an interpretation of Arabic text into Arabic Sign Language for the people with hearing problems on cell phones, for example, Personal Digital Assistants (PDAs). Setiawardhana et al [7] developed android sign language translation on android platform. They use the KNN model to detect the hand gestures.

Promila Haque et al [11] developed a two-Handed Bangla Sign Language Recognition system which recognize 26 sign gestures, they structured the system to three phases data formation, training and classification phase. They extracted image principal component by using Principal Component Analysis (PCA) and used K-Nearest Neighbours as their classification algorithm, they obtained a success rate of 77.8846% by testing 104 images.

The existing research on Sign language recognition systems includes systems which uses data from sensors embedded on hand gloves for hand tracking to know the hand position each time a gesture is made, other system use Kinect device to detect movement of hand and its acceleration, hand gloves with sensors cannot withstand weather changes like rain and humidity, systems which use Kinect device and those which executes only on computer are impractical to take around.

With the development of smartphone and its continues computational power improvement such system become much easier to be accommodated by smartphone, therefor the portability problem will be well addressed and the system will be available for all smartphone users, some similar application are developed like [10] which can run on android devices but these applications exhaust more memory and computational power. We propose a client server system which will address these drawbacks.

We are going to extensively explain a related work in [10], their work was an implementation of American Sign Language recognition through an Android app.

Dataset:

- American Sign Language using 36 symbols
- Alphabets A to Y, 0-9, spacebar.
- Z not involved Z is not a static gesture
- each symbol, 500 training images
- size 200X200 pixels
- black background

process are as follows:

1. Hand gesture capture and skin segmentation
2. Feature Extraction
3. Classification using SVM

A. <u>Hand gesture capture and skin segmentation</u>

The gestures are caught utilizing cell phone camera. The camera is begun and video taken is partitioned into littler frames with the goal that the hand gesture is appropriately recognized.

YCbCr color space id used for Skin segmentation

- Y is the luma (brightness of an image in a video) component,
- Cb blue difference
- Cr red difference.
- conversion formula (RGB to YCbCr)is :

$$\begin{pmatrix} Y \\ Cr \\ Cb \end{pmatrix} = \begin{pmatrix} 0.2290 & 0.5870 & 0.1140 \\ -0.1687 & -0.3313 & 0.5000 \\ -0.5000 & 0.4187 & -0.0813 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + 128$$

$$
\begin{pmatrix} Y \\ Cr \\ Cb \end{pmatrix} = \begin{pmatrix} 0.2290 & 0.5870 & 0.1140 \\ -0.1687 & -0.3313 & 0.5000 \\ -0.5000 & 0.4187 & -0.0813 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + 128
$$

- If pixel value is within the range of Cb and Cr then is part skin
- values outside range non-skin part and will be converted to black pixel (0 values).

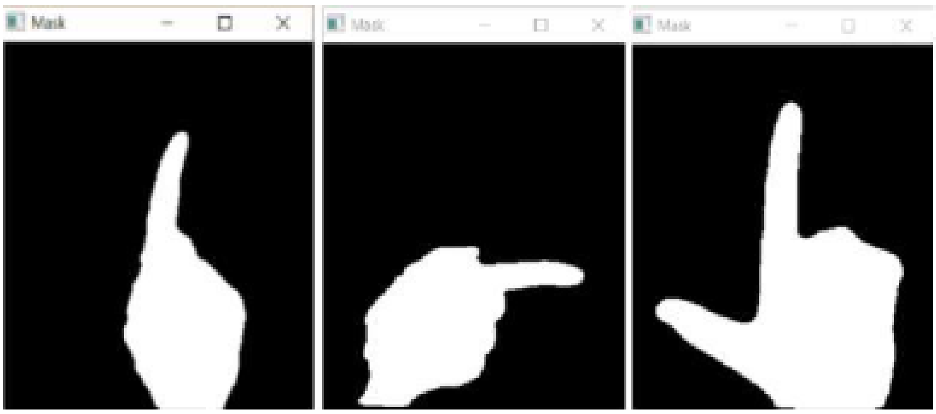


Figure 6 Segmented Image with skin detection

B. <u>Feature Extraction</u>

Histogram of Oriented Gradients is used to extract the features from the pictures after the segmentation of skin process has finished. Histogram of Oriented Gradients counts the number of gradients occurred in neighborhood of segments in an image. Histogram of Oriented Gradients first splits image into smaller cells of pixels. In each cell the direction o its histogram is calculated.in the end Histogram of Oriented Gradients makes a global histogram by scanning over all the generated cells. The main advantages of Histogram of Oriented Gradients over the SIFT method and other similar feature extraction mechanism is not affected by geometric transformations. Histogram of Oriented Gradients Algorithm:

1) Find gradient values.
2) Calculate histogram cells.
3) Change contrast and illumination
4) next step is the block normalization in which a normalization factor is calculated and multiplied to block vector.
5) Finally object recognition which Histogram of Oriented Gradients by using the Support Vector Machines (SVMs).

