A Major Project II Report On

"Persian Sign Gesture Translation to English Spoken Language on Smartphone"

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF DEGREE OF

MASTER OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted By:
Muhammad Reza Jafari
2K18/CSE/24

Supervised by
Dr. Vinod Kumar (Associate Professor)
COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

June, 2020

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

DECLARATION

I, **Muhammad Reza Jafari**, Roll **Number 2k18/CSE/24** do hereby declare this Project Report titled "**Persian Sign Gesture Translation to English Spoken Language on Smartphone**" as my original work except where explicit citations have been made and that it has never been submitted to any Institution of higher learning for academic award.

Signed: ______________________________ Date: 22.06.2020

Muhammad Reza Jafari

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 110042

CERTIFICATE

This is to certify that the project Report titled "**Persian Sign Gesture Translation to English Spoken Language on Smartphone**" has been done under my supervision and is hereby being submitted for examination with my recommendation.

Signed Vinod Kumar 26/6/2020 Date

Dr. Vinod Kumar

Associate Professor

Department of Computer Science and Engineering

ACKNOWLEDGEMENT

The success of Major project requires help and contribution from numerous individuals and the organization. Writing the report of this project work gives me an opportunity to express my gratitude to everyone who has helped in shaping up the outcome of the project. I express my heartfelt gratitude to my project guide Associate Professor Dr. Vinod Kumar for giving me an opportunity to do my Major project work under his guidance. His constant support and encouragement has made me realize that it is the process of learning which weighs more than the end result.

Muhammad Reza Jafari
Roll No. 2K18/CSE/24

Abstract

Hearing impaired and others with verbal challenges face difficulty to communicate with society; Sign Language represents their communication such as numbers or phrases. The communication becomes a challenge with people from other countries using different languages. Additionally, the sign language is different from one country to another. That is, learning one sign language doesn't mean learning all sign languages. To translate a word from sign language to a spoken language is a challenge and to change a particular word from that language to another language is even a bigger challenge. In such cases, there is necessity for 2 interpreters: One from sign language to the source-spoken language and one from the source language to the target language.

There is ample research done on sign recognition, yet this paper focuses on translating gestures from one language to another. In this study, a smartphone approach is proposed for Sign Language recognition, because smartphones are available worldwide. Smartphones are limited in computational power so, a client server application is proposed where most of processing tasks are done on the server side. In client-server application system, client could be a smartphone application that captures images of sign gestures to be recognized and sent to a server. In turn, the server processes the data and returns the translation Sign to client. On the server application side, where most of the sign recognition tasks take place, background of the sign image is deleted, and under Hue, Saturation, Value (HSV) color space is set to black. The sign gesture then separate by detecting the biggest linked constituent in the frame. Extracted feature are in binary form pixels, and Convolutional Neural Network (CNN) is used to classify sign images. After classification, the letter for a given sign is assigned, and by putting the sequence of letters, a word is created. The word translates to target language, in this case English, and the result returns to client application.

Keywords

Sign Language, Gesture Recognition, Computer Vision, Image Processing, Machine Learning, hearing-impaired people, Convolutional Neural Network (CNN).

Table of Contents

List of Tables

Table of Figures

LIST OF ABBREVIATIONS

PSL	Persian Sign Language
CNN	Convolutional Neural Network
ASL	American Sign Language
ArSL	Arabic Sign Language
ANN	Artificial Neural Network
DT	Decision Tree
HSV	Hue Saturation Value
BSL	British Sign Language
LSF	French Sign Language
LSE	Spanish Sign Language
HCI	Human Computer Interaction
RGB	Red Green and Blue
SURF	Speeded Up Robust Features
HOG	Histogram of Oriented Gradients
API	Application Programming Interface

CHAPTER 1

INTRODUCTION

Introduction

A means of communication among people with verbal disability is sign language; sign language is used to represent what they want to share with one another, such as: words, phrases or numbers. Additionally, face and body gestures are also used. Hearing-impaired people have difficulty communicating because most people don’t understand sing language and as sign gesture is different per nation [1] translating them from people is even more difficult. Addressing this gap, captures interest among researchers.

Much research has been done on sign language using different systems and devices, such as gloves with sensors [2], and more complex systems with cameras and Kinect Devices to help record acceleration movement [25]. Majority of the research implemented, used computer as main platform, which is not practical to carry around. The smartphones are the effective devices to address this gap; it can address the challenge of communication between people and verbally challenged people. In this Major project II, we are proposing a system for Persian Sign Language (PSL) base on smartphone platform that can recognize Persian Sign language and create a word and translate the word to English word.

Like natural language, gesture based communication also differs per nation or location. Considering this difference Persian language is no exception. Figure 1 shows the Persian sign Language Alphabets. Gestures communication in Persian, started by Jolia Samiee in 1987 which proposed the first Persian Sign Alphabet which contains 33 Alphabet, in 1994 second Persian Sign Alphabet proposed which contain 38 characters and finally in 1996 the Persian Sign Alphabet finalized with 37 characters.[26]

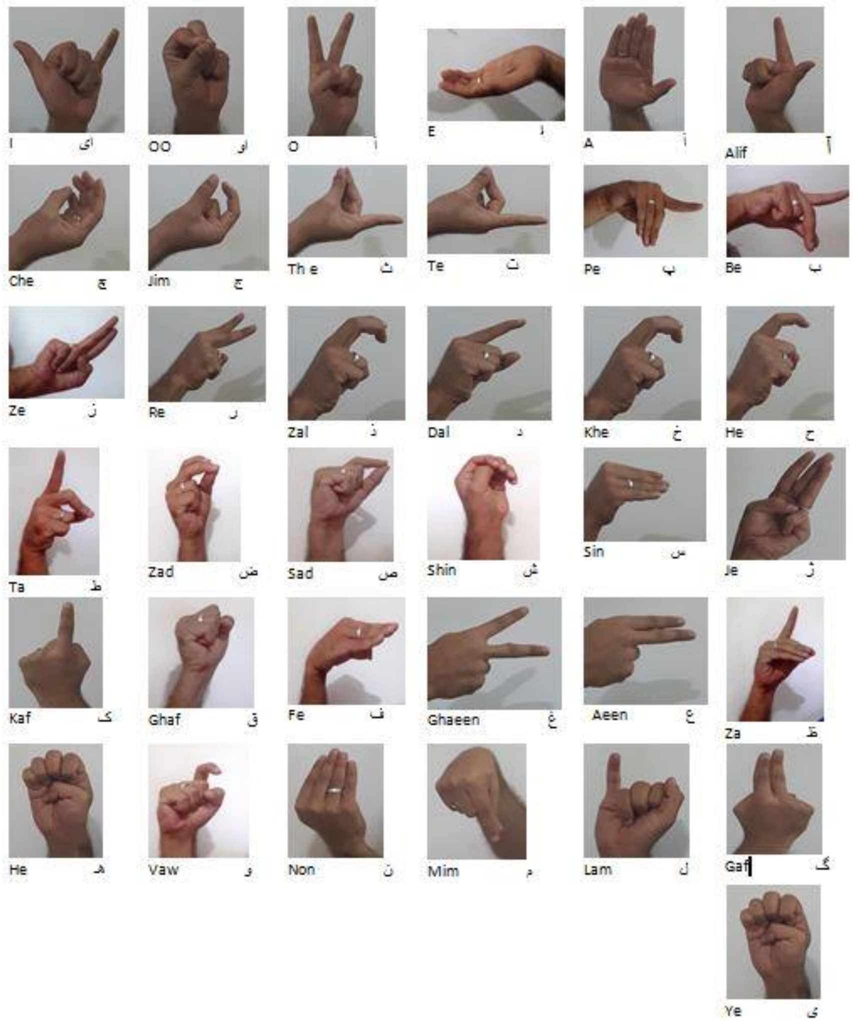


Figure 1 - Persian Sign Language Alphabets

Gesture recognition applications include different stages such as: segmentation, feature extraction and classification. The aim of segmentation is to eliminate the noise in the background, leaving only Region of Interest (ROI), which is the main useful data in the picture. In feature extraction stage, ROI is the focused area to extract from image. These highlights can be edges, shapes, flows, corners, colors, surfaces, etc. The highlight of pictures basically is personality of each communication via gestures motion. In the next stage, the features experience classification is completed, which is used to train the system, and deicide to which group of gestures, the new sign belongs. [4]

