Bi-directional Robotic Communicator for Hearing Impaired

Author
Sobia Fayyaz
Reg. No. 171411

Supervised By
Dr. Yasar Ayaz

Department of Robotics and Intelligent Machine Engineering
School of Mechanical and Manufacturing Engineering
National University of Sciences and Technology
Islamabad, Pakistan
August, 2019

Bi-directional Robotic Communicator for Hearing Impaired

Author
Sobia Fayyaz
Reg. No. 171411

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Robotics and Intelligent Machine Engineering

Thesis Supervisor
Dr. Yasar Ayaz

Thesis Supervisor's Signature: ____________________________________

Department of Robotics and Intelligent Machine Engineering
School of Mechanical and Manufacturing Engineering
National University of Sciences and Technology
Islamabad, Pakistan
August, 2019

DECLARATION

I certify that this research work titled "*Bi-directional Robotic Communicator for Hearing Impaired"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

Sobia Fayyaz

2016-NUST-MS-RIME-00000171411

PLAGIARISM CERTIFICATE

It is certified that PhD/M.Phil/MS Thesis Titled *Bi-directional Robotic Communicator for Hearing Impaired* by Sobia Fayyaz, Reg no. 171411 has been examined by us. We undertake the follows:

a. Thesis has significant new work/knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
b. The work presented is original and own work of the author (i.e. there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.
c. There is no fabrication of data or results which have been compiled/analyzed.
d. There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
e. The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

Name & Signature of Supervisor

Dr. Yasar Ayaz

Signature**:** _______________

COPYRIGHT STATEMENT

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Mr/Ms _____Sobia Fayyaz_______ Registration No. _________171411________ of __**SMME**___ has been vetted by undersigned, found complete in all aspects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature with stamp: ___________________________

Name of Supervisor: _____Dr. Yasar Ayaz___________

Date: ____

Signature of HoD with stamp: _____________________

Date: ____

Countersign by

Signature (Dean/Principal): ________________

Date: ___________________________________

FORM TH-4

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: (Student Name & Regn No.)______________Sobia Fayyaz, Reg no. 171411______________ Titled: ____Bi-directional Robotic Communicator for Hearing Impaired____be accepted in partial fulfillment of the requirements for the award of ______________________degree.

Examination Committee Members

1. Name: AP Sara Baber Signature:________________

2. Name: Dr. Hasan Sajid Signature:________________

3. Name: Dr. Muhammad Naveed Signature:________________

Supervisor's name: Dr. Yasar Ayaz Signature:________________

Date:____________________

____
Head of Department

____
Date

COUNTER SINGED

____
Dean/Principal

____
Date

ACKNOWLEDGMENT

First and foremost, all gratitude and praise belongs to Almighty Allah, the creator, sustainer, protector, and the Lord of the Worlds. May the product of this work be practically emerged in the near future for those in need for social inclusion and decrease in correspondence obstructions. Secondly, my parents for their prayers; encouragement and mental support during my studies. I would like to express my sincere thanks to all my teachers, professors and friends whose words of wisdom helped me to move forward and accomplish the goal.

Dedicated to the hearing impaired community in Pakistan and all over the world. May this effort open up a way for these communities to get access to more facilities specifically in education sector and better social inclusion in the near future.

ABSTRACT

For many years, Gesture recognition and sign language have been a center of research for different domains and use cases. However, Pakistani Sign Language (PSL) failed to gather such interest as compare to other regional variants including American Sign Language (ASL), French Sign Language (LSF), British Sign Language (CSL), to name a few. The proposed system bridge a communication path between the signer and non-signer using Deep Learning technique. The system acquire the gesture information by extracting the local and global features from the static sign images using Convolutional Neural Network (CNN) and then translate the gestures into spoken language for the non-signer. The system is trained on approximately 4500 static images split in train and test sets with given annotations. It is capable of translating the spoken words to gestural language to make it easy for signer to get involved in a conversation.

Developing a robust and better algorithm for the hearing impaired community to facilitate them handling minimum social contact hurdle and get them access to education and employment services is the main purpose of this work. The proposed system demonstrates an accuracy of 90% for twenty-four PSL gesture corresponding to the selection of English alphabets. This research can be considered as a step into a long-term research plan towards transforming and creating a real-time PSL recognition system that is able to assist deaf individuals within their day-to-day lives that deals with static as well as dynamic gestures and minimizing the gap between the Pakistani society and its marginalized deaf community.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

This section will give the background information with respect to gesture based communication and the significance of building up a framework for Pakistani Sign Language (PSL) specifically. This chapter will also briefly introduce sign languages and their evolution in different regions within the communities. A brief overview of the different profit and non-profit associations in charge of PSL advancement will be featured along with their commitments and contributions to the regional deaf community. The progress within the ASL will be contrasted with PSL to set a benchmark for the development, improvement and promotion of PSL within the hard of hearing network. The proceeding sections will define the research aim and scope, which will keep the readers to limit the scope with specified context. Lastly, an overview of the layout of the research will be outlined, which will demarcate the specific agenda and content of discussion for each of the chapter within this research.

1.1 Research Background

1.1.1 Sign Languages and their evolution

Language and cognitive development are correlated with each other that occur at early stages of human life. Hearing people use spoken language to exchange the information and communicate with each other. Oratory facilities are used in spoken languages to produce sounds mapped against specific words and grammatical combinations to convey meaningful information that processed accordingly [227]. People who have hearing disabilities use sign language (SL) to interact with other individuals. Sign language exist and develop independently like other natural spoken languages and it requires the manual-visual physical modality for proper understanding and comprehension of inter-personal communication [226]. For example, ASL has grown autonomously from English language spoken in the United States, having in excess of 22 million (approx. 10% of the population) hard of hearing people [2], [3], out of which many can't appropriately communication, function or work in their daily lives, despite ascend in innovation in technology and related applications [4], [5]. As there is no standard sign language using in all over the world instead almost every region have its own signs to interact with non-hearing individuals. People from different areas of world have variations in their languages and they acknowledge not only spoken language but also sign language, this range surpasses 6,909 and 138 respectively [6]. Despite of great need of sign language in the mid eighteenth century, this language stayed static and was being ignored till 1950s due to lack of proper understanding and assessment tools [7].

Every language has some set of rules of grammar to express information in meaningful way but not everyone was able to recognize what is being conveyed through sign language until Tervoort [8] performed a structural analysis on gestural language. His findings showed that sign language holds its properties independent linguistic system in the different modality, as each spoken word has its corresponding gesture that keeps a constant meaning and specific form, derived independently from the spoken language [8]. Stokoe [9] developed a model incorporating the phonological parameters with variations in hand shapes, arm movements, upper-torso movements and facial expressions [9]. In 1980s, Sign language was given adequate attention and the researcher and community started to work for its promotion and provide facets in order to firmly establish the fact that sign languages were separate languages with their unique linguistic and lexical identities and structures [1].

Children learn the gesture language in the same way as acquiring the spoken language, but phonological substitution, omission and over-generalization are the mistakes usually commit by them [10]-[12]. Beside all basic needs, Education is also necessary for human being to survive in this society. Deaf and dumb need other form of social and educational support, Most sufferers learn gesture representation of words from their guardians as approximately 90% deaf children belongs to hearing parents. So the main source of their learning is home, and sign language can be teach or 'created - with easy gestures' within domestic settings; eventually it will give rise to gestural language (followed with grammatical structure) with cultural values which is then reported to deaf school in Managua, Nicaragua [13]. The awareness regarding gesture based communications inside the general communities is not very wide leading to misconceptions and confusions which increase negative stereotypes and sense of division to deaf individuals[14]. Basic misconceptions hold by other communities are: (i) Gesture language and spoken language carry similar structure and syntax (ii) Signer from one region able to communicate with other regional signer with the same gesture (sign express similar meaning)(iii)Sign language evolved globally in the same manner. All global sign languages have evolved in uniform manner [3], [15]. Although, many other communities from different areas including Desa Kolok (Bali), Negev desert (Israel) and Adamorobe (Ghana) were reported, local hearing population achieved fluency in sign language where genetics caused deafness ratio is high [16], [17]. But in other areas, people are even not in state to understand what is being said using sign language therefore deaf individuals find it difficult to get mingle in communities. However, Deaf community is treated as stigmatized and marginalized within contemporary societies, and there is not actually system that could make the communication reliable and comfortable between local deaf community and hearing population. Same as other regions, Disable or hard of hearing population of Pakistan is concealed and greatly stigmatized leading to various issues including psychological, emotional, economical, and to name a few [18].

1.1.2 Pakistani Sign Language: An Overview

The development of PSL is still in progress. Many institutions and non-profit organizations are playing their role in building the standardized dictionary and lexicon for PSL but unfortunately they are unable to be on the same pace. As a result, regional sign languages from within same country are getting promoted. This section tells about the some of the known contributors to the widespread usage and betterment of PSL will be highlighted, with the contribution and time span they are following. No uniform sign language is being used in Pakistan if we compare this situation to all over the world, even to this date. Following sections will enlighten the gap observed in comparison with ASL along with its promotion and learning techniques in deaf community of United States. However organization taken the responsibility to promoting the PSL are listed: (i) *Sir Syed Deaf Association* (SDA), (ii) *Anjuman Behbood-e-Samat-e-Atfal* (ABSA), (iii) *National Institute of Special Education* (NISE), (iv) *Pakistan Association of Deaf* (PAD) and (v) *Family Education Services Foundation* (FESF) [19].

Two PSL dictionary versions were published by ABSA in the 1990s which introduced a new signs in correspondence of daily life actions, and SDA published approximately 750 signs of PSL Lexicon [19]. But the authors was unable to find any dictionary which were claimed to be publicly available. After that, PAD and NISE did effort to make a training sessions where they emphasized on the deaf learning strategies and they worked for the uniform sign language that could be use all around Pakistan [19]. FESF, works for the development and betterment of deaf community with holistic approach. They offered Job Placement Program(job where hard of hearing people can work comfortably), Parent Training Sessions (they teach parents so that communication habit and basic understanding to express something develops at home in deaf individuals),Teacher Training (focus on methodologies which can be used to teach children with some target or curriculum),Vocational Skills Training Program(life living works such as tailoring, Food course, IT course opportunities for deaf community) and Academic Program (FESF's Deaf Reach Schools across Pakistan) [20]. Moreover, online resources are available now in form of mobile application, compact disk (CD) and specially in hard copy form (book) to learn PSL easily at anywhere, and it is the result of efforts of FESF [20].

1.1.2 American Sign Language: Historical Development and Progress

Laurent Clerc, a french instructor, work at first ever school for deaf-and-dumb in Hartford, Connecticut. He introduced the teaching methods for deaf students so development of American sign language (ASL) was started, also this is the reason of similarity between ASL and LSF in term of associated word and gestures [14]. There was not acceptance of distinction between sign language and spoken language at the time, so there was also introduced the singing technique for US deaf community [21]. Stokoe's [9] research spread the awareness among people and they started to think gestural language as separate from linguistic, so ASL was formally established and declared as a language. Stokoe with contribution of his friend, published a set of lexicon and explained the structure and syntax of ASL [23].

With the ASL, rights of deaf community were promoted and considered on government level that cause the build of policy in 1967 in which educational, social interaction and language competency clauses were mentioned. With the growing interest of communities in sign language, California State University (CA) started educational programs for the deaf community and in the meanwhile, American Sign Language Teachers Association (ASLTA) was also formed for the betterment and improvement of ASL course curriculum [25]. As educational setups took initiative to provide the services to the hard of hearing network, sign language was being introduced as a second language in United States [26], [27].

In support of the population using gestural language, The National Institute on Deafness and Communication Disorders (NIDCD) was formed in 1988 for research purpose and facilitate the groups with other disorders including different senses such as taste, smell, speech, stammering to promote the reliable and effective communication among citizens [28]. The primary agenda of NIDCD has many points, some of them are:(i) conduct research and determine the possible reason (genetics, environmental, toxic and infectious) that cause deafness (ii) Examine the human body capability to regenerate the hearing process (iii) analyzing the perpetual and cognitive processes which can affect the communication, and (iv) to develop rehabilitation devices using growing technology to enable the deaf community deal with social inclusion [28]. In 2007, a group of consultants from diverse fields, sign language trainers, and curriculum maintainers formed the

American Sign Language National Standards committee; whose aim was to draft a standardized and uniform learning to students in US from primary classes to high grade classes [25].

The NIDCD was giving plans occasionally and plan mentioned before was initiative taken by different government and non-government bodies in the US, which eventually permitted the hard of hearing network to teach, work and contribute viably inside the general public or citizens. Simultaneously, with increased number in Bilingual educational programs; deaf individuals were motivated to take part in the society and have higher level of education. These programs helped them to improve contact with hearing people and encouraged them to live a normal life with positive attitude towards opportunities. With all these effort of various acting organizations, ASL became the fourth-most language adopted by individuals for higher education in the US [25].

1.1.3 Need for Promoting PSL education: Lessons Learned from ASL

Regardless of the disability, no matter of what type it is, disable group of people has their rights and they should be given the facility to work for the nation and for their lives. There is a great need for highlighting the PSL in every domain i.e. Education, employment etc. Although the importance of sign-language is known in the world; still in Pakistan, deficient consideration regarding gesture based communication learning is cultivating estrangement inside the deaf community but on the other side they are struggling to lead a normal life and communicate effectively within the society.

Literature on deaf language learning has clearly shows the benefits of sign language education as well as it indicate the loss caused by not fulfill the educational needs and communication desires. Like other languages, learners have to focus on the structure and syntax of the sign language so it is good to teach the children (with hearing loss) as a native language it will make the difference in their lives as they grown up. Great difference was reported in academic performance of children who learned the gestural language at the early years as compared to those who started after time [7], [30], [31]. These findings have been noticed not only for ASL but Danish, Swedish and other gesture based languages also [32]–[35]. Delay in guiding the proper structural sign language to deaf children can have negative impact on their cognitive and linguistic development, vocabulary growth in coming time [42]-[44]. In the society of Pakistan, people with hearing ability holds a negative perception regarding hearing impaired community and that’s why schooling or educational setup is different for these people. The segregation presents a problem of social inclusion and eventually no communication or conversation is taken place among the signer and the non-signer community [18]. Whereas schools in US make it easy for its deaf community to survive as normal as other people do by providing them education medium, employment opportunity and other rights which they deserve as a citizen [45]. There is no stereotype or superiority in term of hearing disability.

Following the positive insights and steps being focused by Pakistan's organizations for the awareness and promotion of PSL among every common society, impressed by the development of ASL in US:

- Awareness regarding education using sign language and breaking out the societal stigma associated with disabilities.
- Acceptance of disabled communities as a culture with their independent identity in Pakistan.
- Eliminating the age-group problem and designing the training programs and curriculum for all ranging from children to their elders.
- Bilingual learning setting to alleviate the participation of marginalized deaf community with the hearing group in Pakistan.
- Funds collection from the public and private supporting organizations to promote the deaf community related programs and make its existence in higher educational framework with bilingual mode.

1.2 Research Aims

The gap between the ASL educational and PSL institutional setup can be easily understandable from the preceding sections. Deaf community of Pakistan is provided with limited opportunities in term of employment and education as well, so it is great to take initiatives to fill the gap between society of signer and non-signer citizen using technology inspired solutions. In contrast to the sign languages in the world, PSL has stayed ignored for the noticeable period in the scholarly world in term of honest approach developing a system able to recognize PSL. Any labor or endeavors made were not fully practical and implementable [46], [47].

This research aims at providing convenience to the deaf community of Pakistan by building a system that able to maps the human gestures to the linguistic sensory information and vice versa. Providing them modality to get understanding and improve communication among combined community. It is an effort to remove the communication hurdles and to giving a recognition to PSL not locally but also in international circle. The prior consideration of the research endeavors should be to provide a practical solution with acceptable accuracy and minimum ratio of false interpretation that can give rise to living a quality life of all members of the society. Consequently, this research is an attempt to develop a AI based software application (Desktop and mobile app) for PSL recognition.