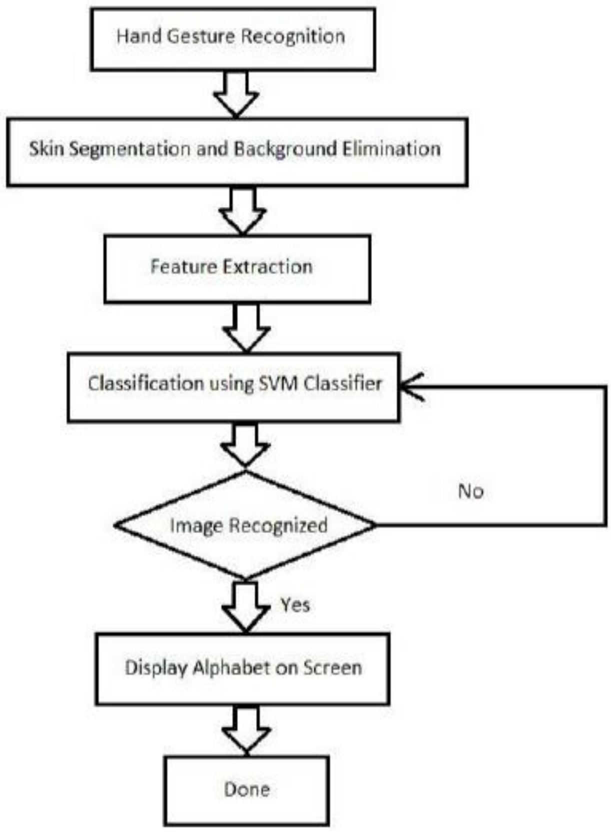


Figure 7 Data Flow Diagram

Vector Machine task is to classify a new dataset given to a particular class that is supposed to belong to, it is a non-probabilistic linear model. In Support Vector Machine, datasets are presented by points on the hyperplane. And the given training data sets are clearly segmented by gap to make the classification further simple and with a higher possible accuracy. For each new data to be predicted it is assigned a class tag according to its point with relation to the gap. Support Vector Machine defines a kernel function for it is the effective solution for such task. To get much good accuracy Support Vector Machine has this concept “soft margin function”, denoted by C, that can control effect of each support vector [11].

gamma is used which is a free parameter in the radial function and is denoted by:

$$K(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2), \gamma > 0$$

support vectors are the xi and xj.

values of C and Gamma are taken here as 2.67 and 5.383, A bigger change in gamma depicts low in variance in model and a high bias.

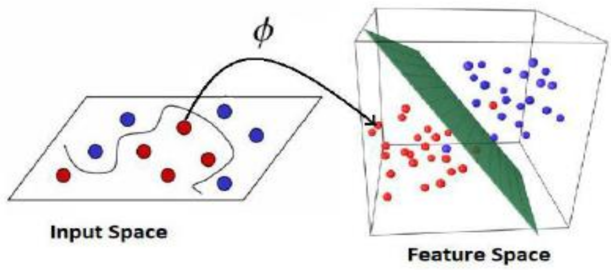


Figure 8 Hyperplane separating the input data into two

The green colored is the hyperplane shown in figure 8. There are two different classes in the box depicted in the figure above with red and blue colors. In the classification phase shift is denoted by Phi.

The figures that will be shown below depict Support Vector Machine recognition of segmented alphabets in Figure 2. The predicated output alphabet is in red color shown at the screen bottom.

The green colored is the hyperplane shown in figure 8. There are two different classes in the box depicted in the figure above with red and blue colors. In the classification phase shift is denoted by Phi.

The figures that will be shown below depict Support Vector Machine recognition of segmented alphabets in Figure 2. The predicated output alphabet is in red color shown at the screen bottom.