There is much work done on sign recognition system with computer platform; however, very limited research has been done on cell phones platforms. The past research based on cell phones have shown big disadvantage in computation and resource limitation. [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

Deaf communities use sign language to communicate through history. One of the earliest sign language found on fifth century by BC from plato's Cratylus in which Socrate says: "If we hadn't a voice or a tongue, and wanted to express thing 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]. The most known system for sign language is fingerspelling till 19th century, which were used to translate words from spoken language to sign language, the first manual alphabet for sign language was developed by Leon (1520-1584). [15]

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

The meantime in 1755, Abbé de l'Épée has founded a school especially for deaf children in France precisely in Paris, Gallaudet University which is the only deaf people's Liberal Arts University in the world, it was founded by Edward Miner Gallaudet in 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 graduates of the Deaf School for children in Paris which was founded by Abbé de l'Épée. [16]

Sign language has generally no relation to its corresponding spoken language but rather there is complex correlation between them and it varies much to the country in which the sign language is used, for examples on this type of scenarios we could observe that American Sign Language (ASL) is basically has been driven from French Sign Language but it is used in America and English-speaking Canada, like whiles Australia, New Zealand and 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, 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 region could be correlated to be the location of Deaf schools in the geographical area not necessarily corresponding to the different number of dialects used in the country. [17][18]

Term "International Sign" is given by International organization and World Federation of Deaf, which is sign language that is used mostly in international events for Deaf like the Deaflympics and also used at the World Federation of Deaf meetings, International Sign also called Gestuno. [19]

2.1.2 Relationship with spoken languages

Ideally people think sign language is visual alphabetic of spoken language, where people with expressing of spoken language in signs which is not right. Misconceptions that sign language are invented by people who have no hearing disability. De l'Épée, Charles-Michel and even Thomas Hopkins Gallaudet teachers of deaf school with no hearing disabilities are incorrectly referred to as inventors of sign languages which is not correct because sign language basically invented by the deaf people themselves who don't or have very less knowledge of spoken languages.

It is obvious spoken language improved by borrowing elements from other 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 language 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 spoken word will be represented by hand shape where it carries same meaning of spoken word.

2.1.3 Acquisition

Children with no hearing disability acquire their native spoken language automatically as they grow, similarly deaf children acquire sign language as they grow up without any problem because of brain plasticity of children and because of that children learn easier that an adult.

The study from University of McGill done regarding who natively acquire sign language and people who learned it. It shows that American Sign Language (ASL) users that learned the language natively have much better performance than those who did learned the language when a task was given to them to represent the video sentence on American Sign Language (ASL), as well the study found that there is different between the two groups in their ASL sentences grammatical morphology, so the study conclude there is a very special and critical period on learning or acquiring sign language. [21]

2.1.4 Interpretation

Translation of sign language to spoken language is done by sign language interpreters, sign language interpreters are used to facilitate the communication among people with verbal disability and the hearing people, sign language have their syntax not similar to spoken language so that will require a considerably higher amount of effort from interpreter's part.

Most of time interpretation sign language and spoken language are been used in same region or country, for example the France spoken French and French Sign Language (LSF), spoken English the United Kingdom and British Sign Language (BSL), The United State spoken English use American Sign Language (ASL), The Spanish Spoken use Spanish Sign Language (LSE), etc. There are interpreters which can translate between non paired sign language and spoken language for example American Sign Language (ASL) and French Sign Language (LSF).

Recently there have been improvement in computer science and artificial intelligence (AI) especially in Computer Vision, algorithms development using deep learning can give ability for machine translation that can automatically translate a short video of sign language to written text which could be some few sentences. [22]

2.1.5 Remote Interpreting

In translation process, ideally an interpreter is supposed to be physically available to do the interpretation between two parties, but after the 2000s there are revolutionary improvement in technology and communication, and it is become easy to provide remote interpretation. Where interpreter can be in different location geographically from other parties (sign gesturer and people who need to be translated to). The technology called Video Remote Interpreter (VRI), the first party which is the sign gesturer and second party which the person to be translated to could be in different location or in same location and the interpreter which is in different location, by help of video communication media between interpreter and first party and audio communication like between interpreter and second party translation can be done.

VRI is used when there is no available interpreter on the site, which is a situation VRI could be used, where two parties are not in same location and need to communication to one another, so in such case all the three parties are in different location the interpreter is used to make available communication between two parties using VRI.

2.1.6 Interpretation on television

In TV programs sign language is provided at corner bottom of TV screen, while the TV program show on full TV screen. For example when a president of Afghanistan appear on the TV in press conference you will see a sign language translator is provided on the corner of screen.

Presenting the sign translation to TV program tend to cause some distraction to some of the people who watch the TV program, for that reason some program are repeated two times one time without sign translation another time with sign translation. For example majority of BBC program in early morning and late nights are with sign translation, also there are some television technologies that can provide the option to show or hide sign translation or even show or hide subtitle for TV programs. [23]

2.1.7 Telecommunications

By means of telecommunication from some decades ago in 1964 New York World's Fair the first communication remotely between two deaf was made each one of them in different city, The company AT&T introduced their videophone which had trademark of "Picturephone". However, due to the bandwidth limitation, video communication was not possible to make till the 2000s when sufficient bandwidth was available.

Now a days Internet provides variety and wide range of communication services and platform which is possible to be used by deaf people to communicate among themselves. Some platform has features especially for deaf people to empower them, for example ability to zoom in to video in order to concentrate on the sign gestures. [24]

2.2 Skin Segmentation

Skin segmentation is the process of recognizing skin-colored picture and locales in a picture or a video. This procedure is mostly used to find area that conceivably has human appearance and appendages in pictures. Skin picture recognition is utilized in vast variety of picture handling applications like face recognition, skin disease recognition, HCI (Human Computer Interaction). The vital key for skin recognition is the skin coloring, although color cannot be the main integral factor because of the variety in skin tone as indicated by various races, for example the light conditions likewise influence the outcomes. Subsequently the skin tone 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 non-colored. One of the basic techniques is to check if each skin pixel falls into characterized color extent or qualities in a few directions of a color space. Skin color into a characterized color extent or qualities in a few direction 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].

2.2.1 The Color Space

The color data showed by color space model into three or four different color constitution. Different application is used different color space like image processing, computer graphics, TV broadcasting, and computer vision. For skin detection, different color space is used. They are: Luminance based color space (YCBCr, YIQ, and YUV), Hue Based color space (HSI, HSV, and HSL), RGB (RGB, normalized RGB),

These models gave more detail in the following sections. 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].