The most advantageous action to develop a system with existing PSL Lexicon prepared by FESF, and also system must be capable of handling the real-time data because the environment could possibly be dynamic and little toleration can become alarming. System developed using Tensorflow framework and Deep Learning Technique. Static gestures were used to accomplish the task of building the system. The software comprised of dual-modules. First module, Convolution Neural Network is trained to translate image feature vector into speech. For second, Speech is interpret to the correlated gesture image. It currently able to interpret the twenty-four PSL English alphabets in closed environment. However, system is extendable if it is given the dynamic gestures data for the complete reliable and efficient communication. it is completely real-time and has been tested in the closed environment with currently certain conditions whose affect can be minimize with the improved gesture data.

1.3 Research Focus and Scope

The disparity in PSL progress in contrast with other developing sign languages is highlighted in the preceding sections. As physical disability is not the sign of being unskilled or less intelligent, it is our responsibility to provide them opportunity (Education, social, employment) and make a platform for them so they can live a quality life being a citizen and also participate in increasing the Economy of Pakistan. The PSL vocabulary is now more than thousand words and maintained in many forms such as books, CDs etc. Dynamic gestures with torso movement and facial expressions are currently not the part of this research but work can be done on it in the near future. In open environment, more data processing need to be done and steps such as (i) hand segmentation and motion tracking (ii) face recognition and expression identification (iii) track variations and movement of body parts such as upper-torso (v) multi-user interfered sign toleration. This work mainly target the English alphabet static gesture recognition from image to speech and speech to image vice-versa. Targeted twenty-four gestures used within the research are displayed below in the figure.

Figure 1.1: English alphabet based static PSL gestures

Dynamic gestures were not included in the research, proposed system work under environment with certain background conditions and more deal with the children learning, game and social inclusion. System is capable of to translate the spoken sentence and divide into chunk of characters of words and display it to help in conversation. It could also be extended to designing the social connecting app for deaf community.

The salient features of the bi-direction communicator for hearing impaired are outlined:

- The system designed for closed environment or indoor conditions.
- The proposed system is able to interpret the 24 gestures that do not contain any other movement, instead it is fixed.
- The system has been tested with tidy background with no shadow
- The proposed system allow a speech input and then convert into corresponding gestures on which is trained.
- Divided into two modules: image to speech (i2s) and speech to image (s2i)

Single Handed Gestures

Double Handed Gestures

Figure 1.2: English alphabet based static and dynamic PSL gestures

1.4 Structure of the Study

This research work is organized into five chapters, where every part centers on a solitary basic viewpoint in satisfactory detail to upgrade understanding the need of building such system and how it is working and can be improved in coming time. Following is a list of chapters clarifying some points that needs attention of all of us:

- Chapter 1 gives a thorough introduction, which will basically establish the understanding of the research topic and the endeavors of contributors and researcher for the concerned community of Pakistan and other countries (US - highlighted among all). Development and changes in programs initiated specifically for the hard of hearing citizens with time. it also explains how could be the problem of learning, employment, significant associations and social inclusion can be minimized.
- Analysis of existing available literature and facts is conducted in chapter 2. It will give the information regarding other gesture languages with their performance, and changes observed with variations in sources. Authenticated information published by verifiable peers will be shared in this part.

- The justification of this conducted research with the methodology adapted in comparison with the existing solutions or techniques is given in chapter. It is researcher's sole intention to provide details related to each step taken, so that it would be easy for the future researchers to make an improved version of the proposed system which can be implementable.
- The results, performance and overall accuracy of the system will be highlighted in chapter 4. The primary challenges will be identified so future work can be centered on eliminating the deficiencies or limits of the system.
- The primary findings and the obtained result of this endeavor will be given in chapter 5. Results obtained using different techniques to get better accuracy and the decision for choosing the best algorithm based on performance will also be shared in context of the proposed system discussed in chapter 1. The limitations and improvements that can affect the results or performance of the system in future work will also be discussed.

Chapter 2. Literature Review

Researches that have been conducted on sign language recognition will be mentioned with details in coming sections of this chapter. As it is a long journey that past researchers had traveled, it might be not possible to cover each and every research conducted in past time. but intention is to broadly cover the major steps and techniques that made the positive difference in the society. There has been a huge number of gesture based applications and also sensors, methods to get data from them, algorithms to run in collaborations with sensors exist and Sign Language Recognition (SLR) is a sub-domain of gesture recognition systems (GRS) but with specific and dedicated hand gestures. Chapter is categorized for two parts: one deal with research concepts for gesture language and second with the developed systems being used for recognition in the world. Figure 3, shows gesture recognition systems with all its sub-classifications including Human-Robot and Human-Computer interaction, Augmented or Virtual Reality (AR), and Sign language (with its variants)

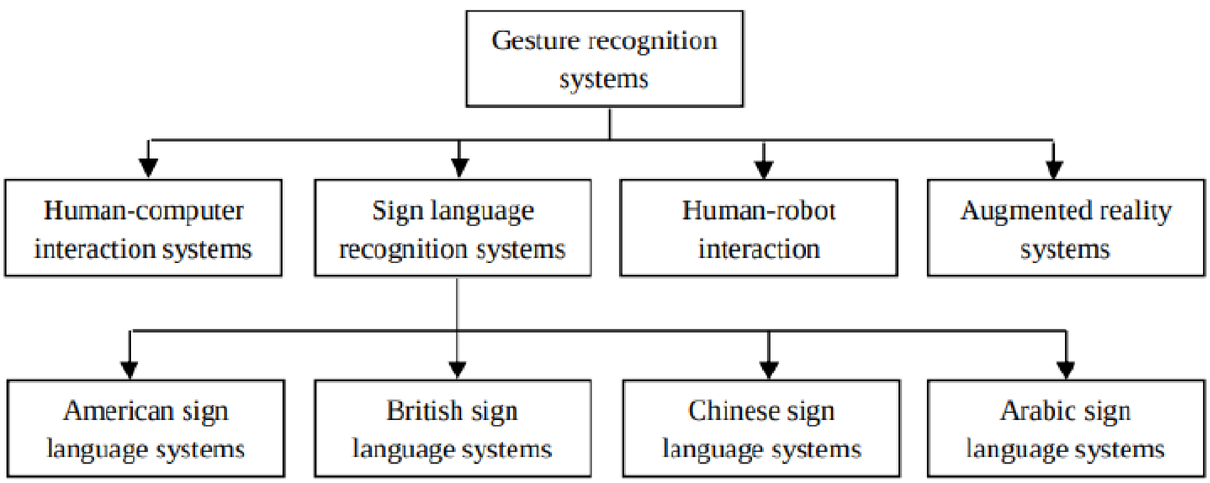

Figure 2.1: Gesture recognition-based systems with its sub-classification (including sign language recognition systems)

2.1 Gesture Recognition Systems

Gestures are a mean of communication for many people which includes facial expressions, head, shoulders or upper-torso movement that collectively known as non-verbal or body language of any individual. Researchers paid wide attention to the gesture and human interacted systems for more than a decade. One major distinction between GR and SL systems is SL works with defined set of gestures of lexicon that are associated with the words and phrases whereas GR typically accept the gestures that are provided to it by researcher depending on the use-case or application where it is being used. Colored gloves with sensors installed are used for the HRI systems, followed by the detection of skin color [49], [50] input shape and input dept details [51],[52], However, 3D hand modeling systems are also active in research area and mostly developed by using Hidden Markov Model (HMM) [53], [54]. HGR system designed and developed by Franklin and colleagues [55] that controls of robot-waiter using gesture commands. Similarly, hand gesture was used to control the vehicle by Guo et al. [56].Gesture interface integrated with stereo-vision developed for robot by Cipolla and Hollinghurst [57]. As improvement in technology is definite, Kortenkamp et al. [58] build another system using HRI that create model of a human roughly to map the real work field. A vision based system for color tracking by hand shape and adaptive algorithm for color tracking done by Waldherr et al. [59]. More robust system with pleasant results developed by Yin and Xie [60] that uses the neural network for getting recognized hand gestures, also voice signals incorporated in the system. Wachs et al. [63] built control system for improved hand gesture recognition using fuzzy C-means (FCM) algorithm with supervision and heuristic labeling (faster convergence, high accuracy).

Correa et al. [64] proposed HGR module for the RoboCup League which includes Skin detection, segmentation, tracking and identification using boosted Bayesian classifiers (better performance than Hidden Markov Models but system). Algorithms was proposed by Posada-Gomez et al. [65] and Hashimoto et al.[66] to control the movement of wheel chairs while Montesano et al. [67] worked with voice signal for the same use-case. Bohme et al. [68] developed a control for service robot operating able to work in indoor environments. Weitzenfield et al. [69] used both vision and auditory information to robot football players. Van den Bergh et al. [72] used robot navigation system along with depth image segmentation of gestures from cluttered background. A robot named SIATRob, developed by Xu and colleagues [74], used 3D depth perception and hand tracking with motion trajectory captured by Kinect Platform which is then input to HMM to learn hand gesture sequences for better and efficient navigation. Another research conducted aimed to force to generate motion patterns of human gestures and re-plan the trajectories in 1-Dimension [76]. Yokoyama et al. [77] worked on collaboration of HRP-2P humanoid and human worker, who was able to grasp the wall panel and used various sensors including tactile, velocity and navigation, rotation monitoring and stereo vision. Sign language was used a primary mode for communication between human robots and human workers, Monte Carlo method used to make robot adapt to human movements and observe unexpected reactions in return [78].

Elgammal et al. [79] proposed a GR system with exemplar approach, probabilistic framework were also used and system was able to represent gestures as a human poses. Fang et al. [53] outlined a efficient and robust approach (in comparison with learning based models which takes much time and computation) to build GR system. After that, Discrete Gaussian derivatives with box filters and blob features for multi-scale feature detection of hand gestures proposed by Bretzner et al. [81] which was giving a high success rate of gesture recognition. Hsieh et al. [82] presented an algorithm (able to detect skin color, Boundaries or edges, Fourier descriptors, SVM for learning) to build a real time system to recognize gestures. Different techniques (algorithm-wise and varied sensors) was being used to build GR systems with greater accuracy. Liu et al. [83] used Dynamic Time Warping (DTW) along with GR algorithm (achieved high accuracy) which require only a sample of single gesture to learn features from it unlike glove-based and camera-based algorithm. Bretzner et al. [81] developed recognition system for household electronic so that it can control remotely by using hand shapes (processing done with shapes such as extraction, tracking etc). In [84], SVM used with another approach, AdaBoost method with Region of Interest (ROI) for detection and extraction, Kanade-Lucas-Tomasi (KLT) for tracking, in order to achieve high accuracy system and distance of about 1-2 meters are required between system and user for better functionality. Abid et al. [85] used SVM with Bag-of-Features model and 3D Histogram of Gradient Orientation (3D-HOG) approach.

There are a wide range of sorts of expanded reality frameworks, some of which effectively apply diverse hand signal strategies and GR frameworks have potential for developing applications for augmented or virtual reality. Multi-image of same scenario based segmentation with Kalman filter (tracking the movement) used by Lahamy and Litchi [87], and the images captured by time-of-flight (TOF) camera. Oil and gas reservoir simulations were designed where objects moved by human gestures in augmented reality [87]. In [88], faster performance was observed to do particular task as compare to peripheral devices [88]. technology gadgets developed by Lambercht and Kruger [89] which enable human to interact with industrial robots in augment reality using hand gestures with other body movements that support the gesture to convey meaning. Kolsch et al. [90] and Lo et al. [91] conducted research on wearable devices or gadgets on which AR systems can be installed or implemented. Kolsch et al. [90] was working on AR systems that not only task gesture data but other signals also such as speech data. Lo et al. [91] developed a touch free gesture system with head mounted display that works in High Dynamic Range (HDR) and neural networks (gesture learning).

First Author	System Type	System Characteristics	System Details
Lo [91]	Augmented reality	3D range camera with FreeGlass	3D GR using NN
Pu [86]	Appliance control application	Wi-See	Whole home GR for multiple applications
Wachs [63]	HRI	Vision-based system	FCM with heuristic labelling
Abid [85]	Home Application	Vision-based system	3D HOG and non-linear SVM
Lahamy [87]	GR	Vision-based system	Using SR4000 for GR
Xu [74]	HRI	Kinect sensor	3D GR using HMM
Lambrecht [89]	HRI using AR	Kinect sensor and vision-based system	2D GR and Object recognition using Kinect
Sohn [78]	HRI	Performing collaborative tasks	GR using Monte Carlo method
Trujillo- Romero [73]	HRI	Vision and speech-based system	GR using NN
Kollorz [92]	GR	Vision-based system using TOF camera	GR using feature classification algorithm
Hikawa [93]	GR	Vision-based system	GR using JRNN
Reifinger [88]	AR	Infrared-based system	GR using HSI and HMM
Liu [83]	GR	uWave, a three-axis accelerometer	allows for personalized GR with single training sample/sign
Correa [64]	GR and HRI	Vision-based service robot	GR using Bayes' Classifier and skin model

Table 2.1: Basic system information about different gesture-based research implementations

2.2 Sign Language Recognition Systems

Sign language, a mode of communication used by people suffering with hearing impairment or deafness. There is a great need to address the communication barriers, facilities and services between the signer and non-signer communities in Pakistan. This non-verbal language based on different but specific hand shapes along with expression corresponding to certain actions and words. Numerous systems with sensor inputs (vision) developed to facilitate the deaf community, But due to lack of awareness and knowledge of sign language causing problem of social isolation. Gesture perform for contact are static and dynamic which taken in specific timeframe with not-changed shape in space and series of time-frames with position change respectively [80], [94]. Many regional sign language exists with variation in hand shape for same action [95].

In recent decades, a thorough literature has been gathered related to worldwide and nearby gesture based languages just as the various strategies applied to resolve the issue of communication through image-based models. Due to wide usage of sign language among deaf community, American Sign Language (ASL) got popularity in United States and it also inspired the other countries to focus on the local hard-of-hearing communities. Starner et al. [97] developed a system (desktop application works with two cameras and able to track the movement in frames) for the recognition of ASL in real time using Hidden Markov Model-based (HMM), it achieved accuracy ~98%. Uebersaux et al. [98] presented a system that functions when provided with a depth data, it recognize the gesture by segmenting hand and estimation the orientation of hand. Cooper et al. [99] used Markov models to analyze the temporal changes in linguistic units, discriminative feature selection using Sequential Pattern to reduce the effect of noise. Foreground detection with Multi-layer Codebook (MCB) with Pixel Based Hierarchical-Feature AdaBoosting (PBHFA) for ASL recognition proposed by Pattanaworapan et al. [100]. It provided the improved results as compare the past researches done as it was capturing the signer back-hand images.