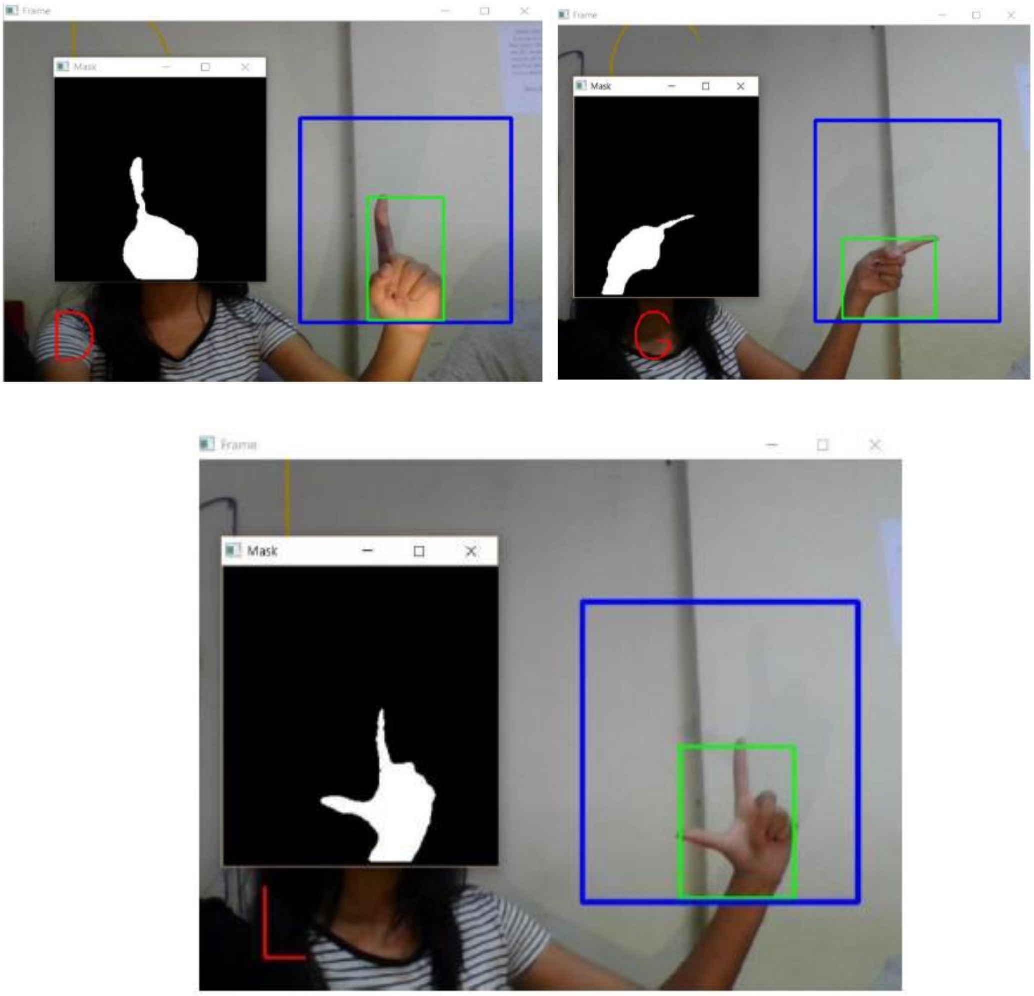


Figure 9 Recognized alphabets of the above segmentation image

Dataset are to be uploaded to cloud after the training and classification process has finished, doing so will serve the phone storage computational resources, as mentioned earlier the disadvantage of using phone as a platform is it constrains when it comes to resources, and as well the datasets will be easily accessed and accommodate future changes.

this constrains we are proposing a client server system to take off the computational task off the smartphone and a server handles all the processing, all what the smartphone needs to do its to capture the image to be predicted and sends it over to the server for prediction and in turn the server replies with the predicted value.

Support Vector Machine (SVM) has accomplished many promising results in accuracy in many of the recent researches, in [10] they used Support Vector Machine for Classification and they obtained an accuracy of about 89.54%, while in [5] SVM is applied to predict 16 Classes of American Sign Language at an average accuracy of 97.13%. Keeping that in mind we use the SVM multi-class classifier for sign prediction on the server side.

The figure below depicts the general architecture of the Arabic Sign Language recognition system on smartphone:

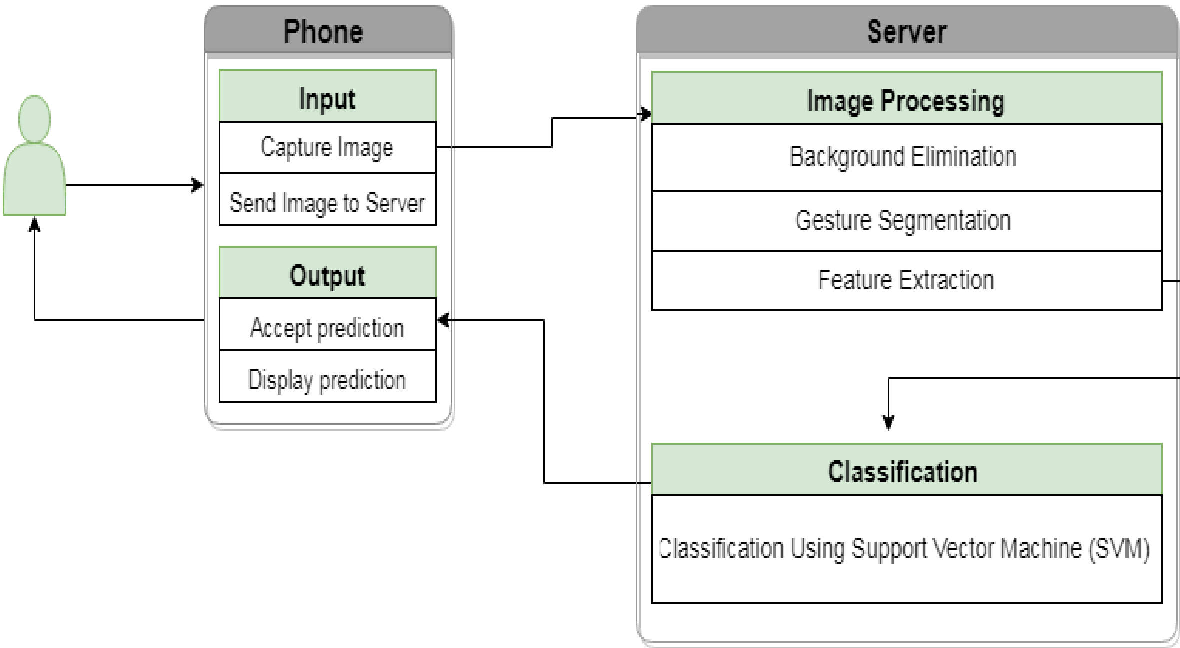


Figure 10 Arabic sign language recognition on smartphone architecture

4.2 System Overview

We are able to recognize 10 Arabic Sign Language in this paper, as depicted in figure 2 below a client server system is implemented, where on the client side a smartphone is used, the user directly interacts with the smartphone application. The application captures gesture image as an input to the system and sends it over to the server through an application programming interface (API) provided by the server application, as well this client-side application is

4.2 System Overview

We are able to recognize 10 Arabic Sign Language in this paper, as depicted in figure 2 below a client server system is implemented, where on the client side a smartphone is used, the user directly interacts with the smartphone application. The application captures gesture image as an input to the system and sends it over to the server through an application programming interface (API) provided by the server application, as well this client-side application is responsible of displaying the predicted value returned by the server, the smartphone application developed runs on android 4.0(API level 14) and higher to be able to cover maximum possible android smartphone users.

The server application provides the client-side phone application services through an API, the server is responsible of basically two major tasks, which are image processing and classification.

4.3 Smartphone application

An android application with a friendly user interface and inbuilt tutorial is developed, it is responsible of:

i) capturing the sign gesture and send it to the server for classification.

ii) receiving and displaying the predicted sign gesture returned by the server to the user.