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

RGB color space is generally 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 utilized by PCs, designs card and LCDs or screens. As in Figure 2 underline 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, different color can be made. Switching the scenario, an explicit color can be divided into its red, green, blue segments. These quantities can be utilized to discover comparative color pixels from the picture. Standardized RGB is a demonstration that is easily acquired from the RGB esteems by straightforward standardization strategy [12].

$$r = \frac{R}{R + G + B}$$

$$g = \frac{G}{R + G + B}$$

$$b = \frac{B}{R + G + B}$$

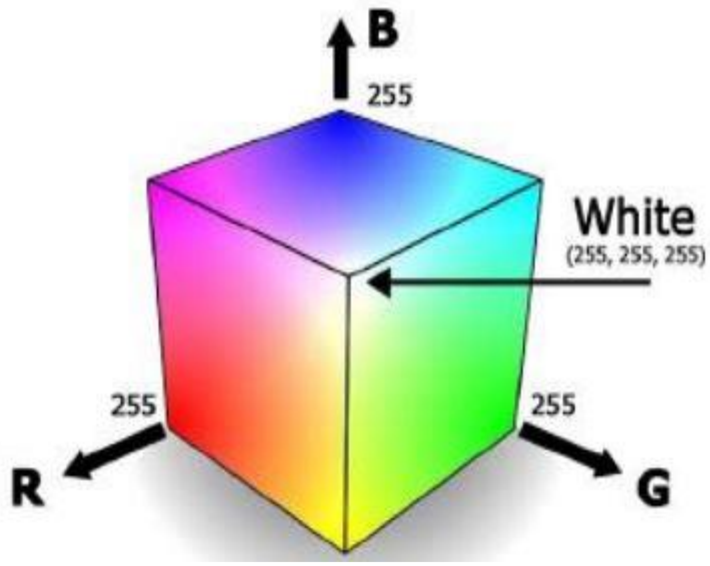


Figure 2 - RGB Color Model

2.2.1.2 YCbCr (Luminance, Chrominance) Color Model

For picture compression European TV studios utilized an encoded non-linear RGB called YCbCr. As appeared in Figure 2 below from [12], color is shown by luma (Which is luminance found from non-linear RGB) developed as a weighted total of RGB values [12]. YCbCr is a normally utilized color space in advanced video area.

Since the portrayal make 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 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 reference esteem.

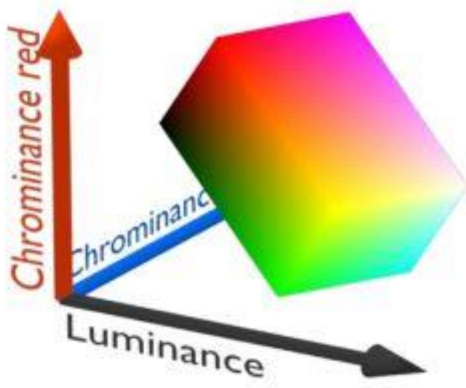


Figure 3 - YCbCr Color Model

Y = 0.299R + 0.287G + 0.11B

Cr = R – Y

Cb = B - Y

2.2.1.3 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) differ 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 00 to 3600 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 (300, √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.

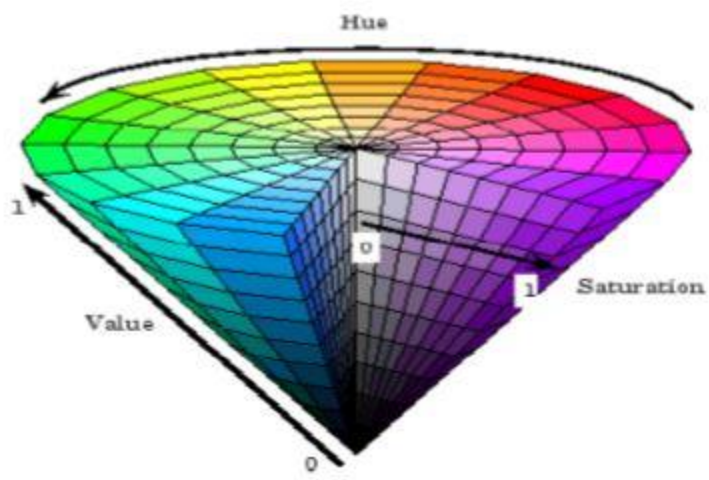


Figure 4 - HSV Color Model

2.3 Feature Extraction

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

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 converting the dimensions higher than ever to encourage the showcase of the picture. The last change result in another histogram a histogram of the change dim dimensions that may enhance the PC's capacity to do edge or limit location and extraction of textual features.

Consider "g" as an random gray level, and "b" as an random interval width. Generally "b" is limited to small values with respect to "g". Consider "N" as the number of pixels whose gray values exist between (g – (b/2), g + (b/2)) from just a few connected sets of points or "components" in the x- axis [13].

N-pixel slices of p(x) are components that determined the segments of p(x). Thinks about a set of N-pixel slices of p(x), in textual picture which consist of big regions of closely uniform textures, the shapes, distribution of components, and sizes between these slices are similar. With this constraint a measure of visibility of details among these slices is 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.

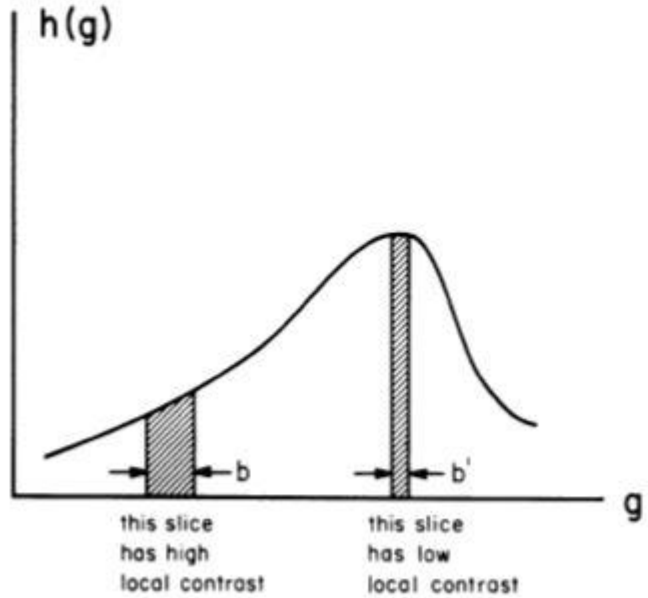


Figure 5 - Relation of local contrast to h(g)

If "b" is small, approximation of histogram h(g) considered as continuous function; "b" and h(g) are related as follows:

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

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 h(g) is shown on Figure 5. [13].