British Sign Language (BSL) recognition system developed by Kadir et al. [102] which was monocular and able to handle more lexicons with less training, it provided 92% accuracy. Key elements used for better recognition of sign language including hand and head detection, activity observed from movement of body and dual level classification [102]. Dreuw et al. [103] proposed a appearance-based system to recognize German finger spelling alphabet under different conditions by individuals. In [104], work on Polish finger gesture recognition was done with images of high resolution and depth information. Colored gloves to build database and system was developed for recognition of Spanish sign language [105]. German Sign Language interpreter system was developed using Gaussian Hidden Markov Model (GHMM) and neural networks (multi-layer perception), and different techniques were used to minimize the word error rate (WER) [106]. Another researches conducted are; Arabic Sign Language (ArSL) recognition system by Al-Roussan et al. [107] with discrete cosine transform and HMM (sinusoids ad frequency approach). Haar-like algorithm, Extracting features by Fourier transform, K Nearest Neighbor (KNN) Algorithm for classification approaches used by Albelwi and Alginahi [108] to classify Arabic Sign Language (ArSL) gestures from consecutive frames with 90.55% accuracy. CyberGlove (GLV) and Flock of Bird (FOB) systems with algorithm (Dempster-Shafer theory) used for ArSL recognition system (accuracy 98.1%) [109]. As the number of population with hearing impairment increasing in India, System has been proposed for Indian Sign Language (ISL) recognition in [110]. Tiwari and Srivastava [111] used vision-based approach (image compression using two-Dimensional Discrete Cosine Transform, Kohonen Self-organizing Feature Map (SOFM)) for ISL communication system. Further, Fuzzy Neural Network and Eigen value-weighted Euclidean distance techniques were used to recognize ISL [112] and [113] respectively. Keskin et al. [117] worked on developing recognition system for static gestures of ASL which based on Kinect platform. Lang et al. [118] and Zafrulla et al. [119] used skeletal information from Kinect to build system whereas Kinect images used by [120] for this task. Oszust and Wysocki [121] used Principal Component Analysis (PCA) and Minimum Entropy Clustering (MEC) approach along with Kinect for Polish sign langauge system. Dynamic gestures was recognized of ASL using Discriminative Exemplar Coding (DEC) with AdaBoost classifier. Development of system to recognize the Sign language used in china for communication among deaf community was active in research and researchers used HMM or ANN [123] and DGMM/HMM-based [124],[125],[126]. They used a large vocabulary set of sentence gestures techniques which was responsible for less output accuracy and response time. Zhang et al. [127] employed dimension reduction technique followed by multi resolution gesture for CSLR with position sensor and CyberGlove (accuracy 92.15%). Support Vector Machines (SVM) classifiers, Gray-level Co-occurrence matrix (GLCM), Hu Moments and SIFT (Scale Invariant Feature Transform) feature descriptors used by Quan et al. [96] for CSL recognition. Surface Electro-Myographic (sEMG) sensors and accelerometers (ACC) which are specifically used to perceive the electric signal from skeletal muscle of human arm and kinematics data, were used to develop a CSLR systems. Kosmidou and Hadjileontiadis [129] worked on intrinsic mode entropy for Greek SLR with sEMG and ACC sensors. Kim et al. [130] research to get better understanding between cameras-based and glove-based methods being used for sign language recognition. Decision trees and random fields with the electric sensors (sEMG and ACC) used for CSLR by Ma et al. [131]. Another research conducted by Li et al. [132] for CSL was done using multi-stream HMM (MSHMM) and provided the promising result (accuracy 96.5%). Unlike other growing and developed sign language used in the different regions of the world, Pakistani sign language stayed unfocused and unexplored in the society as well as in the academic domain. PSL needs investigating techniques to become stable such as discovering its gestures, sentence syntax and phonetics. Initially research conducted on PSL recognition system using Statistical Template

Matching technique by Alvi et al. [46] provided with DataGlove5 inputs. After doing some pre-processing on the provided input, gesture was translated into speech; system was improved from 69.1% to 85% which was good achievement [46]. Later, [47] worked on developing the system for recognition of Urdu sign language, data was collected and categorized with three basic colors i.e RGB and then classification algorithm was applied. This system achieved ~97% accuracy but practically it is not implementable.

First Author	Language	System Type	Accuracy	System Characteristics	Single or multiple signers
Zahedi [101]	ASL	Vision	50-70%	Appearance-based model	Multiple
Mohandes [109]	ArSL	Glove	98%	GLV with FOB system	Single
Albelwi [108]	ArSL	Vision	90.5%	Fourier descriptors with KNN	Multiple
Pattanaworapan [100]	ASL	Vision	60-70%	Pixel based Adaboosting	Multiple
Zhang [127]	CSL	Vision	88-92%	Multi-layer HMM	Multiple
Ma [131]	N/A(CSL assumed)	sEMG	91.5%	HCRF	Multiple
Li [132]	CSL	sEMG	86-97%	Sign-component framework	Multiple
Sun [122]	ASL	Kinect	86%	DEC approach	Multiple
Quan [96]	CSL	Vision	93%	Feature extraction with SVM	Single
Gweth[106]	ASL	Vision	70-84%	PCA with NN	Single
Oszust [121]	PoSL	Kinect	89-95%	Skeletal-and skin color-based features	Single
Cooper [99]	BSL	Vision	76%	HMM with Sequential pattern boosting	Multiple
Kadir [102]	BSL	Vision	84-90%	Visual Description with Boosted classifiers	Multiple

Table 2.2: Information regarding gesture recognition systems

2.3 Similarities and Variations among Sign Language Recognition Systems

Various approaches have been used to recognize the sign language with minimum false prediction. As it was considered in the research field in early stages of sign language [99], it is better to understand the variations among regional gesture language and its characteristics. Sign language the need to be identified, observed by [99] are:

- Classifiers: Specific objects represented by hand shapes.
- Positional Signs: if signs are performed on particular pat of body (Examples include: salute)
- Non-manual Features (NMFs): Facial expressions (lip spacing, eyebrow position) or body movement (shrugging shoulders).
- Directional Verbs: Direction of motion in case of dynamic gesture (Examples include ‘give phone’ or ‘take money’).
- Body shifts: Role-shifting during conversation Example include: changing gaze)
- Adverbs modifying verbs: Observe gesture with velocity as phrases can have same gesture with slightly different meaning such as ‘running quickly’ or ‘running slowly’.
- Finger spellings: signer can spell the word in gestures of alphabets in case of not knowing the gesture of action.

To build any framework or system, data is required from the sensors or devices used for particular use-case. Different methods were being applied to collect the sign language data, research made efforts to give the algorithm for better recognition of sign language. The variations identified in systems developed for gesture-based applications which are studied in [134] are:

- (i) Sensor stimuli (type of sensors used: optic, electric etc).
- (ii) Usage context (conditional changes: indoor, outdoor, cluttered and uncluttered background, gesture velocity)
- (iii) Sensor platform (restrain/unrestrained, single/multiple).

Sign language data collected by individuals or group for various purpose (academic or social) was done by using hardware sensors such as single camera [96], [106], stereo-camera [97], [99], Microsoft® Kinect [121], [122], Gloves (with sensors) [46], [109] and surface-mounted electromyography (sEMG) [131], [132]. With changes in technology, hardware equipment are getting better. Algorithm's complexity for computations is decreased now with improvement of sEMGs and ACCs, which enable the system require less data processing time. Unfortunately, systems developed using sensor gloves data seems impractical to many researchers in dynamic environment [134]. Varied Hardware integrated with GR systems will be highlighted in the succeeding section.

2.3.1 Hardware-based Variations

Sensor gloves or painted gloves were used in the early researches to make the system less computationally expensive. The gloves used for data collection purpose usually installed with more than 15 sensors including CyberGlove, Data-Entry Glove, AcceleGlove and DataGlove which were used to develop SLR system [47], [109], [135]–[138]. As glove-based approach seemed impractical and not robust to many [134], a great literature can be find on vision-based approach for the classification of sign language. AcceleGlove with two-link skeleton of arm, axis accelerometer and angular sensor used by Hernandez et al. [138] and achieved accuracy ~90%. Borghetti et al. [139] presented work to create sensor-based gloves for rehabilitation purpose and developing device for patient recovery. Many researchers [63], [92], [140], used single camera as using multi-camera can elevate computations and analysis, exponentially rise in data, also can affect the overall performance negatively. Despite of this, [57] and [97] used stereo-vision camera (not calibrated and mounted respectively) to analyze the efficiency of the real time system based of data collected. Non-vision techniques were explored using Kinect sensor [116], [141] which provide the 3D model of gesture used for developing recognition system. Augmented reality based software applications developed using three-dimensional range cameras assisting with interaction in real and virtual both environments [143]–[146].

2.3.2 Algorithm-based Variations

Variations in techniques applied to build the sign language recognition system are also observed as it occurred in case of hardware. Although literature focus on techniques rather than emphasizing the method of building dataset or external devices used. The keys to focus on any sign are observing the shape, orientation, motion (trajectory), and position called as temporal features of gesture data [147], [148]. Temporal features are more strongly associated to dynamic gestures [80], [94], [149]–[152]. The steps which is usually followed for SLR includes data acquisition and segmentation, feature extraction and classification (hand tracking and trajectory analysis) [108], [131], [132]. For example,[84] used procedures to separate face and hand (detection) along with SVM classifier. Posture classification using RNN performed by [93], [153]. Bag-of-features model and SVM for performing learning-based classification [85].

Chapter 3: System Features and Overview

The comprehensive details of the system will be provided in this chapter along with its implementation. Sign languages have grown up similar as spoken with its own structure and syntax, But the effort made by contributor for development of PSL is still not enough. we still have to work on on the unique qualities of PSL such as non-manual features, phonemes, context etc. The techniques or descriptors used to prepare the dataset is also provided in this chapter. Some feature engineering also been done to align the sign language data on the same surface. In this work, three different models were used to evaluate the performance of traditional classifiers with state-of-the-art deep learning classifier to make decision which is better for PSL interpretation. The information regarding other components of the system are provided in the proceeding sections.

3.1 Proposed System Design

The overview of the developed system can be shown in Figure 4. The core components of the systems including pre-processing, Feature extraction will be described. Within Figure 4, Both modules and their flow can be differentiated from one another on the basis of arrows and blocks. The different modules and steps used to make the system in working condition is elaborated in the coming subsections:

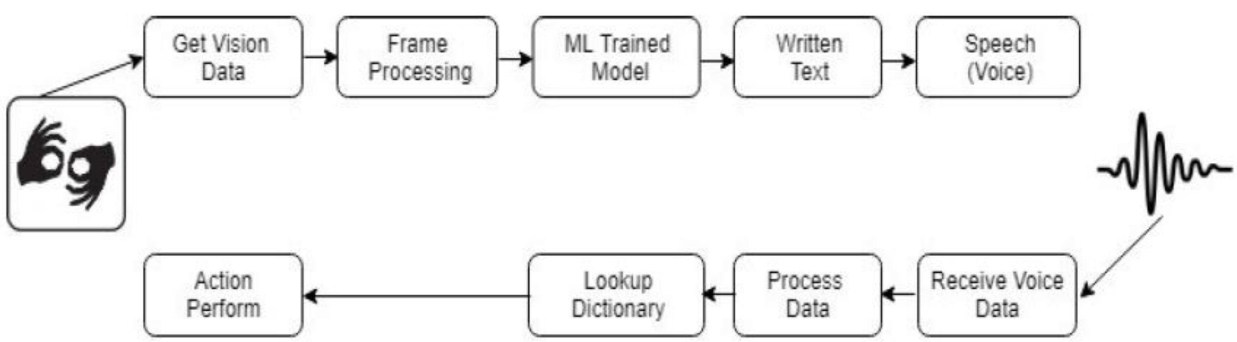

Figure 3.1: The proposed system overview (For both: signer and non-signer)

3.1.1 Dataset Description

We used the twenty-four static signs of English alphabets from sign language which are captured from non-native and native signers. Data is captured under same lighting, uncluttered background and for indoor environment. The complete dataset of static signs comprised of 4500 images. Because of no disparity between the images which belongs to identical classes, training and testing set separated from dataset with percentage of 70 and 30 respectively. Final testing of system was done by capturing the real-time images with slightly changed light and background condition. The signs along with its associated alphabets used in dataset can be seen in Figure 1.

3.1.2 Feature Extraction

Classification problem can be solved if features extracted are efficient. Extraction of feature is significant as it is responsible for extracting the relevant useful information from the images, which is used by classifiers to observe class patterns and variations among them to separate and arrange data features. The applications developed by researchers concerned with human gestures used various techniques to extract the useful features from the input. Many of them used color (RGB, normalized-RGB, HSV, HSI, YCbCr) approach and skin segmentation using different filters [157]–[162]. In the proposed system, data is resized to square image of 96 pixels. As different algorithms, Support Vector Machine (SVM), Multi-layer Perceptron (MLP) and Convolutional Neural Network (CNN), are used (best among them) to achieve the better results of the work. The details and performance of the classfiers will be given in this chapter later, Global and local Features are extracted. Features (points, edges or region patches) helps part of the image in differentiation from its immediate neighborhood [168].

3.1.2.1 Global Features

The illumination can distract the system, so we ignored it to increase the robustness of the system and saved it from false pixel detection caused by noise and shadow [163]. The images are converted into grayscale. Image complete information (i.e extraction of global features including patterns, texture or blob) for the CNN as this state-of-the-art classifier capable of extracting the features itself and don't need any external descriptor, saved in .npy file to provide input to the models. Array of image features (pixel values) and their annotation (associated alphabet label) created using Numpy Library in float Datatype.

3.1.2.2 Local Features

When extracting the local features, following properties should be considered: (i) identifiable (similar shapes should have similar features in all images), (ii) noise resistance (images should be noise free as it could affect the feature), (iii) affine in-variance, (iv) translation, rotation and scale in-variance (transformation), and (v) statistically independent [168], [169]. To evaluate the proposed system performance for traditional classifiers, SVM and MLP, we require features having aforementioned properties. Speeded up Robust Features (SURF) which is a local key-point (feature) descriptor used for feature extraction. These feature are then provided to the SVM and MLP. SURF is known for its result that is not affected by scaling, rotation, illumination etc.

(a) Feature Detection (Interest Point Localization)

Features are distinct point at varied positions in the image such as edges and blobs which are easily comparable. Features also can be points making shapes or continuous curves making connected regions. Detection of features is the fundamental step in classification. Detectors find the feature by analyzing every point in image in order to take decision whether that point should be taken as feature or not (OpenCV, 2011), they are capable of finding the same interest points when provided image with changed conditions (illumination, transformation etc.). SURF detector is based on matrix called "Hessian Matrix". With point X = (x,y) in Image I, The Hessian matrix at X with scale at σ is defined as:

$$H(X,\sigma) = \begin{pmatrix} L_{xx}\ (X,\sigma) & L_{xy}(X,\sigma) \\ L_{xy}(X,\sigma) & L_{yy}(X,\sigma) \end{pmatrix}$$

…………… eq. (1)

Where $L_{xx}(X, \sigma)$ is the convolution of the Gaussian second order derivative σ with the image I at point X, and similarly for Lxy(X, σ) and Lyy(X, σ).

(b) Feature Description (Interest Point Descriptor)

Region comprised of features is extracted and described after detection of features. A high level description can be understand from the example: "blue sky on the top and green grass on the bottom". Similarly, many descriptions found in images are provided based on the neighborhood regions. A feature descriptor or a vector is a n-dimensional vector (n varies with respect to image or objects) which represents the object in image based on measurements which include symbolic such as color, connecting points or shapes.

The interest points and their description are obtained from grayscale images which is grouped together into N centroids by using distance metric i.e Euclidean distance. Bag-of-words (BoW) histograms (K-means Clustering) were computed, we get the carrier of visual words (sparse vector of frequency counts of a dictionary of image features) using BoW model. Spatial and geometric information associated with descriptions generally don't preserve so BoW is used to retain the information. Descriptors mapped to their corresponding visual words with help of dictionary created using BoW. The consequences then passed to the classifiers to learn the word with its features. We used k = 3 (neighborhood) and N = 24 (labels).

Figure 3.2: SURF Key-point extracted from grayscale image

3.1.3 System Learning (Image to speech)

Supervised learning employed in the proposed system, which require a database of images with their corresponding feature vectors and annotations in order to train the system effectively. Earlier researches conducted was mostly relied on Hidden Markov Models (HMMs) [97], [150], [194]–[196], for recognizing the gestures. Recently various machine learning techniques were introduced that can be used for gesture recognition with less processing and less complexity. Also, Variations in neural network [106], [153] which is deep learning network observed. The proposed system was trained using three different approaches to make decision which is giving better accuracy on the provided sign language dataset. Support Vector Machine (SVM) which is a classifier that find a hyperplanes in order to make separation in data and learn it, Multi-layer Perceptron (MLP) that calculate the gap between actual and predicted score during training and improve it for better performance, Convolutional Neural Network (CNN) is a state-of-the-art network which is only used for classification task but many useful work related to speech also be done by it.