Figure 11(a) shows the first displayed activity for the user where there is button to open the camera and capture a sign gesture while Figure 3(b) shows the activity displayed after a prediction is made and returned by the server

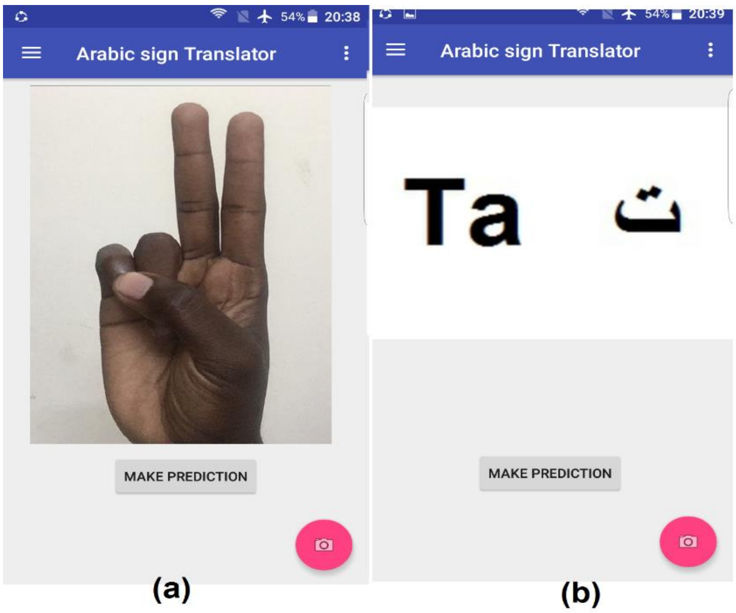
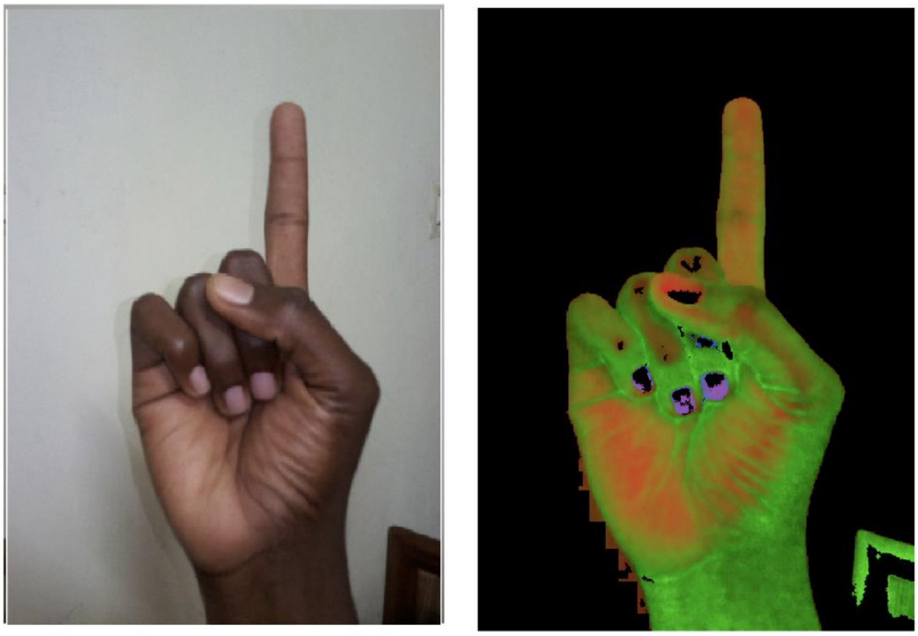


Figure 11 (a) first activity to capture sign image (b) second activity to show predicted alphabet

4.4 Background Elimination

The background of the received image from the smartphone is detected and set to black by converting the image from RGB to HSV color space, two masks are prepared with tuned parameters to match skin color and a series of erosion and dilation using elliptical kernel is made, a final frame is created by combining the effect of both masks, Figure 4(a) shows the raw image while Figure 12(b) shows the image after its background is eliminated and set to black.



(a) (b)

Figure 12 (a) Raw image (b) image after background elimination

4.5 Gesture segmentation

The image from the previews phase where the background is eliminated is first converted to grayscale, though we lose the color data of the original image but this makes our system robust to different lightning conditions, we then set all non-black pixels to white (binarize) and leaving the remaining pixels as black. The hand gesture then is segmented by removing all the connected component in the image only allowing the largest connected component which will the sign gesture, the next step is to remove the arm from the hand gesture and only be remained with the gesture, by assumption about the wrist we cropped 15 pixels from the bottom of the image, then finally the image is centered a resized to 30 x 30px. The whole processing phases are shown in the Figure 13 below:

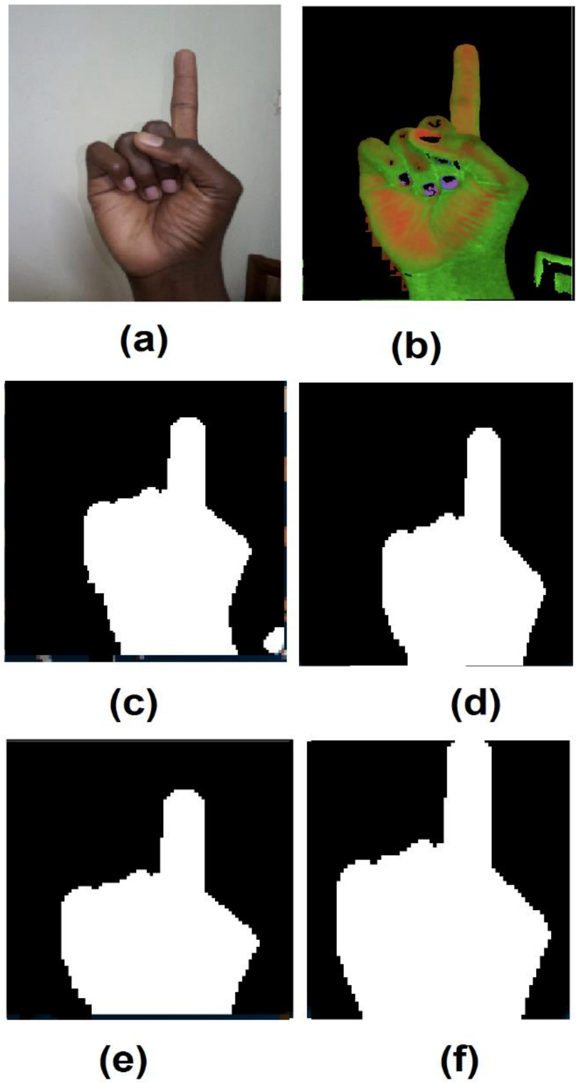


Figure 13 (a) Raw image (b) background removed (c) Binarized (d) segmented hand gesture (e) wrist removed (f) centered

4.6 Feature Extraction

The sign language gesture images are normalized and scaled to 30 x 30px, binary pixels of the images are what we used as features. We found out that scaling down to less than 30 pixels will not contain enough features to classify the Arabic Sign Language gestures efficiently because most of the distinguishing features will be lost. By using 30 x 30px we will be having 900 number of features.

4.7 Classification

Support Vector Machine (SVM) is what we used to classify our sign gesture data sets extracted from the images, SVM is a model used for classification of objects, it is under supervised learning category, its main goal is to find the best hyperplane which can separate data points of different classes, data points lying close to decision hyperplane are called the support vectors, so to obtain a decision boundary support planes will be made to for separating the support vectors of the various classes.

In this paper, 200 images for each of the 10 Arabic Sign Language will go through the image processing where the extracted features from the phase is passed over here for classification, multi class SVM is used to train our SVM model.

CHAPTER 5

Experimental Results

5.1 Results and Analysis

For each of the 10 Arabic Sign Language the model is evaluated, which contains the following Alphabets: Alif, Ba, Ta, Kha, Dal, Dhad, Thah, Ghayn, Lam, and La . We have used a total of 2000 images to train the Support Vector Machine (SVM) classifier. To evaluate the system performance, we have used two different approaches, the first approach is by splitting our training images to 20% testing and 80% training, and we have obtained an accuracy of 91.03% Table 2 shows detailed precision, recall and F-Measures for each class, while for the second approach we have used a total of 200 images 20 samples each taken directly from a smartphone and for each image there is a small scale and rotation difference an accuracy of 92.5% is recorded for the second approach.

We have found out that the accuracy of a given gesture is dependent on another visually similar gesture included in the model, where these similar sign gestures are interchangeably misclassified for example the alphabet “Ghayn” and “Thah”. while sign gestures which are different from others have a higher classification accuracy for example “Lam”, Table 1 below shows the detailed 10 Arabic Sign Language accuracy evaluation.

Table I Recognition rate of 10 Arabic Sign Language gesture

Alphabet	**Accuracy (%)**	**False prediction**	
Alif أ	90	Ba (1)	Dal (1)
Ba ب	85	Ta (1)	Alef (1)
Ta ت	95	Ba (2)	
Kha خ	95	Ghayn (1)	
Dal د	100		
Dhad ض	100		
Thah ظ	85	Ghayn (2)	Ta (1)
Ghayn غ	85	Thah (2)	Dhad (1)
Lam ل	100		
La لا	95	Ba (1)	
Average	92.5%		

Table II Precision, Recall and F-Measures

Alphabet	Precision	Recall	F-Measure
Alif أ	0.86	0.84	0.85
Ba ب	0.76	0.81	0.79
Ta ت	0.83	0.88	0.86
Kha خ	1.00	1.00	1.00
Dal د	1.00	1.00	1.00
Dhad ض	0.95	1.00	0.98
Thah ظ	1.00	1.00	1.00
Ghayn غ	0.88	0.85	0.86
Lam ل	1.00	1.00	1.00
La لا	0.78	0.72	0.75
Total	0.906	0.91	0.909

We have used the following performance metrics:

- **Precision**

Precision which is also known as predictive positive value is a slice of relevant prediction from the total prediction made.

$$Precision = \frac{true\ positive}{true\ positive + false\ positive}$$

- **Recall**

Recall also referred to as sensibility, is a slice of relevant prediction made over the total amount of relevant prediction.

$$Recall = \frac{true\ positive}{true\ positive + false\ negative}$$

To conclude precision and recall are both dependent or based on the measure of relevance.

- **F-Measure:**

F-measure also known as F1 score or F-score, it is the accuracy measure of the prediction made. To compute F1-Score two other metrics are to be considered which are the Recall r and the precision p, where precision is the total number of correct positive prediction made divided by total positive returned prediction while for Recall is divided by relevant predictions(the predictions supposed to be made), F1-score ranges between 0 and 1 perfect f and p will make F1-score of 1 while worst score of p and r gives a 0 F1-Score.

$$F - Measure = \frac{2(Precision * Recall)}{Precision + Recall}$$

- **Support**

This is the count or number of class occurrences in the datasets, if there is less balancing in support values then this is an indication that the structure of the datasets is weak and needs to be improved.