2.4 Classification

Several factor could be affect the classification of image, it is a very complex process, for very a good and successful classification sufficient data to 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 and drawbacks over others. There will be different classification methods, on each classification method selected and used, talking of classification methods many classification approaches are used some include ANN, fuzzy, deep learning and expert system. [25]

Category lists of classification approaches with their examples are as bellow:

1. **Supervised classification approaches:** Decision Tree classifier, ANN, Minimum Distance and Maximum Likelihood.
2. **Unsupervised classification approaches:** K-means Clustering and ISODATA
3. **Pre-pixel Classifier:** SVM, ANN, Minimum Distance and Maximum Likelihood, Decision Tree.
4. **Sub-pixel Classifiers:** Spectral mixture analysis, Fuzzy-set and sub-pixel classifier
5. **Parametric Classifiers:** Linear discriminant analysis and Maximum Likelihood.
6. **Non-parametric 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 sub-pixel classifier.
13. **Spectral Classifiers:** ANN, Minimum Distance and Maximum Likelihood.
14. **Deep learning:** CNN, RNN, Classic Neural Network

Chapter 3

Related Work

3.1 Related research reviewed

Sign gesture is used as part of communication medium in human life. The usage of signs or movements of body are not depend on age, sexual orientation, or ethnicity[3]. Many researchers used different approaches in sign language recognition.

Paper [8], introduces a system that uses mobile phone to recognize American Sign Language (ASL), for skin picture segmentation they have used YCbCr systems. First the system catches the image by an android mobile phone. Histogram of Oriented Gradients (HOG) helps extract features of the image to classify using Support Vector Machine (SVM). Their approach achieve accuracy of 89.54%.

In paper [7], sign language detection based on an android system. For hand detection using OpenCV. For classification K-NN is used, the system detected up to 50cm away from palm of a hand for recognition of gesture.

S. M. Halawani [3] has proposed Arabic Sign Language Translation System (ArSL-TS). His model uses a smartphone to translate Arabic text into Arabic Sign language.

In paper [4], a method for better segmentation can recognize 32 Persian static sign gestures. Their method is used YCbCr color space, sign Gaussian model and Bayes rule. In order to recognize the sign gestures by help of radial distance and Fourier Transform sign gesture extraction, and by help of Euclidean distance to find similarity between hand gesture and training database. The accuracy of the system is 95.62%.

Cheok Ming Jin [5], proposed a smartphone platform for ASL (American Sign Language) detection. He implements Canny edge detection plus seeded region growing in order to segment hand gesture in the picture, for extraction of feature Speeded Up Robust Features (SURF) algorithm and for classification SVM is used. The accuracy of the system for 16 class of ASL is 97.13%.

In [6], a static Persian Sign Language gesture for recognizing some word is presented. It use a digital camera for taking picture input, the system used Discrete Wavelet Transform and Neural Network for feature extraction and classification. The classification accuracy is 98.75%.

[9] This paper present a system to recognize Persian static sign gesture, a digital camera is used for taking input pictures, and feature extraction and classification used Wavelet Transform and Neural Network, the accuracy of the system is 94.06%.

Sakshi Lahoti [10], proposed a smartphone approach in order to recognize the American Sign Language (ASL), YCbCr system used for segmentation of skin in pictures captured by smartphone, for feature extraction they use HOG, and finally for classification SVM is used, the accuracy of system is 89.54%.

Promila Haque [11], proposed two-Hand Bangla Sign the system can recognize 26 sign gestures. Principal Component Analysis (PCA) is used to extract image principle component and for classification K- Nearest Neighbors is used. He used 104 images for testing and achieved Precision 98.11%, Recall 98.11% and F-Measure 90.46%.

In [20], a comparison classification between CNN and SVM show, CNN has better performance compared to SVM. The accuracy of CNN is 90%.

In [27], as shown in the research experiment the CNN improves the performance of classification.

Abbas Muhammad Zakariya, [28] proposed Arabic Sign Language (ArSL) recognition; based on client server approach which client is a smartphone. They use HSV color space for background elimination and SVM for classification and achieved accuracy of 92.5%.

.

Chapter 4

Current Work and preliminary Results

4.1 Current Work and preliminary Results

Gloves and embedded sensors for tracking hand gesture allocated most of Sign language recognition researches to itself, but these gloves cannot use in humidity and rain, and they are not portable because whenever you need to use them you need a computer as well.

With development of smartphones technology and improvement of computational capacity of them, sign language recognition system is easier to apply by smartphones, and the challenge of portability would solve. Some researcher proposed system used smartphones like [10] but still smartphone suffer from computational power in [28] they proposed a client sever approach which is much better in term of computational power of smartphones. In this paper we propose a client-server approach which addresses these disadvantages and using CNN classifier as shown better performance in [20][27].

In this section explain a related work in [10] and [28], their work as an implementation of American Sign Language recognition with an android app.

Dataset:

- 36 symbols used by American Sign Language (ASL)
- Number from 0 to 9, letters from A to Y, and spacebar.
- Letter Z is not in symbols because the system process static gesture sign.
- Each symbol, 500 training images
- Size 200X200 pixels
- Black background

Processes are mentioned bellow:

1. Capture the hand gesture and segmentation of skin
2. Feature Extraction
3. For Classification SVM is used

A. Capture the hand gesture and segmentation of skin
   The gestures are caught utilizing smartphones camera. The camera is begun and video taken is partitioned into littler frames with the goal that hand gesture is appropriately recognized.
   - Y is the luma (image brightness in a video) components,
   - Cb is blue difference
   - Cr is red difference
   - Conversion formula (RGB to YCbCr) is:

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

   - If pixel value is between 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)

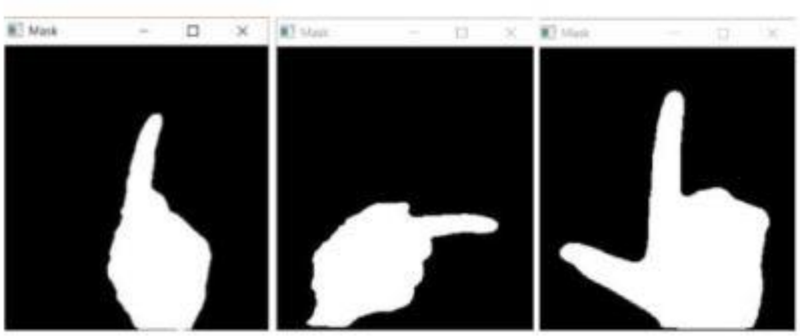


Figure 6 – Segmentation of picture by Skin Detection

B. Feature Extraction

After skin process segmentation has ended, the Histogram of Oriented Gradients (HOG) is used for extraction of feature from the image. The number of gradient that is raised in neighborhood of segments in a picture has counted by HOG. HOG first downsize the image into smaller block of pixels. Histogram of each cell is calculated, and at the end HOG build global histogram by scanning the all the cell that is generated. The benefit of HOG over other method (SIFT and other similar) feature extraction is that geometric transformation did not affected it. Algorithm of HOG is as follow:

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 Support Vector Machines (SVMs).