3.1.3.1 Support Vector Machine (SVM)

Originally, SVM is designed is for binary classification (two classes) [197] whereas the data is comprised of multi-classes (twenty-four alphabet gestures). One-against-all multi-class method of SVM is used for training purpose. SVM is capable of finding the possible hyperplane (decision surface) by computing a linear function (non-linear in case of multi-class) to map a sample to its related vector in high dimensional feature space. The classification rule of SVM for the binary class:

$$y_k = \begin{cases} +1, & x_k \in \alpha_1 \\ -1, & x_k \in \alpha_2 \end{cases}$$

…………… eq. (2)

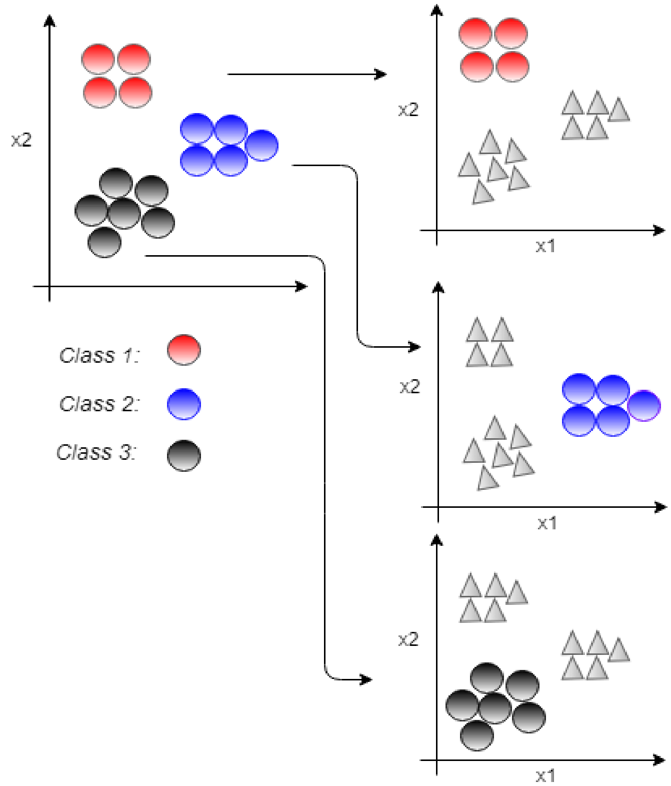

Figure 3.3: SVM multi-classification sample plots

Input vector transforms into feature space using the following equation:

$$f(x) = (w.\sigma(xk)) + b$$

…………… eq. (3)

where σ: ƦƤ → Ƒ and w ∈ Ƒ, while the (.) denotes dot product and conditioned by

$$yk\ (w.\sigma(xk)) + b - 1 >= 0$$

…………… eq. (4)

The binary hyperplanes in high dimensional space F given as [199]

$$\text{Plane1} : w.\sigma(xk) + b = +1$$

$$\text{Plane2} : w.\sigma(xk) + b = -1$$

…………… eq. (5) & eq. (6)

In multi-class SVM, Single large formulation is developed from all input vectors data providing L separate SVM models where L denotes the total classes. SVM configures the multi-class to binary class problem so the number of models obtained can be calculated by L(L − 1)/2) [203].

3.1.3.2 Multi-Layer Perceptron (MLP)

MLP is layered structure whose activities and operations are stimulated from organic neural arrangement of human beings. This model used for classification, non-parametric regression, non-linear function estimation or to simulate the behaviors [228]. A perceptron is a neuron that have weighted information signals (input) and bias to yield a output signal by using activation function.

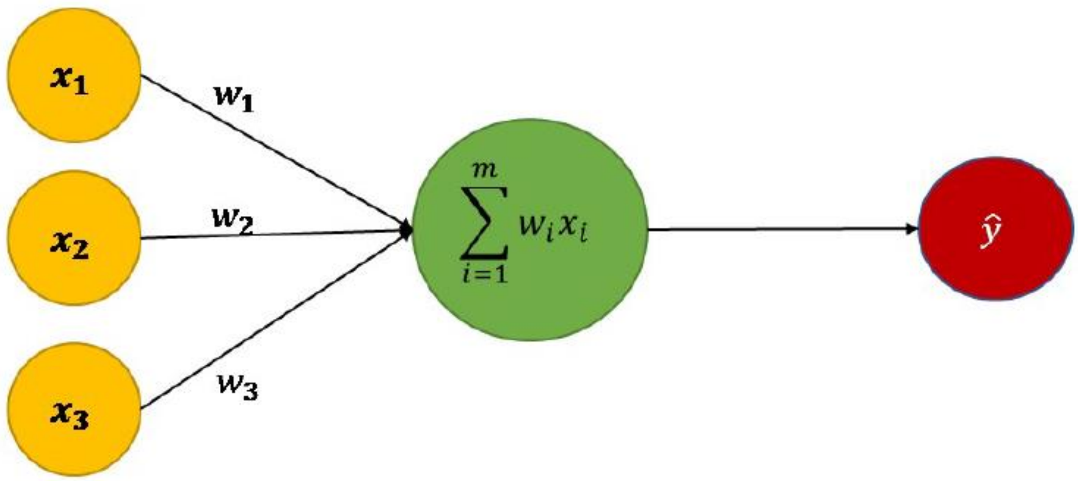

Figure 3.4: Single neuron with its associated layers

The weights of inputs are then summed up and distributed or circulated through an activation function (Hidden layer) which governs the threshold for neuron enactment. The MLP frequently used for classifying the data as they have capacity to learn the pattern with data features and associate it to the output. MLP have propensity to choose highlights of features with temporal or varied resolutions and gather them into higher-order features. MLP neural network, Figure 8, has three layers (i) Input (through which feature vectors are applied) (ii) Hidden layer (where activation function is employed to learn the pattern formation from features) (iii) Output layer (where the predicted score is given associated with features or input provided).

Using input neurons, Data is passed to the MLP network. Hidden layer is the succeeding of the Input layer where learning process takes place using activation function (variety of these functions available including Sigmoid, Relu etc. activates for different ranges), result is then provided at output layer by the network. There can be several input neurons depending on the dataset, similarly the number of neurons at output layer depicts the number of classes to be classify. An iterative method Stochastic gradient descent (SGD) used with MLP to optimize an objective function (minimize cost function) with best suitable properties for training. MLP posses forward-backward process i.e receive input, apply computations, pass results towards output layer refers to forward pass; Error calculated at fully neurons connected layer by comparing the actual and expected result which is back propagated to network instead of giving output, it moves back to update the weights and learn the pattern effectively refers as backward pass. Weights update with respect to their amount of contribution in the calculated error. These processes continuously iterate the training sample (data) until the required epoch or performance threshold reached.

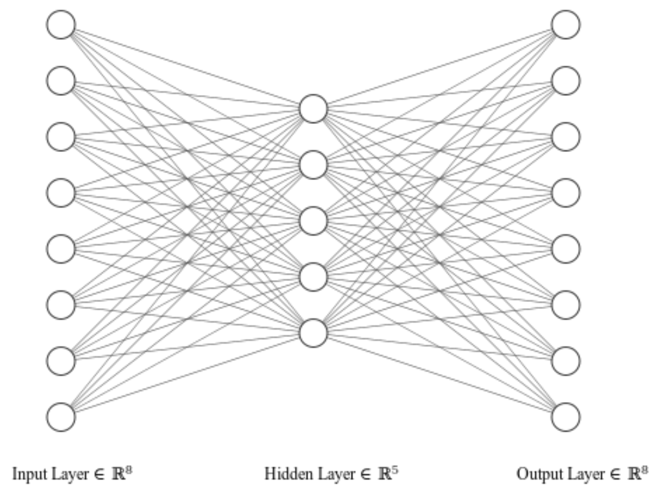

Figure 3.5: MLP Architecture with random units at each layer. Input neurons are connected to each unit in hidden layer and every unit of hidden layer contribute to produce output.

3.1.3.3 Convolutional Neural Network (CNN)

Convolution neural network (CNN) which is a deep learning based classifier received enough admiration from the researchers for recognizing the scenes it has previously viewed (Supervised learning). CNN is good at obtaining the useful and effective features from the image data and c classify the data; therefore manual selection and extraction of data is not needed. The three main factors on which CNN performance rely are (i) local receptive fields (neighborhood area around each neuron; neuron from one layer will connect to only those neuron in next layer which are spatially close) (ii) spatial sub-sampling and activation (non-linearity function, translation, keeps system below computational limit and increase expressive power of model) (iii) weight sharing (forcing the neurons of one layer to share the weights). The features extracted by CNN assure invariance to transformation [229]. CNN has tendency to learn from diverse scenes but not in case of varied databases; Efficiency of network, Computation cost and time involved clearly depends on data. The basic architecture of CNN can be seen in Figure 9.

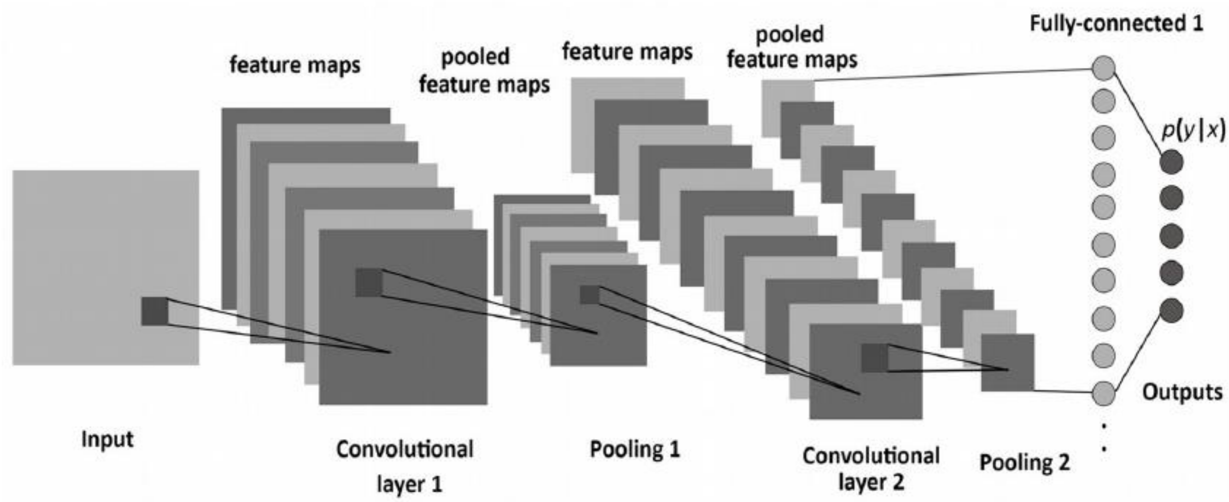

Figure 3.6: Organized CNN basic architecture

An organized CNN can have a set of layers that either can be convolutional, sampling (pooling) or fully connected layer. Other layers can be dropout and additional activation. A convolution operation on input data is applied at convolutional layer, while pooling layer is used to alter (shrink or increase) the spatial size of input (representation) from preceding layer that causes translation of features. The neurons from the fully connected layer are tied up to the succeeding layer neurons. CNN have sparse connectivity among neurons in hidden layers therefore relatively easy to train in contrast of other networks. Convolution layer uses a filter apply over the image (input). The convolution emulate the every single neuron response to feature map (visual stimuli). The number of learnable parameters (computation cost and time complexity) can be reduced (or increase) by using appropriate number of filters, weights and the size of feature vectors.

CNN uses a loss function to minimize the gap found between the actual and predicted score of an input and optimizer to increase the accuracy of the network. CNN parameters chosen on hit and trial for the classification of sign language. Different number of layers (depth of a network) with varied filter sizes were tested to achieve the better prediction. In the proposed system, 2D Convolution layers (depth:4) used with 64 and 32 filters with window size of 3 followed by Max Pooling layer of size 2. Xavier weights are applied to layer, and Rectified Linear Unit (ReLU) activation functions are used for non-linearity. To prevent the overfitting on the sign language dataset which is captured for indoor environment under slightly changed conditions, dropout layers is also used (1.0 and 0.5). Softmax (normalized exponential) function used to minimize the loss of the network. The Softmax loss formulation is followed by equation:

…………… eq. (7) & eq. (8)

$$Loss = \frac{1}{N}\sum_{i=1}^{N} -\log\left(\frac{e^{f_{i},y_{i}}}{\sum_{j=1}^{C} e^{f_{i,j}}}\right)$$

$$f_j(z) = \frac{e^{z_j}}{\sum_{k=1}^{C} e^{z_k}}$$

N = total number of training examples

C = total number of classes

It receive training feature vector z and puts its results (range through 0 to 1, real valued numbers) in vector making a sum of 1, which is the mean loss for every single training vector, xi, to generate the complete loss. Softmax provide result with probability distribution for each gesture in dataset. Adam optimizer which is a variant of SGD (convergence rate, divide learning rate for each weight) is used to achieve the better accuracy; the learning rate (step size or gradient to move towards global minima) used is 10^{-4}. Mini batch learning approach is used to avoid high-biased and high-variance conditions. Network is trained for 20,000 steps.

3.1.4 System Learning (speech to Image)

The proposed system is able to work for signers as well as non-signers. The first module of the system is trained by the researcher whereas functions of second module (speech to image) is performed using the Python API that support conversion of spoken language to the written text (language defined by developer). It takes the voice signal and connect to the API with the accessed token over WiFi and send the voice data to the cloud server. It then receive the speech text extracted from given data. Database of images for twenty-four English alphabets along with their labels (corresponding letter) is maintained by researcher. Speech text is divided on the basis of words which then further split into character array. Program iterate over each letter inside words and look up in the dictionary for the associated gesture image to display on interface with sequences to continue interaction among hearing impaired and non-signer. The flow of the speech to image module can also be understand from the Figure 10 given below:

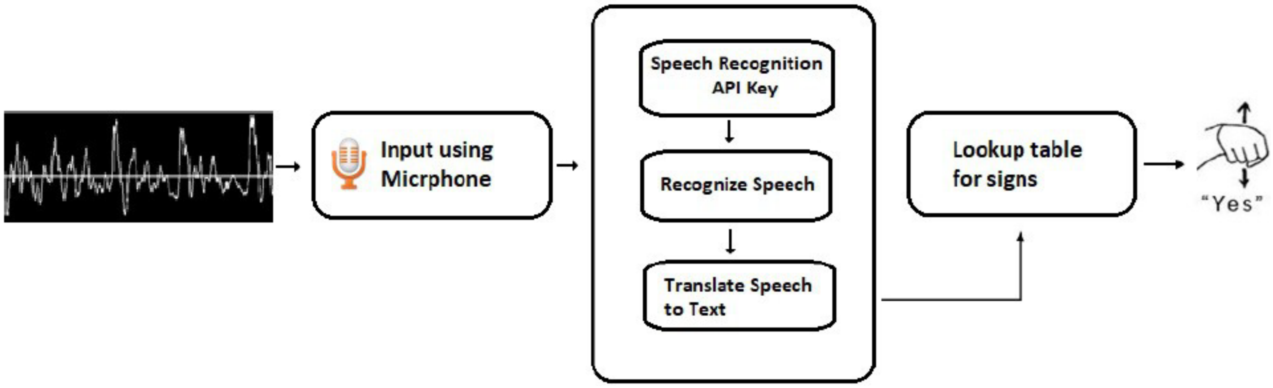

Figure 3.7: The sequence flow for speech to image module

3.2 Algorithmic Flow of the Proposed System

The theoretical overview of the system has been highlighted in the previous sections. The detail of every step including preprocessing (image transformation such as scaling or rotation) , feature engineering (conversion of color modes of the data provided, flattening the image and saving them as vector or extracting the interest point using external SURF detector and descriptor, generating the BoW histograms) and network selection (traditional approaches and deep learning network) were also discussed in the preceding outline. However the abstract details of steps followed to train the different networks (one to choose) are to be given in this section. Algorithm is the set of steps or instructions to be followed which shows the flow of the system. Three approaches used SVM, MLP and CNN; where SVM and MLP provided with surf and non-surf features of the data to observe if the performance of these traditional approaches improves. Whereas CNN was responsible for the whole process itself including the feature engineering (extraction). The concise and sequential flow of the each approach used to build the proposed system with better accuracy is as follow:

(a) Support Vector Machine (SVM) training

<u>Without SURF Features</u>

(i) Resize the image to the size of 96 x 96

(ii) Convert RGB image into grayscale

(iii) Image to Numpy array for future operations and label to categorical floats

(iv) Flatten the image array

(v) Normalizing the image array
(vi) Split the data in train and test sets 70% and 30% respectively
(vii) Pass the data to the SVM ('ovo' decision function shape used) to fit the training set and provide performance score using test set
(viii) Confusion matrix plotted to visualize the performance of classifier

With SURF Features

(i) Resize the image to the size of 96 x 96
(ii) Convert RGB image into grayscale
(iii) Image to Numpy array for future operations and label to categorical floats
(iv) Detect and compute the key-points and descriptors from images using SURF
(v) Generating a cluster model (K-means) using the training instances along with N=24 (number of clusters)
(vi) Split the data (BoW histograms) in train and test sets 70% and 30% respectively
(vii) Pass the data to the SVM (linear kernel were used) to fit the training set and provide performance score using test set.
(viii) Confusion matrix plotted to visualize the performance of classifier

(b) Multi-layer Perceptron (MLP) Network training

Without SURF Features

(i) Resize the image to the size of 96 x 96
(ii) Convert RGB image into grayscale
(iii) Image to Numpy array for future operations and label to categorical floats
(iv) Flatten the image array
(v) Normalizing the image array
(vi) Split the data in train and test sets 70% and 30% respectively
(vii) Pass the data to the MLP to fit the training set and provide performance score using test set
(viii) Confusion matrix plotted to visualize the performance of classifier

With SURF Features

(i) Resize the image to the size of 96 x 96
(ii) Convert RGB image into grayscale
(iii) Image to Numpy array for future operations and label to categorical floats
(iv) Detect and compute the key-points and descriptors from images using SURF

(v) Generating a cluster model (K-means) using the training instances along with N=24 (number of clusters)

(vi) Split the data (BoW histograms) in train and test sets 70% and 30% respectively

(vii) Pass the data to the MLP to fit the training set and provide performance score using test set.