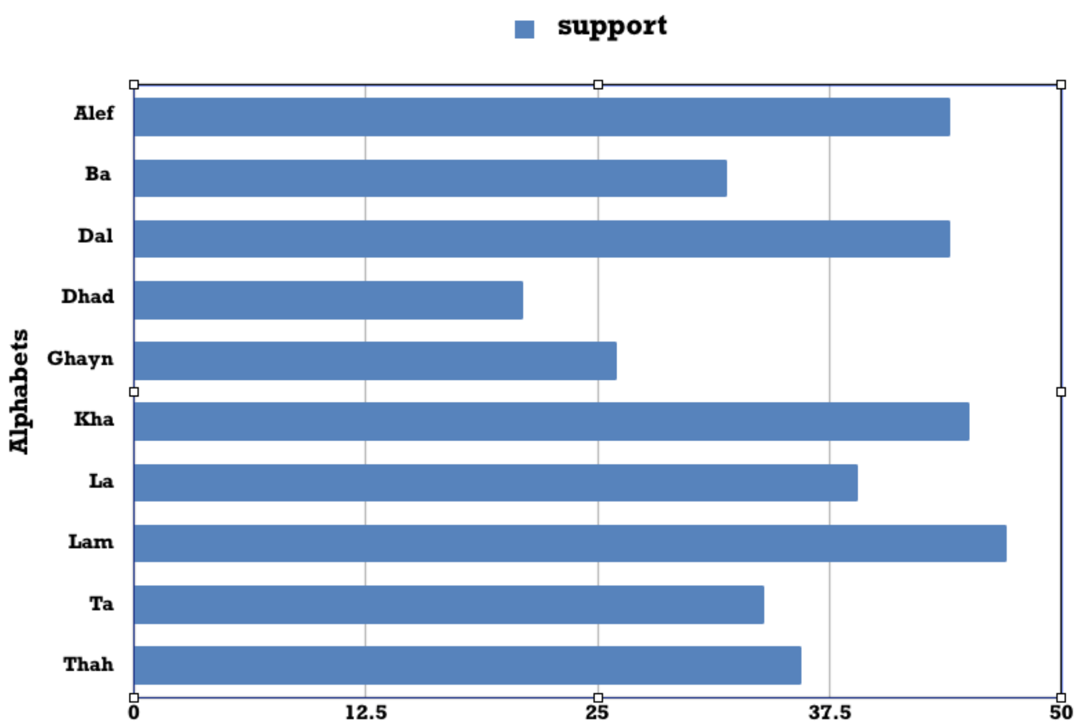


Figure 14 Support scores for recognized Alphabets

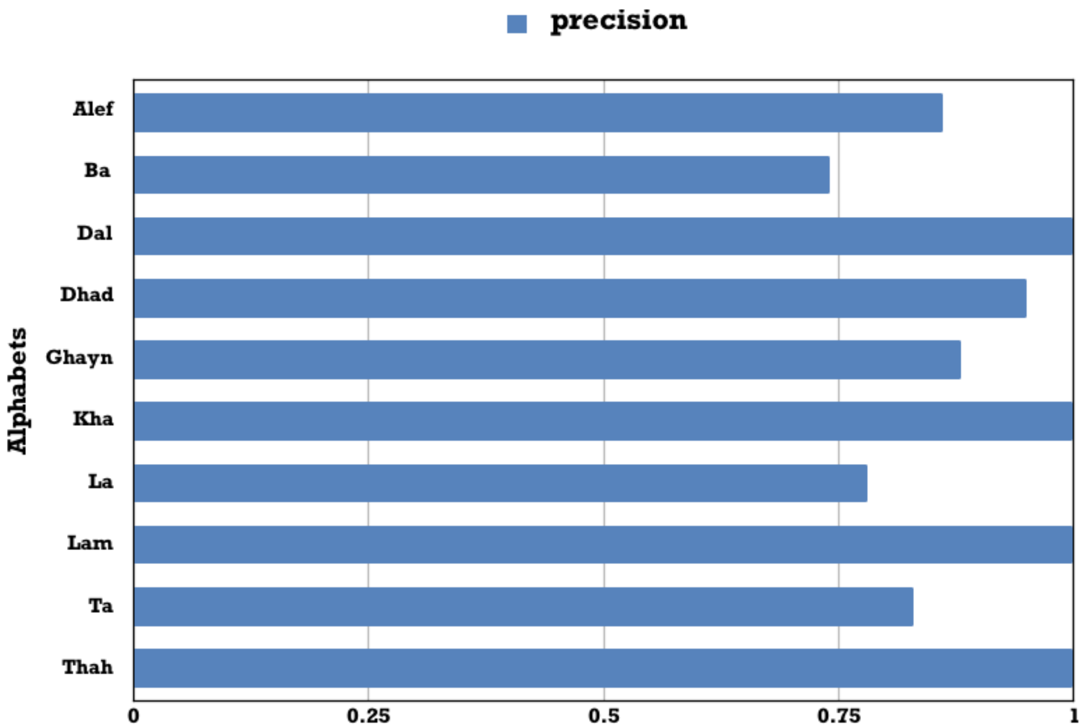


Figure 15 Precision scores for recognized Alphabets

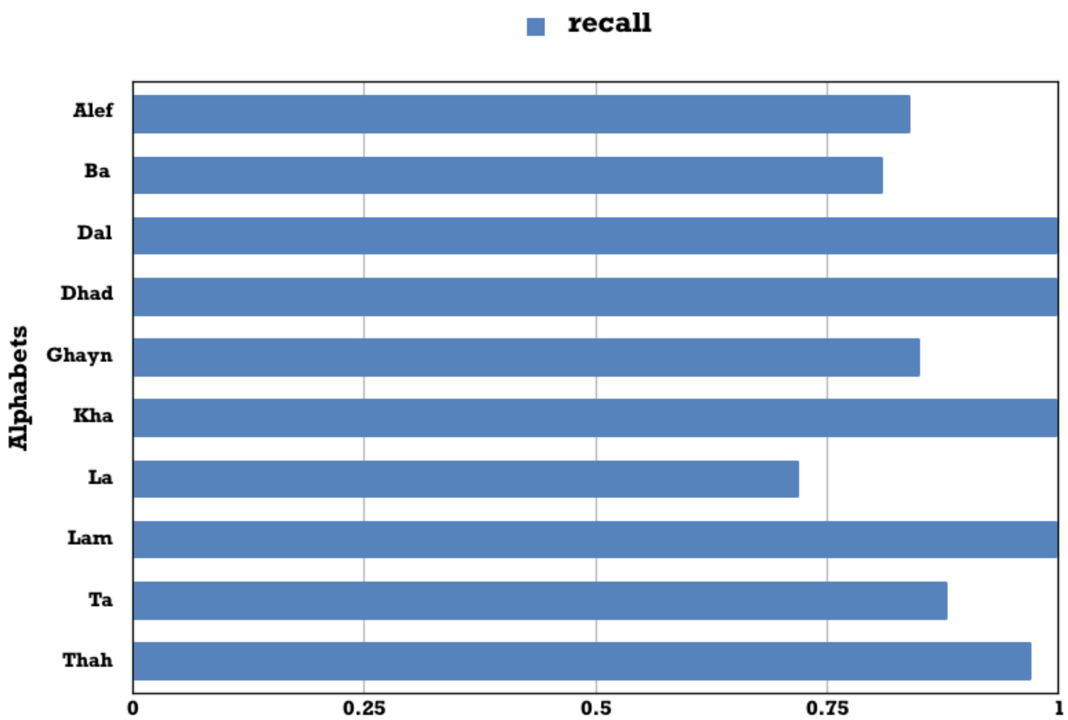


Figure 16 Recall scores for recognized Alphabets

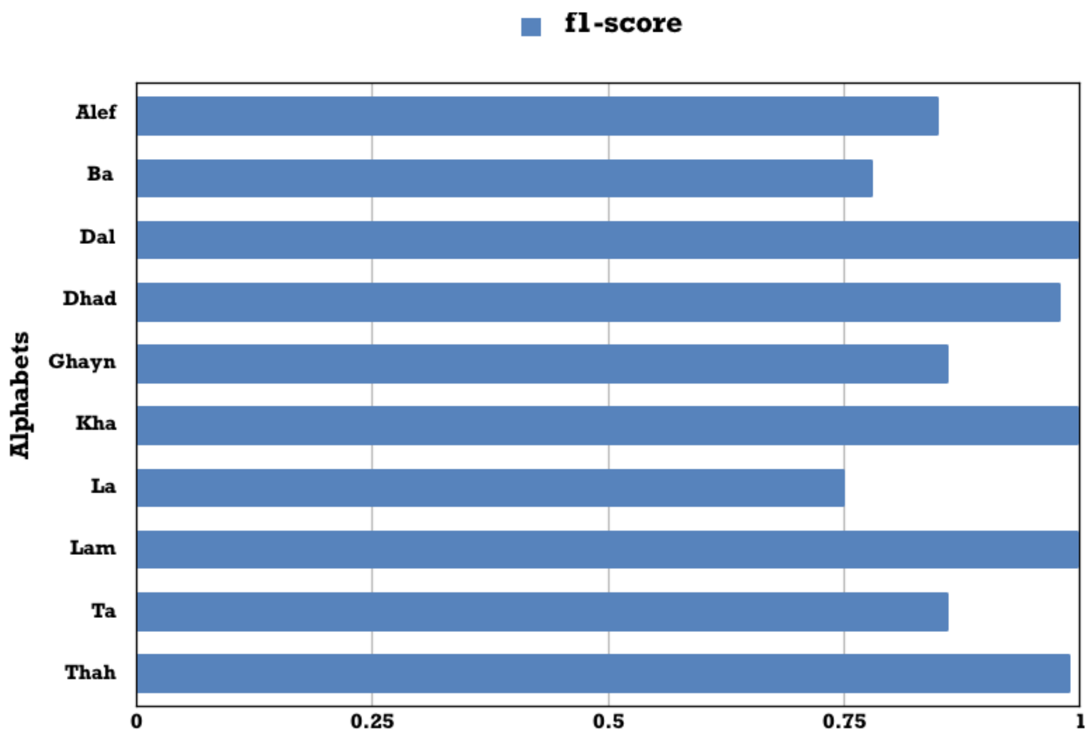


Figure 17 F1-scores for recognized Alphabets

CHAPTER 6

Conclusion and future work

6.1 Conclusion and future work

Many recent research on sign language recognition are using computers as their implementation platform as mentioned above, which are impractical to carry around for the purpose of gesture recognition because just like the natural language sign language as well is as casual and needed by the user at any point in time when he needs to communicate, the most effective way to address this problem is to introduce a more portable device as a platform, and mobile phone are so far the suitable devices to be used in such case due to their advantages over the computer platforms such as portability, availability, affordability and ease of use to mention but few.

We are proposing in this major project I an Arabic Sign recognition system on smartphone platform, previous research of sign language translation on smartphone have shown that the limitation in processing power in smartphone is one of its major constraints [6].To address this constrains we are proposing a client server system to take off the computational task off the smartphone and a server handles all the processing, all what the smartphone needs to do is to capture the image to be predicted and sends it over to the server for prediction and in turn the server replies with the predicted value.