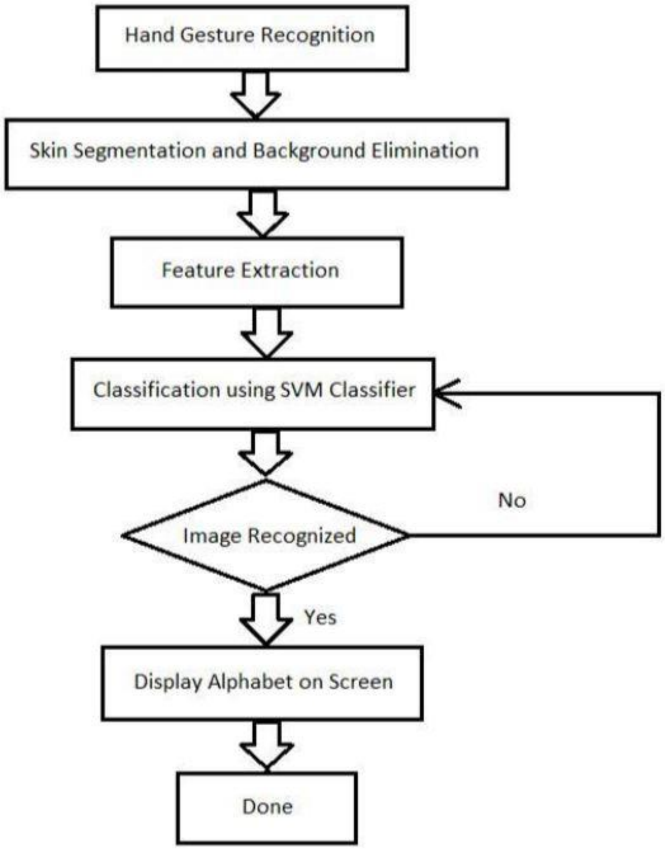


Figure 7 - Diagram of Data Flow

C. Classification using SVM

New images is classified to its class by Support Vector Machine (SVM), it is a non-probabilistic linear method. Datasets are shown by points on the hyperplane in SVM. Training dataset is segmented by gap in order to make more simple classification. Tag of classes is based on relation to the gap. A kernel function is defined by SVM in order to make such a task effective. “Soft margin function” concept that denoted by C is used for better accuracy by SVM, that control 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 vector 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 high bias.

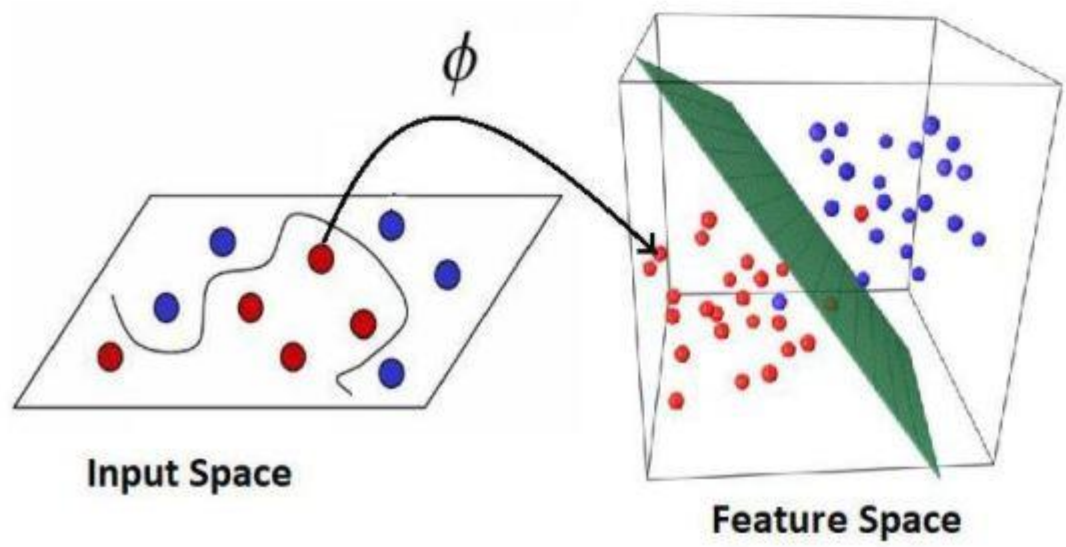


Figure 8 - Hyperplane is used to classify the input

The green colored page is the hyperplane 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.

SVM recognize the segmented alphabets as shown in figures bellow. The output alphabets in red are prediction of SVM in figures.

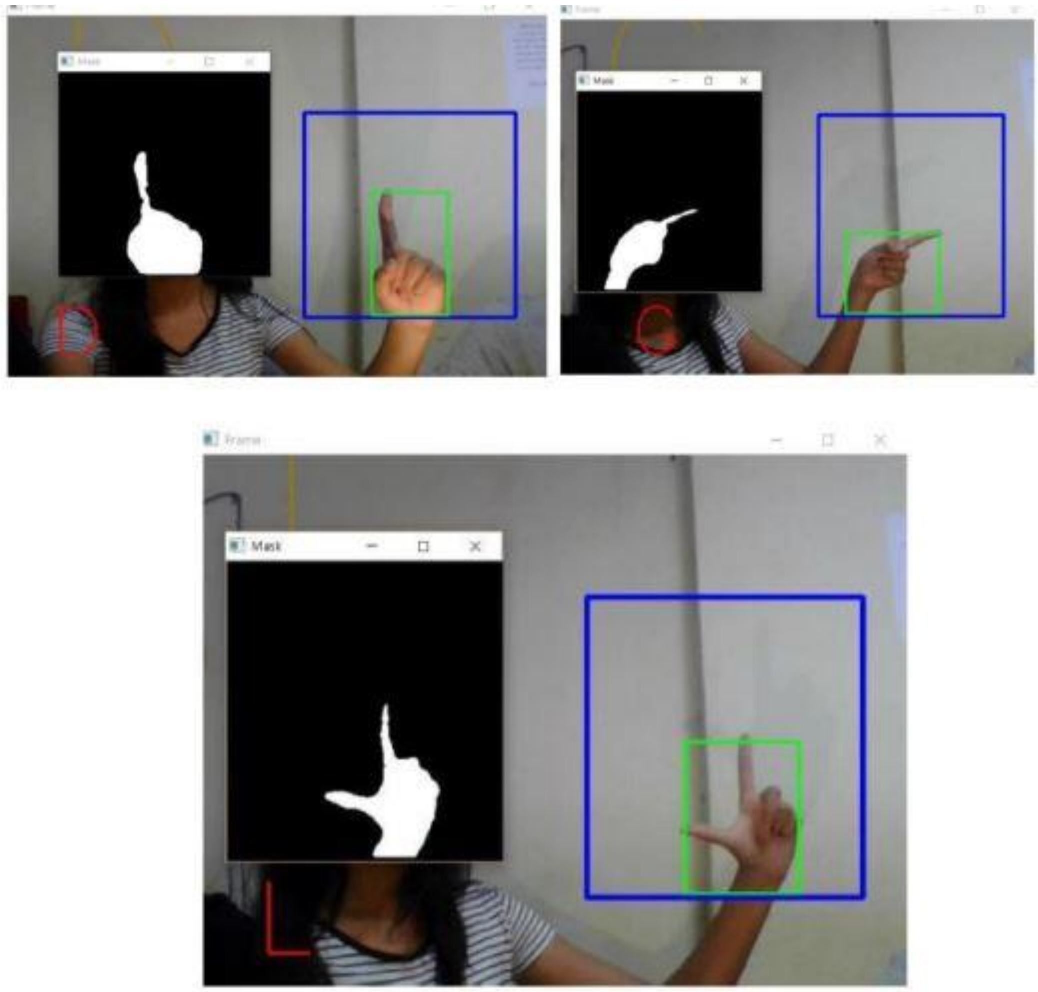


Figure 9 - Recognized alphabets of the above segmentation image

After finishing the process of training and classification the datasets are uploaded in could, so smartphones resources would be saved and drawback of using smartphones platform solved and the datasets are available for further use.

To overcome this drawback, in this paper a client-server recognition system is proposed to overcome low computational power of smartphone that all process handles by server. Smartphone catch the picture as input. The input pictures send for classification to server. The server answers the call by sending back the result of prediction.