(viii) Confusion matrix plotted to visualize the performance of classifier

(c) Convolutional Neural Network (CNN) training

(i) Resize the image to the size of 96 x 96

(ii) Convert RGB image into grayscale

(iii) Image to Numpy array for future operations and label to categorical floats

(iv) Reshape the input array (all instances) to (-1,96,96,1)

(v) Split the whole set into percentage of 70 and 30 for training and testing respectively

(vi) Prepare a network comprised of convolutional layers (filter number, filter size, activation, kernel initialization, padding) followed with pooling layers (window size) and dropout layer. Flatten layer and fully connected (1024 units), Softmax loss and Adam optimizer.

(vii) Compile and fit the model, iterate over data for 20,000 steps

(viii) Save the model and weights

(xi) Freeze the Tensorflow graph in .protobuf format with node name (placeholder names used during training, reduce model size)

(x) Optimize the graph (reducing the model size therefore can be used in mobile devices as well)

```
---------
Variables: name (type shape) [size]
---------
layer1/kernel:0 (float32_ref 3x3x1x64) [576, bytes: 2304]
layer1/bias:0 (float32_ref 64) [64, bytes: 256]
layer2/kernel:0 (float32_ref 3x3x64x64) [36864, bytes: 147456]
layer2/bias:0 (float32_ref 64) [64, bytes: 256]
layer3/kernel:0 (float32_ref 3x3x64x32) [18432, bytes: 73728]
layer3/bias:0 (float32_ref 32) [32, bytes: 128]
layer4/kernel:0 (float32_ref 3x3x32x32) [9216, bytes: 36864]
layer4/bias:0 (float32_ref 32) [32, bytes: 128]
dense/kernel:0 (float32_ref 1152x1024) [1179648, bytes: 4718592]
dense/bias:0 (float32_ref 1024) [1024, bytes: 4096]
dense_1/kernel:0 (float32_ref 1024x24) [24576, bytes: 98304]
dense_1/bias:0 (float32_ref 24) [24, bytes: 96]
Total size of variables: 1270552
Total bytes of variables: 5082208
```

Figure 3.8: CNN Model summary

Chapter 4: Results

The sign language recognition system being proposed deals with 24 different English alphabets gestures, the complete dataset was comprised of 4500 images from which training and testing set was prepared with the ration of 70% and 30% respectively. The selection was random for the sets belonging to different gestures. Different outputs from each module used with in the proposed system will be outline in this chapter, which will allow the researcher to understand the effective function and network for maximizing the classification performance on sign language dataset. Various techniques were used to make system working with less false prediction, Analysis of results of each implemented technique will be done in this chapter. The performance and accuracy of the proposed system will be analyzed in terms of graphical results (visual plots). For result, two metrics including accuracy on validation or test set and top-5 accuracy (if predicted score lie in the five uppermost score). Confusion matrix, a plot to visualize the performance of a classifier, is generated for the traditional approaches (SVM and MLP) regardless of SURF features were supplied to them or not. Accuracy and loss metrics over epochs were considered to evaluate the performance of a CNN that can be seen in this chapter.

4. 1 Support Vector Machine (SVM) Results

In this section, results of the Support Vector Machine (SVM) classifier will be highlighted with the comparison to itself and other techniques used when provided with global features and local features (extracted using SURF) of sign language dataset and annotations. The accuracy provided by SVM when it was given the non-SURF features was 56%. It was improved by the factor of 27% when given with SURF features of sign language dataset which made the overall accuracy of the proposed system 83%.

The values ranging from 0 to 1 in confusion matrix defines the accuracy. Value close to 1 means accuracy is high. It can be seen in Figure 12 that the convergence is low, the reason could be the similarity between the gestures. For example: A is nearly similar to the S, can be seen in Figure 1. SVM was unable to note the slightly changed feature (Thumb position of A and S) from grayscale image flatten vector (all samples). Whereas high convergence, Figure 13, is performed by SVM when it works with SURF and BoW model as they extract the every viewed feature from the images therefore account the good accuracy.

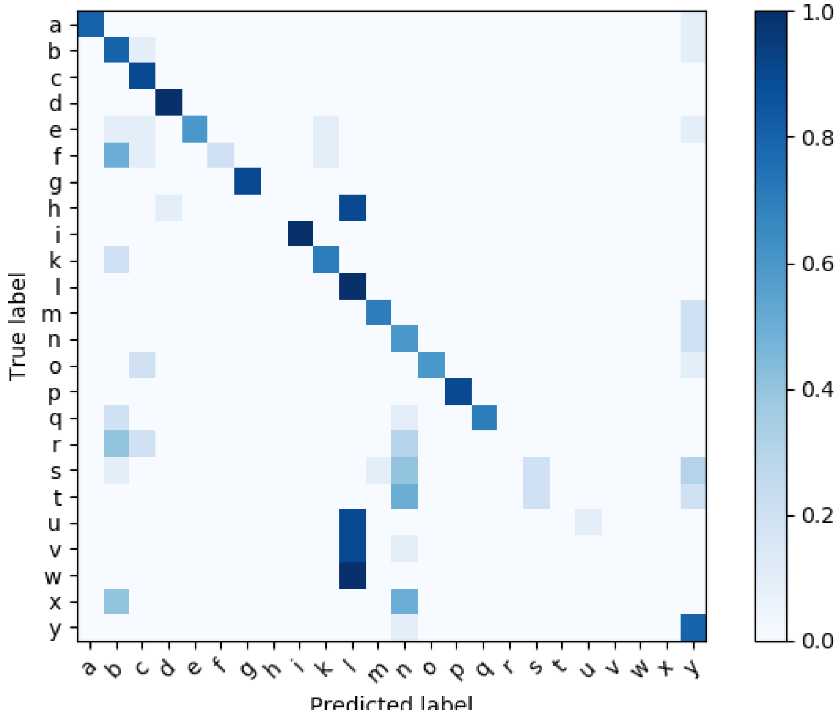

Figure 4.1: Confusion Matrix for SVM classifier when given with non-surf features

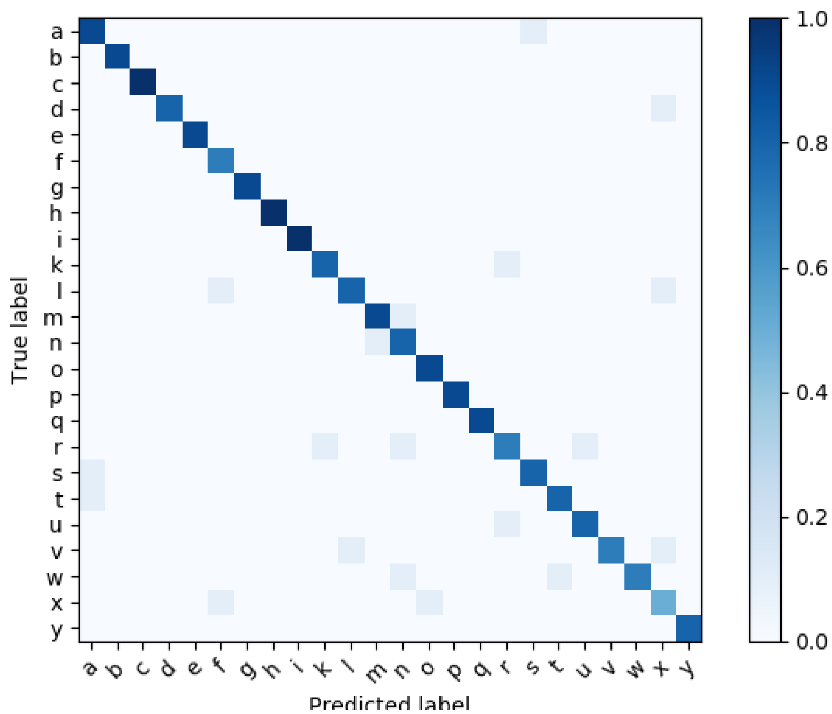

Figure 4.2: Confusion Matrix for SVM classifier when given with surf features

4. 2 Multi-layer Perceptron (MLP) Results

In this section, results of the Multi-layer Perceptron (MLP) classifier will be highlighted with the comparison to itself and other techniques used when provided with global features and local features (extracted using SURF) of sign language dataset and annotations. The accuracy provided by MLP when it was given the non-SURF features was 58%. It was improved by the factor of 28% when given with SURF features of sign language dataset which made the overall accuracy of the proposed system 86%.

The values ranging from 0 to 1 in confusion matrix defines the accuracy. Value close to 1 means accuracy is high. It can be seen in Figure 14 that the convergence is low, the reason could be the similarity between the gestures. For example: N is nearly similar to the T, can be seen in Figure 1. MLP was unable to note the slightly changed feature (Thumb position of N and T) from grayscale image flatten vector (all samples). Whereas high convergence, Figure 115, is performed by MLP when it works with SURF and BoW model as they extract the every viewed feature from the images therefore account the good accuracy.

Groupings	Similarity Feature	Gesture Classes
1	Clenched fist (fully or partially)	*A, S, T*
2	Index finger pointing Diagonally	*G, I*
3	Two fingers pointing with thumb inside diagonally	K, P
4	Pose with two consecutive same fingers	U, H

Table 4.1: Four groups of gesture classes with similar gestures

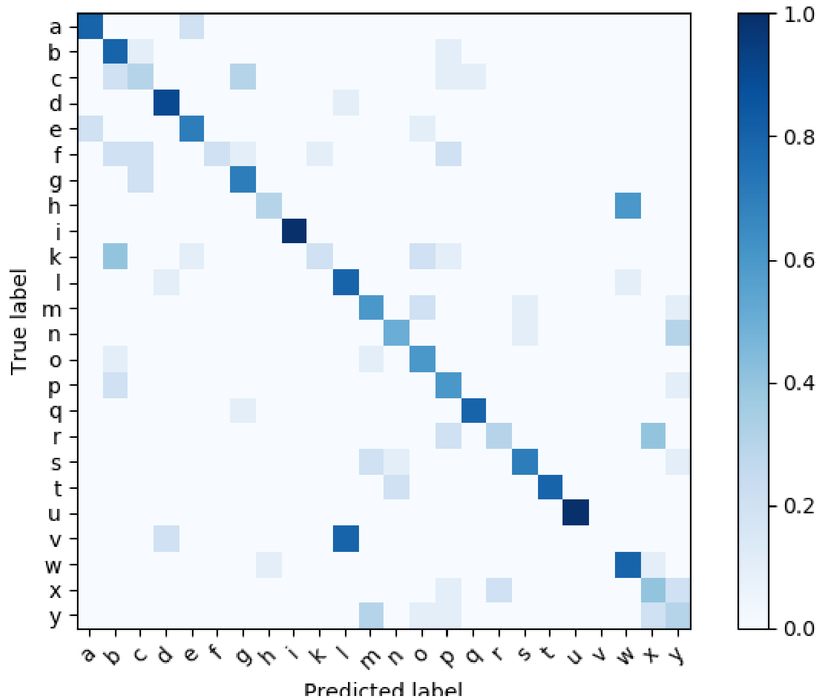

Figure 4.3: Confusion Matrix for MLP classifier when given with non-surf features

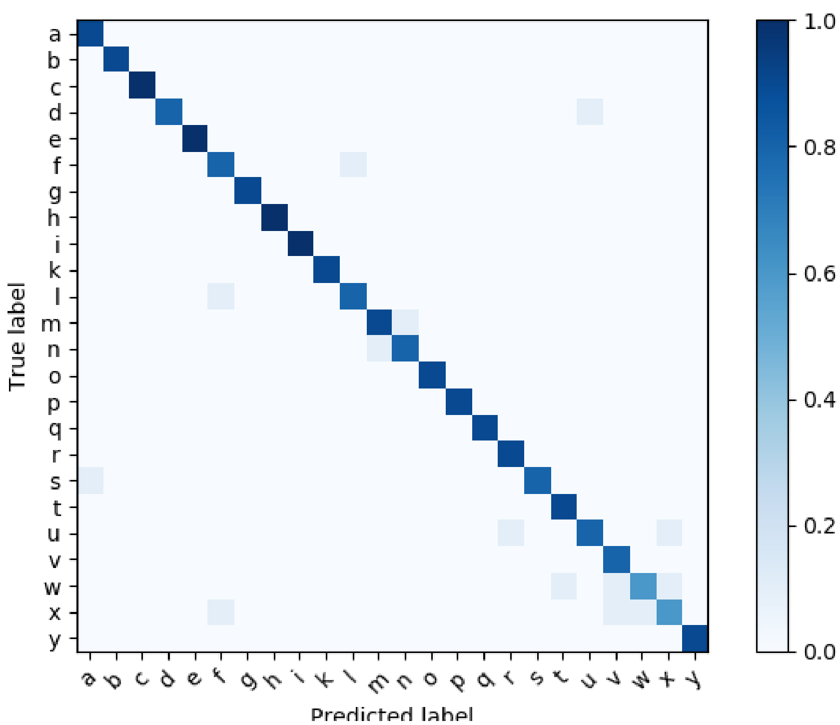

Figure 4.4: Confusion Matrix for MLP classifier when given with surf features

4. 3 Convolutional Neural Network (CNN) Results

The following graph shown in Figure 16 and Figure 17 are for the accuracy and loss results given by CNN. Adam optimizer and Softmax loss function used during the training of the network. CNN is provided with the image dataset of sign language which is of 96 x 96 squared images, Network extract the useful features from the input given to it. It was trained for 20,000 epoch. With every steps, accuracy progressively increased for both training and testing set which can be viewed in Figure 16. Similarly, loss gradually decreased with increase in iterations for all instances that depicts the good performance of the network. CNN was not given any external support, any detector or descriptor for feature extraction, still it produce the better prediction than the traditional approaches. The accuracy achieved by CNN is 94% and loss is 0.00568.

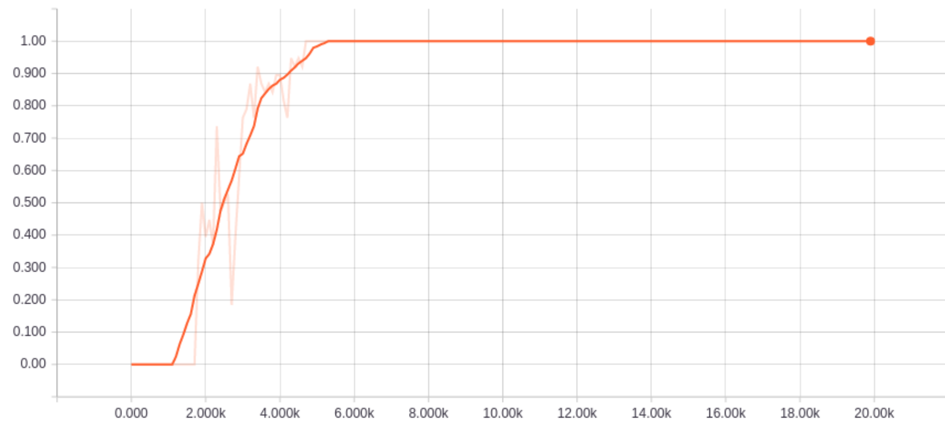

Figure 4.5: CNN Accuracy graph for sign language

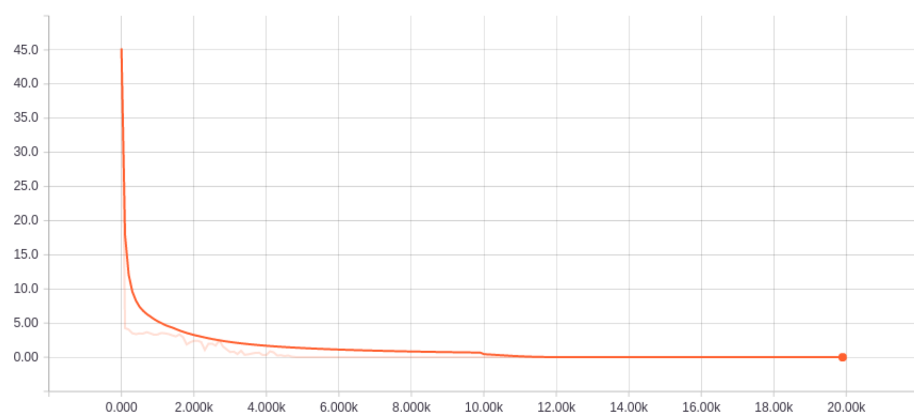

Figure 4.6: CNN loss graph for sign language

4. 4 Proposed System Network Selection

The following bar graph shown in Figure 18 clearly states that CNN has performed well for the classification of sign language with accuracy of 94%. CNN model used in the final developed software application for sign language recognition.