We have extracted features from sign gesture images by normalizing and scaling the images to 30 x 30px and using the binary pixels as features, we then used SVM for Classification. Our approach is limited to 10 Arabic Sign language gestures only, the accuracy drops down when we include more gestures where the model tends to interchangeably misclassifies gestures which are visually similar but other gestures which do not look visually similar are more accurately classified, our future work will be the expansion of our model to recognize more alphabets and also to improve our algorithm to obtain higher accuracy and speed.

Appendix

Arabic Sign Language Recognized

References

[1] Munib. “Android sign language (ASL) recognition based on Hough transform and neural networks”, Expert Systems with Applications

[2] Davi Hirafuji Neiva, Cleber Zanchettin. “Gesture Recognition: A Review Focusing on Sign Language in a Mobile Context”, Expert Systems with Application.

[3] S. M. Halawani. "Arabic Sign Language Translation System on Mobile Devices", *International Journal of Computer Science and Network Security (IJCSNS),*

[4] M. Huenerfauth. "Generating American Sign Language Classifier Predicates For English-To ASL Machine Translation", PhD dissertation, University of Pennsylvania, Department of Computer and Information Science, Philadelphia, PA, USA,

[5] Cheok Ming Jin, Zaid Omar, Mohamed Hisham Jaward. “A Mobile Application of American Sign Language Translation via Image Processing Algorithms” 2016 IEEE Region 10 Symposium (TENSYMP), Bali, Indonesia.

[6] T.J. Joshi, S. Kumar, N.Z. Tarapore, and V. Mohile, "Static Hand Gesture Recognition using an Android Device," International Journal of Computer Applications.

[7] Setiawardhana, Rizky Yuniar Hakkun, Achmad Baharuddin “Sign Language Learning based on Android for Deaf and Speech Impaired People” International Electronics Symposium (IES).

[8] Mahesh M, Arvind Jayaprakash, Geetha M “Sign language translator for Mobile Platforms” IEEE .

[9] K. Sangeetha and B. K. L, "Gesture Detection for Deaf and Dumb People," *International Journal of Development Research.*

[10] Sakshi Lahoti, Shaily Kayal, Sakshi Kumbhare, Ishani Suradkar, Vikul Pawar.” Android based American Sign Language Recognition System with Skin Segmentation and SVM” 9th ICCCNT Bengaluru Bengaluru, India.

[11] www.rodrigo-rivera.com.

[12] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia. “Human Skin Detection Using RGB, HSV and YCbCr Color Models” ICCASP/ICMMD. Advances in Intelligent Systems

[13] jack sklansky, senior member,” Image Segmentation and Feature Extraction” IEEE transactions on systems, man, and cybernetics.

[14] Bauman, Dirksen (2008). Open your eyes: Deaf studies talking. University of Minnesota Press. ISBN 978-0-8166-4619-7.

[15] Nielsen, K.E.. (2012). A Disability History of the United States. Beacon Press. ISBN 9780807022047.

[16] "How Sign Language Works". Stuff You Should Know. 2014-02-06. Retrieved 2019-03-26.

[17] "SIL Electronic Survey Reports: Spanish Sign Language survey" (PDF). Sil.org. Archived (PDF) from the original on 2012-10-20. Retrieved 2012-09-30.

[18] Lucas, Ceil, Robert Bayley and Clayton Valli. 2001. Sociolinguistic Variation in American Sign Language. Washington, DC: Gallaudet University Press.

[19] Cf. Supalla, Ted & Rebecca Webb (1995). "The grammar of international sign: A new look at pidgin languages." In: Emmorey, Karen & Judy Reilly (eds). Language, gesture, and space. (International Conference on Theoretical Issues in Sign Language Research) Hillsdale, N.J.: Erlbaum, pp. 333–352; McKee R. & J. Napier J. (2002). "Interpreting in International Sign Pidgin: an analysis." Journal of Sign Language Linguistics5(1).

[20] Perlmutter, David M. "What is Sign Language?" (PDF). LSA. Archived (PDF) from the original on 12 April 2014. Retrieved 4 November 2013.

[21] Mayberry, Rachel. "The Critical Period for Language Acquisition and The Deaf Child's Language Comprehension: A Psycholinguistic Approach" (PDF). ACFOS. Archived(PDF) from the original on 2017-12-01.

[22] Huang, Jie; Zhou, Wengang; Zhang, Qilin; Li, Houqiang; Li, Weiping (2018-01-30). Video-based Sign Language Recognition without Temporal Segmentation (PDF). 32nd AAAI Conference on Artificial Intelligence (AAAI-18), Feb. 2-7, 2018, New Orleans, Louisiana, USA. arXiv:1801.10111. Archived (PDF) from the original on 2018-03-29.

[23] “Sign Language on Television". RNID. Archived from the original on 2009-04-17. Retrieved 2008-01-30.

[24] Bell Laboratories RECORD (1969) A collection of several articles on the AT&T Picturephone Archived 2012-06-23 at the Wayback Machine (then about to be released) Bell Laboratories, Pg.134–153 & 160–187, Volume 47, No. 5, May/June 1969.

[25] D.LU, Q.WENG “A survey of image classification methods and techniques for improving classification performance”, International Journal of Remote Sensing, Vol. 28, No. 5, 10 March 2007.

[26] Promila Haque, Badhon Das, Nazmun Nahar Kaspy “Two-Handed Bangla Sign Language Recognition Using Principal Component Analysis (PCA) And KNN Algorithm”, International Conference on Electrical, Computer and Communication Engineering (ECCE) , 2019.

List of Publication

1. Abbas Muhammad Zakariya, Rajni Jindal "***Arabic Sign Language Recognition System on SmartPhone***", 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India. IEEE 2019.