Support Vector Machine (SVM) has accomplished many promising result 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 with accuracy 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 Persian Sign Gesture Translation to English on Smartphone:

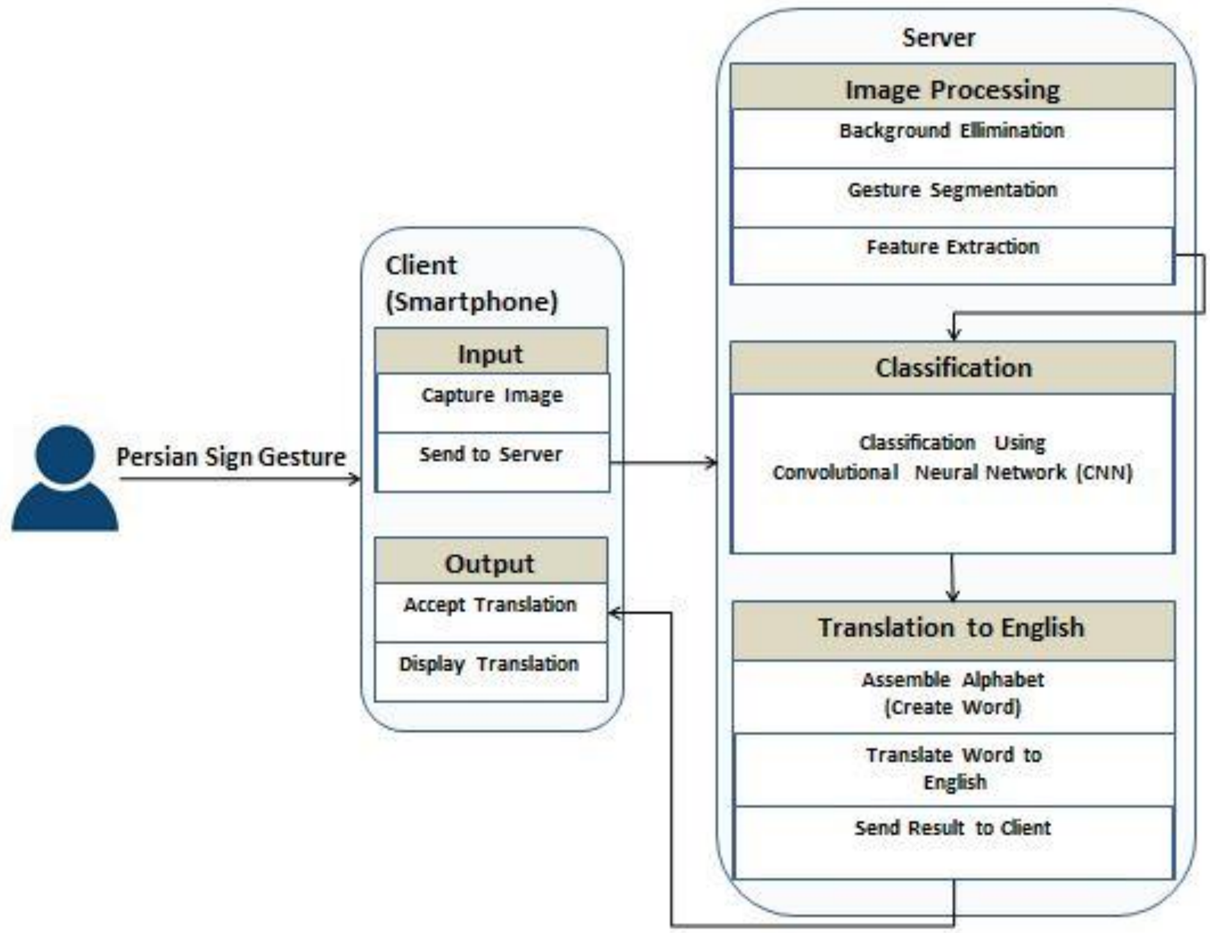


Figure 10 - Machine Translation from Persian Sign to English Spoken Language on Smartphones

4.2 System Overview

A client server recognition system is developed as shown in figure 10, smartphone is used as client; user in client side could interact directly with smartphone application. Sign gesture picture capture by android application as input and send it through Application Programming Interface

(API) to server. On the other hand, on server side receives the picture from the client-server. After predicting and translating the text from sign gesture, server API sends it back to the client API responsible to show the text on the screen of smartphone.

Client has two main responsibilities: first captures the sign gesture image as input for server and displays the prediction text on the screen of smartphone, while server has three main responsibilities: first preprocesses the input images, second, classifies and finally translates the predicted text to the target language.

4.3 Smartphone Application

A smartphone application is used by name of IP Webcam that could access camera of smartphone and send it to server in other to further process, and after processing the images in server the result which is the Persian word and its translation to English would save in a MySQL server. The client (smartphone) could access the result by clicking to the link and a query will execute to fetch the result from MySQL server and show it to client. Some screenshots from the mobile application show in figure 11.

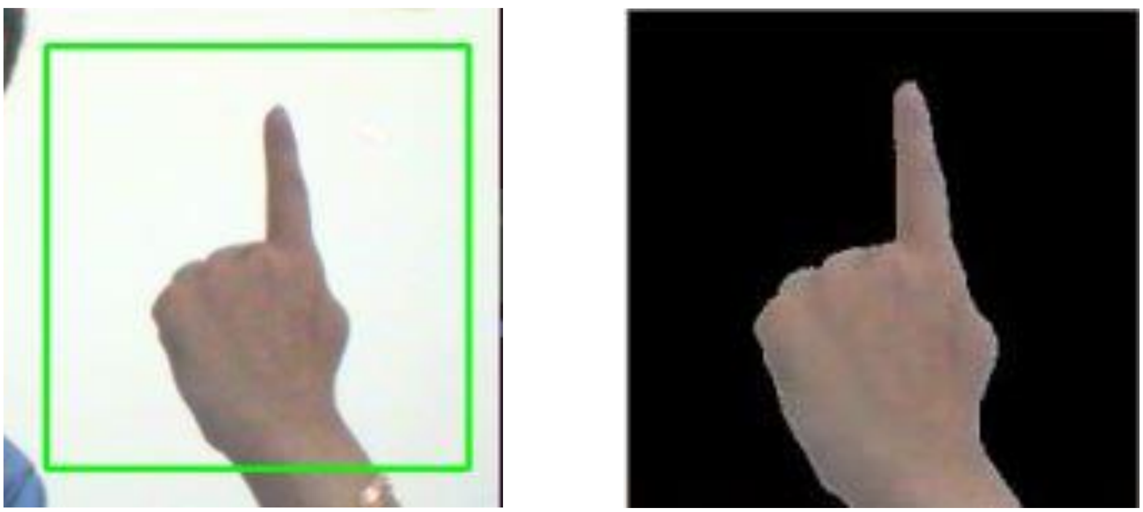
(a) (b) (c)

Figure 11 (a) IP WebCam, (b) MySQL database, (c) Result of translation

4.4 Background Elimination

Images made of pixels; each pixel is shown by number between 0-255 in RGB color space which is changeable to any other color space.

The background of input picture which was sent from smartphone is discovered and it change to black. The picture transforms from one color space to another in this case from RGB to HSV, so the skin color is detected and series of dilation and erosion using elliptical kernel is made. The final frame is created by combining effect of two masks as shown in figure 12.