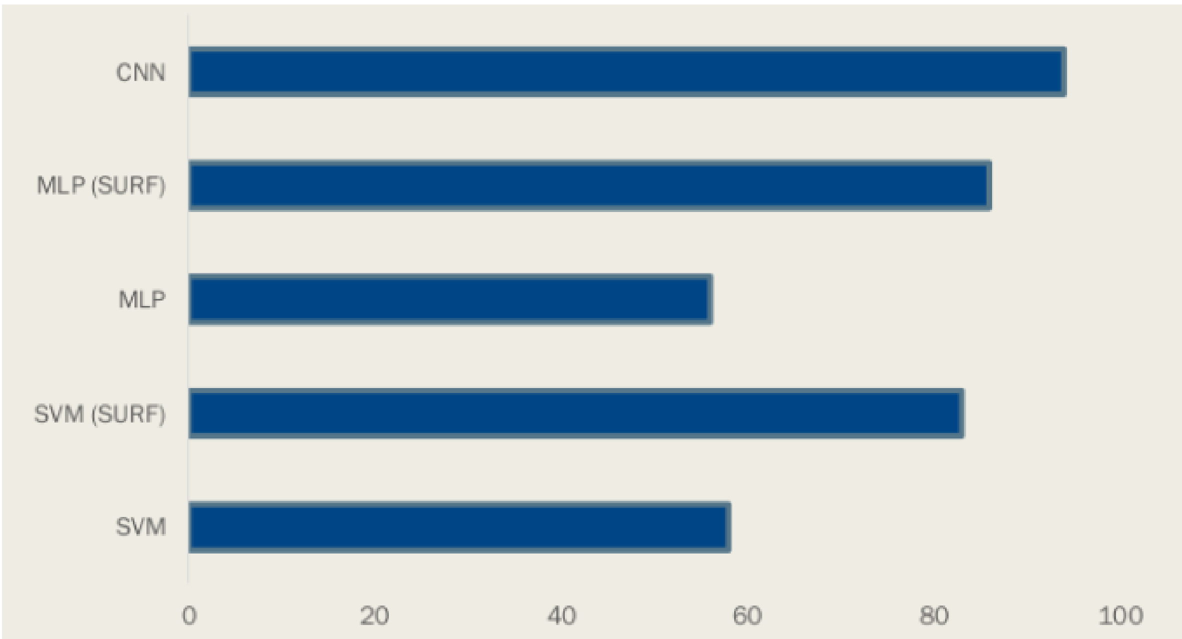

Figure 4.7: Bar graph to show performance of every technique used to develop a sign language recognition system (one technique to choose for proposed system)

Model	**Accuracy**	
	SURF	**Without SURF**
MLP	86%	58%
SVM	83%	56%
CNN	94% -- CNN extracted abstract features by itself (No SURF applied)	

Table 4.2: Accuracy of each model used

Chapter 5: Conclusion and Future Works

The final chapter of this research in which all the findings, shortcomings and challenges with focus on recognition of sign language will be concluded for the better understanding of the future researchers that will enable them to take decision about choosing the network and techniques for any of the task as per the characteristics of the dataset. This research lay the foundations in the future developers of the sign language recognition system. Summary of the previous chapters will be listed in this chapter. The limitations exist in the proposed system will also be outlined and the methods or endeavor need to improve the system will also be discussed in this chapter. Currently the proposed system is designed for indoor communication or education purpose for the deaf community (children) which can be improved to implement in society at large scale.

5.1 Review of Previous Chapters

In chapter 1, the most significant aspects of the gesture language learning and hard of hearing people education were mentioned related to PSL and ASL, where PSL is the nearby communication language among deaf community in Pakistan for which the system is proposed and ASL was picked as benchmark with respects to the extensive exertion, educational initiatives, activities and efforts in the US that at last advances deaf education, and social incorporation inside the US communities, empowering ASL to turn into the fourth-most showed language in the nation. The education effort for deaf and promotion of sign language in Pakistan is insufficient and inadequate, however there is requirement for improving national access towards education for the hearing impaired people, which will cultivate the society towards betterment, and reduce the problem of social inclusion, consideration and acceptability for the deaf community of Pakistan.

Chapter 2 gives a comprehensive list of various sources related to the classification or recognition of the sign language. The different applications developed by researchers that functions with gestures and the fields of research in which such frameworks have been utilized have also examined. The literature led us to the finding that gestures have well communicative property and gestures are not only assisting the hearing impaired community but also intelligent and autonomous systems have been developed that received the commands from the non-verbal language and execute activities. Therefore, effective gesture recognition with less tolerance is still active in research.

The third chapter of this research concentrate on briefing the methodology used in the proposed recognition system. The diverse equipment and programming or algorithm varieties delineated in the literature have been reviewed and examined, with the goal of incorporating the suitable technique in this research. The model of the proposed system has been appropriately discussed, while the reason for using the adopted procedure has also been given.

The results obtained of every technique is evaluated in the fourth chapter. The performance and accuracy of the model chosen for the proposed system provided in the fourth chapter.

5.2 Limitations

The proposed system for sign language recognition has some significant constraints like all practical systems. Usually systems are developed with a particular arrangement of parameters, settings, context and restricting conditions at which the system test and works with reasonable standard. One of the previous constraints inside PSL research was the absence of a standardized PSL database containing pictures and video arrangements for static and dynamic signs indicating different hand shapes. The list of improvements or constraints are: (i) system designed for images, can be developed for the video sequences (ii) system works for 24 English static gestures of alphabets, can be trained for digits and Urdu alphabets as well (iii) dataset captured for indoor environment with uncluttered background, dataset can be updated with the images of diverse background and light conditions (iv) accuracy is not guaranteed for the changed or improved cases.

5.3 Current Research and its Implications

The developed recognition system is provided with standard image input to predict the 24 alphabet gesture for English that originated by PSL. The input must be of uncluttered background with clear gesture so that information can be extracted effectively. The ability to recognize the gesture is 94% in the system developed. The system has two phases (i) for the signer, to interpret the gesture into spoken language (ii) for the non-signer, spoken language to the hand gesture display. The system is able to function in noise-free (for non-signer) indoor (for deaf) environment. The system can be considered as critical but transitional step in developing the real-time PSL recognition system that can be implementable in the dynamic environment. The proposed system is a medium for hearing impaired and non-signers to interact with each other and overcome the barriers. However, system is currently not enough mature to reduce social stigmatization but can be used for deaf children as a learning source.

5.4 Direction of Future Research

This research will be pursued with the intense focus to facilitate the deaf community and can open new aspects of research. while efforts to eliminate the different lacks of the proposed system can be done by the researcher and future endeavors. A varied feature sets can be examined on the system to reduce the misclassification rate and achieve more satisfaction about implementation of the system. The methodologies and appropriate ways can be explored to overcome the problem of the comparable and wandering pattern among some gestures. The time and computation complexity along with response time can be improve by integrating the appropriate equipment or hardware. With time, endeavors will be made to gather video-based sign language data from enormous number of native and non-native deaf individuals and incorporate some advancement in the recognition system of sign language that enable the system to work in dynamic environment with non-fixed position or shaped signs in real-time. The effort will be done on expansion of lexicon dictionary of PSL and maintaining a database. A significant effort should be made for the optimization of very module used for PSL recognition system, enable system to adapt from environment and should be flexible enough for the future improvement.

References

[1] H. Van der Hulst and A. Mills, “Issues in sign linguistics: Phonetics, phonology, and morpho-syntax,”*Lingua*, vol. 98, pp. 3–17,1996.

[2] F.R.Lin,J.K.Niparko,andL.Ferrucci,“Hearing loss prevalence in the United States,” *Archives of Internal Medicine*, vol. 171, no. 20, pp. 1851–1852, 2011.

[3] S. Lidell,*Grammar, Gesture and Meaning in American Sign Language*. London: Cambridge University Press,2003.

[4] M. Huenerfauth, “American Sign Language Generation: Multimodal NLG with Multiple Linguistic Channels,” in *Proceedings of ACL Student Research Workshop*, Ann Arbor, Michigan,2005b.

[5] M. Huenerfauth, “American Sign Language Spatial Representation for an Accessible User-Interface,” presented at the 3rh International Conference on Universal Access in Human-Computer Interaction, Las Vegas, Nevada, 2005a.

[6] N. S. Khan, A. Abid, K. Abid, U. Farooq, M. S. Farooq, and H. Jameel, “Speak Pakistan: Challenges in Developing Pakistan Sign Language using Information Technology,”*South Asian Studies*, vol. 30, no. 2, pp. 367–379, 2015.

[7] L. Simms, S. Baker, and M. D. Clark, “The Standardized Visual Communication and Sign Language Checklist for Signing Children,”*Sign Language Studies*, vol. 14, no. 1, pp. 101–124, 2013.

[8] B. Tervoort,*Structure le analyse van visueel taalgebruik binnen een grope dove kinderen*. Amsterdam: North Holland Publishing Company,1953.

[9] W. C. Stokoe,*Sign language structure: An outline of the visual communication systems of the American Deaf*. Silver Spring, MD: Linstok Press, 1960.

[10] K.E.Conlin, G.R.Mirus, M.Claude,and R.P.Meier,*The acquisition of first signs:Place, hand shape and movement*. Mahwah, NJ: Lawrence Erlbaum,2000.

[11] G. Morgan and B. Woll, Eds.,*Directions in sign language acquisition*. Amsterdam: Benjamins,2002.

[12] R. P. Meier, “The acquisition of verb agreement: Pointing out arguments for the linguistic status of agreement in signed languages,”in *Directions in sign language acquisition*, M.G. Woll, Ed. Amsterdam: Benjamins,2002.

[13] J. Kegl, A. Senghas, and M. Coppola, “Creation through contact: sign language emergence and sign language change in Nicaragua,” in *Language creation and language change:creolization, diachrony, and development*, M. DeGraff, Ed. Cambridge, MA: MIT Press, 1999.

[14] R. Pfau, *Hand waving and head shaking? On the linguistic structure of sign languages*. Barcelona: Institute d’Estudis Catalans,2010.

[15] W. C. Stokoe,“Sign Language Structure:An Outline of the Visual Communication Systems of the American Deaf,”*Journal of Deaf Studies and Deaf Education*, vol. 10, no. 1, pp. 3– 37, 2005.

[16] G. Marsaja, *DesaKolok-A deaf village and its sign language in Bali, Indonesia*. Mijmegen: Ishara Press,2008.

[17] V. Nyst, “A descriptive analysis: Adamorobe Sign Language (Ghana),” University of Amsterdam, Amsterdam,2007.

[18] B.Akram and R.Bashir,“Special Education and Deaf Children in Pakistan: An Overview,” *Journal of Elementary Education*, vol. 22, no. 2, pp. 33–44.
[19] A. A. Haseeb and A. Illyas, “Speech Translation into Pakistan Sign Language,” Blekinge Institute of Technology, Karlskrona, Sweden, 2012.
[20] Family Educational Services Foundation (FESF) Annual Report, “FESF Annual Report 2013-2014.”2014.
[21] S. Wilcox and P. P. Wilcox,*Learning to See: Teaching American Sign Language as a Second Language*. Washington, DC: Gallaudet University Press,1997.
[22] W. C. Stokoe, C. Croneberg, and D. Casterline, *A dictionary of American Sign Language on linguistic principles*. Washington, DC: Gallaudet University Press,1965.
[23] E. Klima and U. Bellugi,*The Signs of Language*. Cambridge, MA: Harvard University Press,1979.
[24] R. L. Cooper, *Language planning and social change*. Cambridge: Cambridge University Press,1989.
[25] G. Ashton, K. Cagle, K. B. Kurz, W. Newell, R. Peterson, and J. E. Zinza, *Standards for Learning American Sign Language*. Washington, DC: American Sign Language Teachers Association.
[26] J. G. Kyle,*Sign and school*. Cambridge, MA: Harvard University Press,1987.
[27] M. Strong,*Language and deafness*. London: Cambridge University Press,1988.
[28] National Institute on Deafness and Other Communication Disorders (NIDCD) Report, “Strategic Plan Fiscal Year 2006-2008,” NIDCD, Rockville, MD,2008.
[29] M. Strong and P. Prinz, “Is American Sign Language skill related to English literacy?,” in *Language Acquisition by Eye*, C. Chamberlain, J. P. Morford, and R. I. Mayberry, Eds. Mahwah, NJ: Lawrence Erlbaum,2000.
[30] M. Strong and P. Prinz, “A study of relationship between American Sign Language and English literacy,”*Journal of Deaf Studies and Deaf Education*, vol. 2, pp. 37–46,1997.
[31] P. Prinz and M. Strong, “ASL proficiency and English literacy within a bilingual deaf education model of instruction,”*Topics in Language Disorders*, vol. 18, no. 4, pp. 47–60, 1998.
[32] S. N. Mahshie,*Educating Deaf children bilingually. With insights and applications from Sweden and Denmark*. Washington, DC: Gallaudet UniversityPress.
[33] S.S.Gibson, A.Small, and D.Mason, “Deaf bilingual bicultural education,”in *Encyclopedia of languge and education.*, Dordrecht, Netherlands: Kluwer Publishers,1997.
[34] S.Goldwin-Meadow and R.I.Mayberry, “How do profoundly deaf children learn to read?,” *Learning Disabilities Research & Practice*, vol. 16, no. 4, pp. 222–229, 2001.
[35] R.I.Mayberry and E. Lock, “Age constraints on first versus and second language acquisition:Evidence for linguistic plasticity and epigenesis,”*Brain and Language*,vol.87, pp. 369–384, 2003.
[36] S. P. Quigley and P. Paul,*Language and deafness*. San Diego, CA: College-Hill Press, 1984.
[37] R. B. Wilbur,*American Sign Language: Linguistics and applied dimensions*. San Diego, CA: College-Hill Press,1987.
[38] R.B.Wilbur,“The Use of ASL to Support the Development of English and Literacy,” *Journal of Deaf Studies and Deaf Education*, vol. 5, no. 1, pp. 81–104, 2000.

[39] R. B. Wilbur,*Readable English for hearing-impaired students: Multi-level guidelines for linguistically controlled reading materials*. Boston, MA: WGBH Educational Foundation, 1980.
[40] R. B. Wilbur, W. C. Goodhart, and D. R. Fuller, “Comprehension of English modals of hearing-impaired students,”*Volta Review*, vol. 91, pp. 5–18,1989.
[41] P. Spencer, “Language at 12 and 18 months: characteristics and accessibility of linguistic models,”in *The world of deaf in fants: A longitudinal study,* K.Meadow-Orlans, P.Spencer, and L.Koester, Eds. New York, NY: Oxford University Press,2004.
[42] M.Magnuson,“Infants with congenital deafness: On the importance of early sign language acquisition,”*American Annals of the Deaf*, vol. 145, no. 1, p. 6,2000.
[43] B. Schick, J. de Villiers, P. de Villiers, and B. Hofmeister, “Theory of mind: Language and cognition in deaf children,”*The ASHA Leader*, vol. 7, no. 22, pp. 6–14,2002.
[44] V.Volterra, J.Iverson, and M.Castrataro, “The development of gesture in hearing and deaf children,” in *Advanced in the sign language development of deaf children*, B. Shick, M. Marschark, and P. Spencer, Eds. New York, NY: Oxford University Press,2006.
[45] T. Humphries, “Schooling in American Sign Language: A Paradigm Shift from a Deficit Model to a Bilingual Model in Deaf Education,”*Berkeley Review of Education*, vol. 4, no. 1, pp. 7–33, 2013.
[46] A.K.Alvi, M.Y.B.Azhar, and U.Mehmood, “Pakistan Sign Language Recognition Using Statistical Template Matching,”*International Journal of Information Technology*, vol. 1, no. 1, pp. 1–12.
[47] H. Tauseef, M. A. Fahiem, and S. Farhan, “Recognition and Translation of Hand Gestures of Urdu Alphabets Using Geometric Classification,” presented at the Second International Conference in Visualization, 2009, pp.213–217.
[48] Family Educational Services Foundation (FESF), “Pakistan Sign Language Alphabets,” 2013. .
[49] N. Dang-Binh, E. Shuichi, and T. Ejima, “Real-Time Hand Tracking and Gesture RecognitionSystem,”in*ProceedingsofJournalofGraphics,VisionandImageProcessing*, 2005, pp. 19–21.
[50] C. Manresa, J. Varona, R. Mas, and F. Perales, “Real-Time Hand Tracking and Gesture Recognition for Human-Computer Interaction,”*Electronic Letters on Computer Vision and Image Analysis*, vol. 5, no. 3, pp. 96–104,2005.
[51] X.Liu and K.Fujimura, “Hand gesture recognition using depth data” in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, 2004,pp. 529–544.
[52] A. Angelopoulou, J. Garcia-Rodgriquez, and A. Psarrou, “Learning 2D Hand Shapesusing theTopologyPreservingmodelGNG,”in*Proceedings of European Conference onComputer Vision*, 2006, pp.313–324.
[53] Y.Fang, K.Wang, J.Cheng, and H.Lu, “A Real-Time Hand Gesture Recognition Method,” in *Proceedings o 2007 IEEE International Conference on Multimedia and Expo*, 2007, pp. 995–998.
[54] Q. Chen, N. D. Georganas, and E. M. Petriu, “Real-time Vision-based Hand Gesture Recognition using Haar-like Features,” in *Proceedings of Instrumentation and Measurement Technology Conference*,2007.

[55] D. Franklin, R. E. Kahn, M. J. Swain, and R. J. Firby, "Happy patrons make better tippers- Creating a robot waiter using Perseus and the animate agent architecture," in *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 1996, pp.14–16.

[56] D.Guo, Y.H.Yan, and M.Xie, "Vision-based hand gesture recognition for human-vehicle interaction,"in *5th International Conference of Control, Automation, Robotics and Visions*, 1998, pp. 151–155.

[57] R. Cipolla and N. J. Hollinghurst, "Human-robot interface by pointing with uncalibrated stereo vision,"*Image and Vision Computing*, vol. 14, no. 3, pp. 171–178,1996.

[58] D. Kortenkamp, E. Huber, and R. P. Bonasso, "Recognizing and interpreting gestures on a mobile robot," in *Proceedings of Artificial Intelligence Workshop*, 1996, pp.915–921.

[59] S. Waldherr, R. Romero, and S. Thrun, "A gesture-based interface for human-robot interaction,"*Autonomous Robots*, vol. 9, no. 2, pp. 151–173, 2000.

[60] X. Yin and M. Xie, "Estimation of the fundamental matrix from uncalibrated stereo hand images for 3-D hand gesture information,"*Pattern Recognition*, vol.36, no.3, pp.567–584, 2003.

[61] S. S. Ghidary, M. Hattori, S. Tadokoro, and T. Takamori, "Multi-modal human robot interaction for map generation," in *Proceedings of IEEE Intelligent Robots and Systems*, 2001, pp. 2246–2251.

[62] S. Sato and S. Sikane, "A human-robot interface using an interactive hand pointer that projects a mark on the real work space," in *Proceedings of the IEEE InternationalConference on Robotics and Automation*, 2000, pp. 589–595.

[63] J. P. Wachs, H. Stern, and Y. Edan, "Cluster Labeling and Parameter Estimation for Automated Setup of a Hand-Gesture Recognition," *IEEE Transaction on Systems, Manand Cybernetics- Part a: Systems and Humans*, vol. 35, no. 6, pp. 932–944,2005.

[64] M. Correa, J. Ruz-del-Solar, J. Verchae Lee-Ferng, and N. Castillo,*Real-Time HandGesture Recognition for Human-Robot Interaction*. Berlin: Springer-Verlag,2009.

[65] R. Posada-Gomez, L.-H. Sanchez-Medel, G. A. Hernandez, and A. Martinez-Sibaja, "A Hand Gesture System of Control for An Intelligent Wheelchair," in *Proceedings of 4thInternational Conference on Electrical and Electronics Engineering*, 2007, pp.68–71.

[66] M. Hashimoto, K. Takahashi, and M. Shimada, "Wheelchair control using an EOG- and EMG-based gesture interface,"in *Proceedings of IEEE/ASME International Conference on Advanced Mechatronics*, 2009, pp.1212–1217.

[67] L. Montesano, J. Minguez, J. Alcubierre, and L. Montano, "Towards the adaptation of a robotic wheelchair for cognitive disabled children," presented at the IEEE International conference on Intelligent Robotic Systems, 2006, pp.710–716.

[68] H.-J. Bohme, T. Wilhelm, J. Key, C. Schauer, C. Schroter, H.-M. Grob, and T. Hempel, "An approach to multi-modal human-machine interaction for intelligent service robots,"*Robotics and Autonomous Systems*, vol. 44, pp. 83–96,2003.

[69] A. Weitzenfeld, C. Ramos, and P. Dominey, "Coaching Robots to Play Soccer via spoken langauge,"in*Robocup2008:Robot Soccer World Cup XII*,Berlin:Springer Verlag,2009.

[70] X.-H. Wu, M.-C. Su, and P.-C. Wang, "A hand-gesture-based control interface for a car-robot," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 4644–4648.

[71] Y. Muhammad and T. Nishida, "Learning interaction protocols using augmented bayesian networks applied to guided navigation," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 4119–4126.

[72] M. van den Berg, D. Carton, R. D. Nijs, N. Mitsou, C. Landsiedel, K. Kuehnlenz, D. Wollherr, L.V. Gool, and M. Buss, "Real-time 3D Hand Gesture Interaction with a Robot for Understanding Directions from Humans," in Proceedings *of IEEE International Symposium on Robot and Human Interactive Communication*, 2011, pp.357–362.

[73] F.Trujillo-Romero, F.E.Luis-Perez, and S.O.Caballero-Morales,"Multimodal Interaction for Service Robot Control," presented at the 22nd International Conference on Electrical Communications and Computers, 2012, pp.305–310.

[74] D.Xu, Y.-L.Chen, C.Lin, X.Kong, and X.Wu,"Real-time Dynamic Gesture Recognition System based on Depth Perception for Robot Navigation," in *Proceedings of 2012 IEEE International Conference on Robotics and Biomimetics*, 2012, pp.689–694.

[75] P. Evrard, E. Gribovskava, S. Calinon, A. Billard, and A. Kheddar, "Teaching Physical Collaborative Tasks: Object Lifting Case Study with a Humanoid,"in *Proceedings of IEEE International Conference on Humanoid Robots*, 2009, pp. 399–404.

[76] O. Stasse, P. Evrard, N. Perrin, N. Mansard, and A. Kheddar, "Fast foot prints re-planning and motion generation during walking in physical human-humanoid interaction," in *Proceedings of 9th IEEE-RAS International Conference on Humanoid Robots*, 2009, pp. 284–289.

[77] K. Yokoyama, H. Handa, T. Isozumi, Y. Fukase, K. Kanedo, F. Kanehiro, Y. Kawai, F. Tomita, and H. Hirukawa, "Cooperative Works by a Human and a Humanoid Robot," in *Proceedings of International Conference on Robotics and Automation*, 2003, pp. 2985–2991.

[78] K. Sohn, Y. Kim, and P. Oh, "2-Tier Control of a Humanoid Robot and Use of Sign Language Learned by Monte Carlo Method," presented at the The 22nd IEEE IEEE International Symposium on Robot and Human Interactive Communication ,2013,pp.547–552.

[79] A.M.Elgammal, V.D.Shet, Y.Yacoob, and L.S.Davis,"Learning dynamics for exemplar- based gesture recognition," in *Proceedings of Computer Vision and Pattern Recognition*, 2003, pp. 571–578.

[80] S. B. Wang, A. Quattoni,L.P. Morency, and D. Demirdijian, "Hidden conditional random fields for gesture recognition," presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, pp.1521–1527.

[81] L. Bretzner, I. Laptev, and T. Lindeberg, "Hand Gesture Recognition using Multi-Scale Color Features, Hierarchical Models and Particle Filtering," in *Proceedings of Face and Gesture*, 2002, pp.423–428.

[82] C.-T. Hsieh, C.-H. Yeh, K.-M. Hung, L.-M. Chen, and C.-Y. Ke, "A Real Time Hand Gesture Recognition System Based on DFT and SVM," presented at the 8th International Conference on Information Science and Digital Content Technology, 2010, pp.490–494.

[83] J.Liu, Z.Wang, L.Zhong, J.Wickramasuriya, and V.Vasudevan, "uWave: Accelerometer- based personalized gesture recognition and its applications," in *Proceedings of IEEE International Conference on Pervasive Computing and Communications*, 2009, pp.1–9.

[84] S.-H. Lee, M.-K. Sohn, D.-J. Kim, and H. Kim, "Smart TV Interaction System Using Face and Hand Gesture Recognition," presented at the IEEE International Conference on Consumer Electronics, 2013, pp.173–174.

[85] M. R. Abid,L.B. S. Melo, and E. M. Petriu, "Dynamic Sign Language and Voice Recognition for Smart Home Interactive Application," presented at the IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2013, pp.139–144.

[86] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-Home Gesture Recognition using Wireless Signals," presented at the MobiCom'13, Miami, FL,2013.
[87] H. Lahamy and D. Litchi, "Real-time hand gesture recognition using range cameras," in *Proceedings of the 2010 Canadian Geomatics Conference and Symposium of Commission I*, Calgary, Canada, 2010, pp.38–40.
[88] S.Reifinger, F.Wallhoff, M.Ablassmeier, T.Poitschke, and G.Rigoll,"Static and Dynamic Hand-Gesture Recognition for Augmented Reality Applications," in *Human-Computer Interaction*, J. Jacko, Ed. Berlin: Springer-Verlag,2007.
[89] J. Lambrecht and J. Kruger, "Spatial Programming for Industrial Robots based on Gestures and Augmented Reality,"presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp.466–472.
[90] M. Kolsch, R. Bane, T. Hollerer, and M. Turk, "Multi-modal Interaction with a Wearable Augmented Reality System,"*IEEE Computer Graphics and Applications*,pp.62–71,2006.
[91] R.Lo,A.Chen, V.Rampersad, J.Huang, H.Wu, and S.Mann,"Augmented reality system based on 3D camera self-gesture sensing," in *IEEE International Symposium on Technology*, 2013, pp. 20–31.
[92] E.Kollorz ,J.Penne, J.Hornegger, and A.Barke,"Gesture recognition with a time-of-flight camera,"*International Journal of Intelligent Systems Technologies and Applications*, vol. 5, no. 3/4, pp. 334–343, 2008.
[93] H. Hikawa and Y. Araga, "Study on Gesture Recognition System using Posture Classifier and Jordan Recurrent Neural Network," in *Proceedings of International Joint Conference on Neural Networks*, 2011, pp.405–412.
[94] L. P. Morency, A. Quattoni, and T. Darrell, "Latent-dynamic discrminative models for continuous gesture recognition," in*Proceedings of IEEE Conference on Computer Visionand Pattern Recognition*, 2007, pp.1–8.
[95] C. Padden and T. Humphries,*Deaf in America: Voices from a culture*. Cambridge, MA: Harvard University Press,1988.
[96] Y. Quan, P. Jinye, and L.Yulong, "Chinese sign language recognition based on gray-level co-occurence matrix and other multi-feature fusion,"in *Proceedings of IEEE Conference on Industrial Electronics and Applications*, 2009, pp.1569–1572.
[97] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk- and wearable computer-based video," in *Proceedings of IEEE Transactions ofPatterns Analysis and Machine Intelligence*, 1998, pp.1371–1375.
[98] D.Uebersax, Gall,M.vanden Berg, and L.vanGool,"Real-time sign language letter and work recognition using depth data," in *Proceedings of IEEE International Conference onComputer Vision Workshop*, 2011, pp.383–390.
[99] H. Cooper, E.-J. Ong, N. Pugeault, and R. Bowden, "Sign language units using sub-units," *Journal of Machine Learning Research*, vol. 13, pp. 2205–2231, 2012.
[100] K. Pattanaworapan, K. Changongthai, and J.-M. Guo, "Hand Gesture Recognition Using Code book Model and Pixel-Based Hierarchical-Feature Adaboosting," in *Proceedings of13th International Symposium on Communications and Information Technology*, 2013, pp. 544–548.
[101] M.Zahedi, D.Keysers, and H.Ney,"Pronunciation clustering and modeling of variability for appearance-based sign language recognition," in *Proceedings of Gesture in Human-Computer Interaction and Simulation*, 2006, pp.68–79.T. Kadir, R. Bowden, E.-J. Ong, and A. Zisserman, "Minimal training, large lexicon, unconstrained sign language recognition," in *Proceedings of the British Machine Vision Conference*, 2004, pp.939–948.

[102] P. Dreuw, T. Deselaers, D. Rybach, D. Keysers, and H. Ney, "Tracking using dynamic programming for appearance based sign language recognition," in *Proceedings of 7thInternational Conference on Face and Gesture Recognition*, 2006, pp.293–298.
[103] J. Marnik, "The polish finger alphabet and posture recognition using elastic graph matching,"*Advances in Soft Computing*, vol. 2, no. 45, pp. 454–461, 2007.
[104] I.Incertis, J.Garcia-Bermejo, and E.Casanova,"Hand Gesture Recognition for Deaf People Interfacing," in *Proceedings of 18th IEEE International Conference on Pattern Recognition*, 2006, pp.100–103.
[105] Y. Gweth, C. Plahl, and H. Ney, "Enhanced continuous sign language recognition using PCA and neural network features," in *CVPR Workshop on Gesture Recognition*,2012.
[106] M. Al-Roussan, K. Assaleh, and A. Tala'ac, "Video-based signer-independent Arabic sign language recognition using Hidden Markov Models,"*Applied Soft Computing*, vol. 9, no. 3, pp. 990–999.
[107] N. R. Albelwi and Y. M. Alginahi, "Real-Time Arabic Sign Language (ArSL) Recognition," in *Proceedings of International Conference on Communications and Information Technology*, 2012, pp.497–501.
[108] M. Mohandes and M. Deriche, "Arabic Sign Language Recognition by Decisions Fusion using Dempster-Shafer Theory of Evidence," in*Proceedings of Computer Communication and IT Application Conference*, 2013, pp.90–94.
[109] R.J.K.Mahesh,S.Mahishi,R.Dheeraj,S.Sudheender,andN.V.Pujari,"FingerDetection for Sign Language Recognition," in*Proceedings of the International Multi-Conference ofEngineers and Computer Scientists*, 2009, pp.1–5.
[110] D. Tiwari and S. J. Srivastava, "A Visual Recognition of Static Hand Gestures in Indian Sign Language based on Kohonen Self-Organizing Map Algorithm,"*International Journal of Engineering and Advanced Technology*, vol. 2, no. 2, pp. 165–170, 2012.
[111] P. R. Futane and R. V. Dharaskar, "Video Gestures Identification and Recognition Using Fourier Descriptors and General Fuzzy Minmax Neural Network for Subset of Indian Sign Language,"in*Proceedingsof12thInternationalConferenceonHybridIntelligentSystems*, 2012, pp. 525–530.
[112] J. SIngha and K. Das, "Indian SIgn Language Recognition Using Eigen Value Weighted Euclidrean Distance Based Classification Technique,"*International Journal of AdvancedComputer Science and Applications*, vol. 4, no. 2, pp. 188–195,2013.
[113] V.Adithya, P.R.Vinod, and U.Gopalakrishnan,"Artificial Neural Network Based Method for Indian Sign Language," in *Proceedings of 2013 IEEE Conference on Information and Communication Technologies*, 2013, pp.1080–1085.
[114] R. Zhou, J. Yuan, and Z. Zhang, "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera," in *Proceedings of the 19th ACM International Conference on Multimedia*, 2011, pp. 1093–1096.
[115] A. Agarwal and M. K. Thakur, "Sign Language Recognition using Microsoft Kinect," in *Proceedings of Sixth International Conference on Contemporary Computing*, 2013, pp. 181–185.