(a) (b)

Figure 12 - (a) input Picture, (b) Picture after setting background to black

4.5 Gesture Segmentation

The picture from previous process where the background was changed to black is first converted into grayscale, although the color of original picture is lost, this process increase the robustness of system to variety of lighting conditions. Pixels that are not black binarize (change to white) and remaining pixels is black. Now the segmentation of sign gesture is begun. First only the largest component in picture detected which is sign gesture and rest of the component delete. Resize the image into 64*64 pixels. The whole process is shown in figure 13.

(a)

(b)

(c)

(d)

Figure 13 - (a) Input picture, (b)Background Elimination, (c) Binarized, (d) Segment sign gesture and resize

4.6 Feature Extraction

The sign gesture pictures are normalized and resized to 64*64px. Our features are binary pixels that are used for other steps. Resizing the picture to 64px is helping us to reduce the number of features and reduce the load of work for server and also it has good enough features in order to use for classification without losing the important feature. The number of features in 64*64px is 4096.

4.7 Classification

For classification of our sign gesture we use Convolutional Neural Network (CNN). CNN is used to classify the extracted feature from previews steps. CNN is a multilayered neural network; it has especial architecture in order to discover complex features in data. It's mostly used in image recognition, powering vision in robots and self-driving vehicles.

CNN has three main steps: 1. Convolution, 2. ReLu, 3. Pooling, 4. Flattening, 5. Full connection. A convolution is a combination of integration of two functions that show you how one function modifies the other. ReLu, rectifier function is used to increase non-linearity in CNN. Images made of variety of objects that are not linear to one another. Image classification would be linear problem without applying this function which it is actually a nonlinear one. Pooling enables the CNN to identify features in different images, irrespective of different in light condition in images and angles. In this paper Max pooling is used to preserve the main features while also reducing the size of the image. Flattening, after pool features achieved, flattening that transforming the whole pooled features map matrix to a single column which is used to fed to neural network phase. Full Connection, after flattened feature map is done it goes through neural network. This has input layer, the fully connected layer, and the output layer. Fully connected layer is similar to Hidden layer in ANNs and output layer is used to predict the classes. (The information go through network and prediction of error is calculated, the error is then back propagated through the system to improve the prediction) [29].

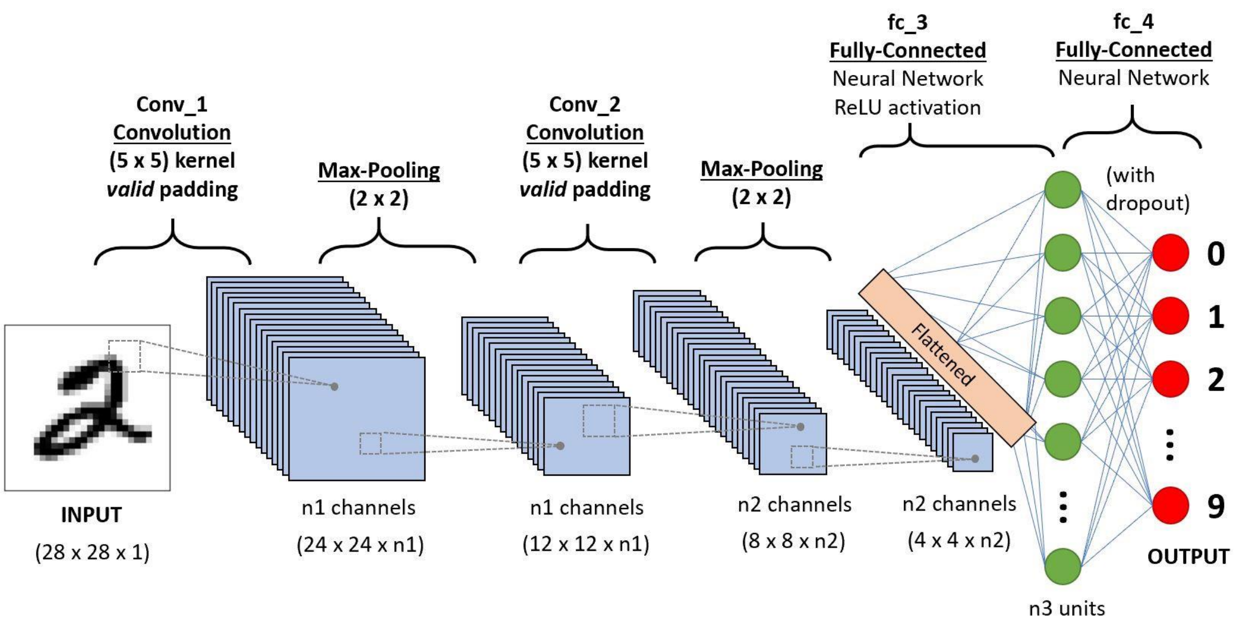


Figure 14 - CNN Algorithm

4.8 Translation to English

The words recognized by CNN classifier are placed in an array in order to construct a word. Then by using a bilingual dictionary, in this case Google-translate, library of Python, translating a Persian word to an English word. The word that is translated is sent to the client (Smartphone), as shown in figure 11.

The word which is used to translate is combination of 10 Alphabetic letters : A(آ), B (ب), C (س), D (د), Gh (غ), K (ک), N (ن), O (أ), T (ت), and Y (ی).

Chapter 5

Experimental Results

5.1 Result and Analysis

For evaluation we used 10 Persian Sign Language. These Alphabetic sign language are as follow: A(آ), B (ب), C (س), D (د), Gh (غ), K (ک), N (ن), O (أ), T (ت), and Y (ی). In other to evaluate we used 2000 pictures in total to train the Convolutional Neural Network (CNN) classifier. For evaluation system performance, we split our data images to 20% testing and 80% training. The accuracy which achieved by CNN is 98%. Precision, Recall and F-Measure for each sign gesture shown in Table 1.

The output of CNN classifier assemble and create a Persian word which by help of bilingual dictionary translate to English word, and English word is sent to client (Smartphone), inorder to show to user.

Table 1 Precision, Recall, F-Measure

Letter	Precision	Recall	F-Measure	Support
A	0.98	1.00	0.99	40
B	1.00	1.00	1.00	40
C	1.00	1.00	1.00	40
D	1.00	1.00	1.00	40
GH	1.00	1.00	1.00	40
K	1.00	0.88	0.93	40
N	1.00	0.97	0.99	40
O	0.87	0.97	0.92	40
T	1.00	1.00	1.00	40
Y	1.00	1.00	1.00	40
Accuracy			**0.98**	**400**

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 mode.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

- **Recall**

  Recall refer to sensibility, it is a slice of relevant prediction made over all amount of relevant prediction.

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Nagative}$$

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

- **F-Measure**

  It is the accuracy measure of the prediction mode. The calculation of F-Measure or F1-score used two metrics which is Recall and Precision, where Precision is the total number of correct positive prediction divided by total positive returned prediction while Recall is divided by relevant predictions (the predictions supposed to be made), F-Measure range between 0 and 1, prefect recall and precision make F-Measure 1 while worst precision and recall make F-Measure 0.

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

After recognizing the sign gesture, the system will put them in an array to create a word in Persian and by help of a bilingual dictionary in this case translate library in python, the system translate the word from Persian in to English, the English word will send to client Android API in order to show to user.

Chapter 6

Conclusion and Future Work

6.1 Conclusion and future work

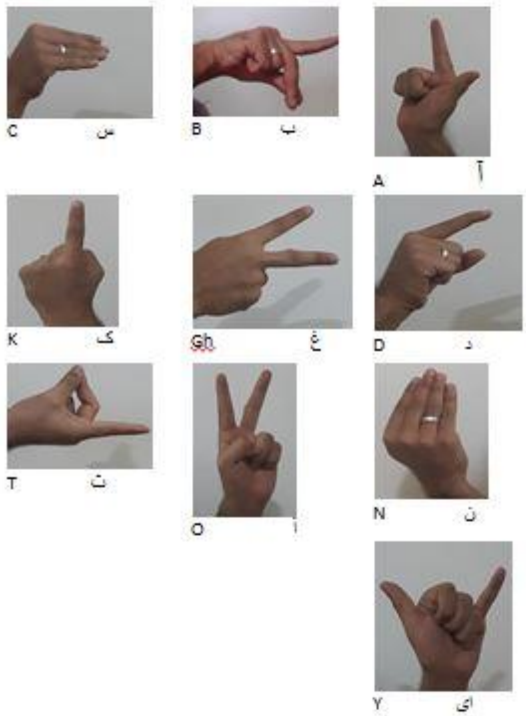
There are many Computer base systems for sign language recognition as mentioned in this paper, but they are not, practical because not possible to carry around for sign gesture recognition, that is needed for any time at any point to communicate with others. The best way to address this gap is smartphones, which are portable, available, and easy to use.

In conclusion, this paper discusses methods of Persian Sign Language gestures translation to English text on smartphones. As stated earlier, the major problem with smartphones is computational power [6], which in this case a client-server system proposes to overcome this constraint. To improve the performance of system a CNN classifier is used, and to translate from Persian to English a bilingual dictionary is used.

Carrying out the research, I propose, for features extraction from picture of sign gesture normalizing the picture and to reduce the number of feature resize it to to 64*64 pixels and for robustness of system binary pixels used as features and using CNN for classification. We have used 10 Persian Sign language gestures, achieving accuracy of 98% which is better than any other work mentioned in this paper. The future work is to increase the number of alphabetic feature and translate the sentences instead of words, and achieve higher accuracy.

Appendix

Persian Sign Gesture Recognition



References

1. Lewis, M. Paul; Simons, Gary F.; Fennig, Charles D., eds. (2013), "Deaf sign language", Ethnologue: Languages of the World (17th ed.), SIL International, archived from the original on 2013-11-26, retrieved 2013-12-03.

2. Mohandes, M., & Deriche, M. (2013, April). Arabic sign language recognition by decisions fusion using Dempster-Shafer theory of evidence. In 2013 Computing, Communications and IT Applications Conference (ComComAp) (pp. 90-94). IEEE.

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

4. Jalilian, B., & Chalechale, A. (2014). Persian sign language recognition using radial distance and Fourier transform. Int. J. Image, Graphics and Signal Processing, 1, 40-46.

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. Sarkaleh, A. K., Poorahangaryan, F., Zanj, B., & Karami, A. (2009, November). A Neural Network based system for Persian sign language recognition. In 2009 IEEE International Conference on Signal and Image Processing Applications (pp. 145-149). IEEE.

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. Karami, A., Zanj, B., & Sarkaleh, A. K. (2011). Persian sign language (PSL) recognition using wavelet transform and neural networks. Expert Systems with Applications, 38(3), 2661-2667.

10. Lahoti, S., Kayal, S., Kumbhare, S., Suradkar, I., & Pawar, V. (2018, July). Android based American Sign Language recognition system with skin segmentation and svm. In 2018 9th

International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

11. Haque, P., Das, B., & Kaspy, N. N. (2019, February). Two-Handed Bangla Sign Language Recognition Using Principal Component Analysis (PCA) And KNN Algorithm. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-4). IEEE.

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, H. D. L. (Ed.). (2008). Open your eyes: Deaf studies talking. U of Minnesota Press.

15. Nielsen, K. E. (2012). A disability history of the United States (Vol. 2). Beacon Press.

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. Shin, Y., & Balasingham, I. (2017, July). Comparison of hand-craft feature based SVM and CNN based deep learning framework for automatic polyp classification. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 3277-3280). IEEE.

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, J., Zhou, W., Zhang, Q., Li, H., & Li, W. (2018, April). Video-based sign language recognition without temporal segmentation. In Thirty-Second AAAI Conference on Artificial Intelligence.

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. Jalilian, B., & Chalechale, A. (2014). Persian sign language recognition using radial distance and Fourier transform. Int. J. Image, Graphics and Signal Processing, 1, 40-46.

26. Fakhr Kambiz. (1397/2018), Amozeshe zabane nasenawayan, ertebat ba afrade kar va lal, The research team professor Fakhr

فخرالدین کامبیز، (1397)، آموزش زبان ناشنوایان، ارتباط با افراد کر و لال. تیم پژوهشی استاد فخر.

27. Li, Y., Zhang, H., Xue, X., Jiang, Y., & Shen, Q. (2018). Deep learning for remote sensing image classification: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(6), e1264.

28. Zakariya, A. M., & Jindal, R. (2019, July). Arabic Sign Language Recognition System on Smartphone. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.

29. Mwiti, D. (2018, May 8). Convolutional Neural Networks: An Intro Tutorial. Retrieved from https://heartbeat.fritz.ai/a-beginners-guide-to-convolutional-neural-networks-cnn-cf26c5ee17ed

Number of publication:

1. Muhammad Reza Jafari, Dr. Vinod Kumar, "Natural Language Interface System for Querying Wikidata", International Journal of Science and Research (IJSR), https://www.ijsr.net/search_index_results_paperid.php?id=SR20506232010, Volume 9 Issue 5, May 2020, 539 – 542.
2. Muhammad Reza Jafari , Dr. Vinod Kumar, 2020, Persian Sign Gesture Translation to English Spoken Language on Smartphone, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 06 (June 2020).

A means of communication among people with verbal disability

ORIGINALITY REPORT

7%	%	7%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Jack Sklansky. "Image Segmentation and Feature Extraction", IEEE Transactions on Systems Man and Cybernetics, 1978 Publication	3%
2	Sakshi Lahoti, Shaily Kayal, Sakshi Kumbhare, Ishani Suradkar, Vikul Pawar. "Android Based American Sign Language Recognition System with Skin Segmentation and SVM", 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018 Publication	1%
3	Ayad R. Abbas, Ayat O. Farooq. "Chapter 15 Human Skin Colour Detection Using Bayesian Rough Decision Tree", Springer Science and Business Media LLC, 2018 Publication	1%
4	Abbas Muhammad Zakariya, Rajni Jindal. "Arabic Sign Language Recognition System on Smartphone", 2019 10th International Conference on Computing, Communication and	1%

26/6/2020

Networking Technologies (ICCCNT), 2019
Publication

5	"Cognitive Cities", Springer Science and Business Media LLC, 2020 Publication	<1%
6	"Intelligent Computing, Networking, and Informatics", Springer Science and Business Media LLC, 2014 Publication	<1%
7	Bahare Jalilian, Abdolah Chalechale. "Persian Sign Language Recognition Using Radial Distance and Fourier Transform", International Journal of Image, Graphics and Signal Processing, 2013 Publication	<1%
8	Khamar Basha Shaik, P. Ganesan, V. Kalist, B.S. Sathish, J. Merlin Mary Jenitha. "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space", Procedia Computer Science, 2015 Publication	<1%

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		

Vinodkumar 26/6/2020