[116] C. Keskin, F. Kirac, Y. E. Kara, and Akarun, “Real time hand pose estimation using depth sensors,” in Proceedings *of the IEEE International Conference on Computer Vision Workshop*, 2011, pp.1228–1234.

[117] S. Lang, M. Block, and R. Rojas, “Sign language recognition in Kinect,” in *Proceedings of11th International Conference on Artificial Intelligence and Soft Computing*, 2012, pp. 394–402.

[118] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, “American sign language recognition with the Kinect,” in *Proceedings of the 13th International Conference on Multi-modal Interfaces*, 2011, pp.279–286.

[119] K.K.Biswas and S.K.Basu, “Gesture recognition using Microsoft Kinect,”in *Proceedingsof 5th International Conference on Automation, Robotics and Applications*, 2011, pp.100– 103.

[120] M. Oszust and M. Wysocki, “Recognition of Signed Expressions Observed by Kinect Sensor,” in *Proceedings of 10th International Conference on Advanced Video and Signal Based Surveillance*, 2013, pp.220–225.

[121] C.Sun, T.Zhang, B.-K.Bao, C.Xu, and T.Mei,“Discriminative Exemplar Coding for Sign Language with Kinect,”*IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1418–1425, 2013.

[122] W.Jiangjin and G.Wen,“Sign Language Recognition Method on ANN/HMM,”*Computer Science and Application*, vol. 9, pp. 1–5, 1999.

[123] W. Jiangjin and G. Wen, “A Hierarchical DGMM Recognize for CSLR,”*Journal of Software*, vol. 11, no. 11, pp. 515–525,2000.

[124] A.Hyvarinen,“A family of fixed-point algorithms for independent component analysis,”in *Proceedings of IEEE International Conference on Acoustics,Speech and Signal Processing*, 1997, pp. 3917–3920.

[125] G. Fang, W. Gao, X. Chen, C. Wang, and J. Ma, “Signer-independent continuous sign language recognition based on SRN/HMM,” in *Gesture and Sign Language in Human-Computer Interaction*, I. Wachsmuth and T. Sowa, Eds. Berlin: Springer-Verlag,2001.

[126] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, “Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2005, pp. 786–791.

[127] X. Chen, “Feasibility of building robust surface electromyography-based hand gesture interfaces,”2009.

[128] V. E. Kosmidou andL.J. Hadjileontiadis, “Sign language recognition using intrinsic mode sample entropy on sEMG and accelerometer data,”*56*, vol. 12, no. 2879–2890,2009.

[129] J. Kim, J. Wagner, M. Rehm, and E. Andre, “Bi-channel sensor fusion for automatic sign language recognition,” in *Proceedings of IEEE International Conference on AutomaticFace Recognition*, 2008, pp.1–6.

[130] D.Ma, X.Chen, Y.Li,J.Cheng, and Y.Ma,“Surface Electromyography and Acceleration Based Sign Language Recognition Using Hidden Conditional Random Fields,” in *Proceedings of 2012 IEEE International Conference on Biomedical Engineering and Sciences*, 2012, pp.535–540.

[131] Y.Li, X.Chen, X.Zhang, K.Wang, and Z.I.Wang,“ASign-Component-Base Framework for Sign Language Recognition Using Accelerometer and sEMG Data,”*IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2695–2704,2012.

[132] C. Vogler and D. Metaxas, “A framework for recognizing the simultaneous aspects of American Sign Language,”*Computer Vision and Image Understanding*, vol. 81, pp. 358–384, 2001.

[133] S. Berman and H. Stern, “Sensors for Gesture Recognition Systems,”*IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, vol. 42, no. 3, pp. 277–290, 2012.

[134] J. Kramer and L. Leifer, “The Talking Glove: An Expressive and Receptive ‘Verbal’ Communication aid for the deaf,deaf-blind and non-vocal,”*SIGCAPH Newsletter*,vol.39, pp. 12–16, 1988.

[135] F. S. Sidney and G. E. Hinton, “Glove Talk- A Neural Network Interface between a Data-Glove and a Speech Synthesizer,”*IEEE Transactions on Neural Networks*,vol.4,no.1,pp. 2–8, 1993.

[136] J.L.Hernandez, N.Kyriakopoulos, and R.Linderman,“A Multi-Class Pattern Recognition of Practical Finger Spelling Translation,” in *Proceedings of IEEE International Conference on Multi-modal Interfaces*, 2002, pp.195–190.

[137] J.L.Hernandez, N.Kyriakopoulos, and R.Linderman,“A New Instrumented Approach for Translating American Sign Language into Sound and Text,” in *Proceedings of 6thInternational Conference on Automatic Face and Gesture Recognition*, 2004, pp.547–552.

[138] M.Borghetti, E.Sardini, and M.Serpelloni,“Sensorized Glove for Measuring ha n d Finger Flexion for Rehabilitation Purposes,”*IEEE Transactions on Instrumentation and Measurements*, vol. 62, no. 12, pp. 3308–3314,2013.

[139] V. Pitsikalis, S. Theodorakis, and P. Maragos, “Data-Driven Sub-Units and Modeling Structure for Continuous Sign Language Recognition with Multiple Cues,” in *Proceedings of the IEEE International Conference Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Languages Technologies*, 2010, pp.196–203.

[140] C. Conly, P. Doliotic, P. Jangyodsuk, R. Alonzo, and V. Athotsos, “Toward a 3d body part detection video dataset and hand tracking benchmark,” in*Proceedings of the 6thInternational Conference on Pervasive Technologies Related to Assistive Environment*, New York, NY, 2013.

[141] P. Jangyodsuk, C. Conly, and V. Athitsos, “Sign Language Recognition Using Dynamic Time Warping and Hand Shape Distance Based on Histogram of Oriented Gradient Features,”*Proceedings of the 7th International Conference on Pervasive TechnologiesRelated to Assistive Environments*, pp. 1–6,2014.

[142] S. Mann, J. Nolan, and B. Wellman, “Sousveillance: Inventing and using wearable computing devices for data collection in surveillance environments,”*Surveillance &Society*, vol. 1, no. 3, pp. 331–335,2002.

[143] S. Mann, “Sousveillance: inverse surveillance in multimedia imaging,” in *Proceedings of the 12th Annual ACM International Conference on Multimedia*, 2004, pp.177–180.

[144] S. Mann, J. Nolan, and B. Wellman, “Blind Navigation with a wearable range camera and vibrotactile helmet,” in *Proceedings of the 19th ACM International Conference on Multimedia*, 2011, pp.1325–1228.

[145] R. C. H. Lo, S. Mann, J. Huang, V. Rampersad, and T. Ai, “High dynamic range video image processing for digital glass,” in *Proceedings of the 20th ACM International Conference on Multimedia*,2012.

[146] R. Battison, “Phological deletion in American sign language,”*Sign Language Studies*, vol. 5, pp. 1–19, 1974.

[147] W.C.Stokoe,“Sign language structure,”*Annual Review of Anthropology*,vol.9,no.1,pp. 365–390, 1980.
[148] V. I. Pavlovic, R. Sharma, and T. S. Huang, “Visual interpretation of hand gestures for human-computer interaction: A Review,”*IEEE Transactions in Pattern Analysis andMachine Intelligence*, vol. 19, pp. 677–695,1997.
[149] A.D.Wilson and A.F.Bobick,“Parametric hidden markov models for gesture recognition,”*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 884– 900, 1999.
[150] G. Ye, J. J. Corso, and G. D. Hager, “Gesture Recognition Using 3D Appearance and Motion Features,” in *Proceedings of Computer Vision and Pattern Recognition Workshop on Real-Time Vision for Human-Computer Interaction*, 2003, pp.160–166.
[151] M. Sigalas, H. Baltzakis, and P. Trahanias, “Gesture recognition based on arm tracking for human-robot interaction,” presented at the IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS), 2010, pp. 5424–5429.
[152] Y. Araga, M. Shirabayashi, K. Kaido, and H. Hikawa, “Real Time Gesture Recognition Using Posture Classifier and Jordan Recurrent Neural Network,” presented at the IEEE World Congress on Computational Intelligence, Brisbane, Australia, 2012.
[153] R. M. Haralick and L.G. Shapiro, “Image Segmentation Techniques,”*Computer Vision,Graphics, and Image Processing*, vol. 29, no. 1, pp. 100–132,1985.
[154] W. K. Pratt,*Digital Image Processing*. Hoboken, NJ: Wiley,2007.
[155] R. C. Gonzales and R. E. Woods,*Digital Image Processing*, 3rd ed. New York, NY: Prentice-Hall,2008.
[156] B. D. Zarit, B. J. Super, and F. K. H. Quck, “Comparison of five color models in skin pixel classification,” in*Proceedings of International Conference on Computer Vision*, 1999, pp. 58–63.
[157] H. Yao and W. Gao, “Face detection and location based on skin chrominance and lip chrominance transformation from color images,”*Pattern Recognition*, vol. 34, pp. 1555–1564, 2001.
[158] M. J. Jones and J. M. Rehg, “Statistical Color Models with Application to Skin Detection,” *International Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.
[159] S.L.Phung, A. Bouzerdoum, and D. Chai, “Skin segmentation using color pixel classification: analysis and comparison,”*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 148–154,2005.
[160] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H.W.Jensen, and M.Gross,“Analysis of Human Faces using a Measurement-Based Skin Reflectance Model,”*ACM Transactions on Graphics*, vol. 25, no. 3, pp. 1013–1024, 2006.
[161] F. H. Xiang and S. A. Suandi, “Fusion of Multi Color Space for Human Skin Region Segmentation,”*International Journal of Information and Electronics Engineering*, vol. 3, no. 2, pp. 172–174, 2013.
[162] X. Zabulis, H. Baltzakis, and A. A. Argyros, “Vision-based Hand Gesture Recognition for Human Computer Interaction,” in *Universal Access Handbook*, Mahwah, NJ: Lawrence Erlbaum, 2009.
[163] D. G. Lowe, “Distinctive image features from scale-invariant key-points,”*International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110,2004.

[164] B. Triggs, “Detecting key-points with stable position, orientation, and scale under illumination changes,” presented at the Eight European Conference on Computer Vision, 2004, pp. 100–113.
[165] K. Mikolajszk, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, and T. Kadir, “A comparison of affine region detectors,”*International Journal of Computer Vision*, vol. 65, no. 1–2, pp. 43–72, 2005.
[166] R. Szeliski,*Computer Vision: Algorithms and Applications*. London: Springer, 2009.
[167] T.Tuteylaars and K.Mikolakczyk,“Local Invariant Feature Detectors:A Survey,” *Foundations and Trends in Computer Graphics and Vision*,vol.3,no.3,pp.177–280,2008.
[168] Y. Mingqiang, K. Kidiyo, and R. Joseph, “A Survey of Shape Feature Extraction Techniques,” in *Pattern Recognition Techniques, Technology and Applications*, P.-Y. Yin, Ed. Vienna, Austria: In Tech Publishers,2008.
[169] M.-K. Hu, “Visual pattern recognition by moment invariants,”*IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.
[170] F.Timm and T.Martinetz,“Statistical Fourier Descriptors for Defect Image Classification,” presented at the International Conference on Pattern Recognition, 4190-4193,2010.
[171] L. F. Costa and R. M. Cesar,*Shape analysis and classification: Theory and practice*. Baco Raton, CA: CRC Press,2001.
[172] I. Kunttu, L. Lepisto, J. Rauhamaa, and A. Visa, “Multi-scale Fourier descriptor for shape-based image retrieval,” presented at the IEEE International Conference on Pattern Recognition, 2004, pp.765–768.
[173] C.Dalitz, C.Brandt, S.Goebbels, and D.Kolanus,“Fourier descriptors for broken shapes,” *EURASIP Journal on Advances in Signal Processing*, vol. 161, pp. 1–11, 2013.
[174] D. Zhang and G. Lu, “A Comparative Study of Fourier Descriptors for Shape Representation and Retrieval,” presented at the The 5th Asian Conference on Computer Vision, 2002, pp. 646–651.
[175] I. Kunttu and L. Lepisto, “Shape-based retrieval of industrial surface defects using angular radius Fourier descriptor,”*IET Image Processing*, vol. 1, no. 2, pp. 231–236,2007.
[176] D. Zhang and G. Lu, “Study and evaluation of different Fourier methods for image retrieval,”*Image and Vision Computing*, vol. 23, pp. 33–49,2005.
[177] M.S.Nixon and A.Aguando, *Feature Extraction and Image Processing.* Philadelphia,PA: Elsevier,2007.
[178] M.S.Hitam, S.AbuBakar, W.N.Jawahir, and W.Yussof,“Content-Based Image Retrieval Using Zernike Moments for Binary and Grayscale Images,” in *Moments and Moments Invariants- Theory and Applications*, G. A. Papakostas, Ed. Greece: Science Gate Publishing,2014.
[179] H. S. Kim and H.-K. Lee, “Invariant Image Watermark Using Zernike Moments,”*IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 766–775, 2003.
[180] R. Bracewell,*The Fourier Transform and its Applications*. New York, NY: McGraw-Hill, 1965.
[181] J. Altmann and H. J. P. Peitbock, “A fast correlation method for scale-and-translation-invariant pattern recognition,”*IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 46–57,1984.
[182] R. Mukundan and K. R. Ramakrishnan, *Moment Functions in Image Analysis*. Singapore: World Scientific,1998.

[183] P.T.Yap, R.Paramesran, and S.H.Ong, "Image analysis by Krawtchouk moments,"*IEEE Transaction on Image Processing*, vol. 12, no. 11, pp. 1367–1377,2003.
[184] C.-H. Teh and R. T. Chin, "On Image Analysis by the Methods of Moments,"*IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 496–512, 1988.
[185] R. Palaniappan, P. Raveendran, and S. Omatsu, "Noise tolerant moments for neutral network classification," in *Proceedings of International Joint Conference on Neural Networks*, 1999, pp.2802–2807.
[186] Z. J. Miao, "Zernike moments-based image shape analysis and its application,"*Pattern Recognition*, vol. 21, no. 2, pp. 169–177,2000.
[187] A.Khotanazad and Y.H.Hong,"Invariant Image Recognition by Zernike Moments,"*IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, 1990.
[188] S. O. Belkasim, M. Shridhar, and M. Ahmadi, "Pattern Recognition with Moment Invariants: A Comparative Study and New Results,"*Pattern Recognition*, vol. 24, no. 12, pp. 1117–1136, 1991.
[189] J. Wood, "Invariant pattern recognition: a review,"*Pattern Recognition*, vol. 29, no. 1, pp. 1–17, 1996.
[190] H. Zhu, M. Liu, H. Ji, and Y. Li, "Combined invariant to blur and rotation using Zernike descriptors,"*Pattern Analysis Applications*, vol. 13, pp. 309–319,2010.
[191] J.Flusser, T.Suk, and B.Zitova, *Moments and Moments Invariants in Pattern Recognition*. Hoboken, NJ: Wiley,2009.
[192] K. M. Hosny, "On the computational aspects of affine moment invariant for gray-scale images,"*Applied Mathematics and Computation*, vol. 195, pp. 762–771,2008.
[193] G.Rigoll, A.Kosmala, and S.Eickeler,"High Performance Real-Time Gesture Recognition Using Hidden Markov Models," in *Proceedings of Gesture Workshop*, 1998, pp.69–80.
[194] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden markov models,"*Image and Vision Computing*, vol. 21, no. 8, pp. 745–758, 2003.
[195] R.Yang and S.Sarkar,"Gesture recognition using hidden markov models from fragmented observations," presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, vol. 1, pp.766–773.
[196] C. Cortes and V. Vapnik,*The Nature of Statistical Learning Theory*. New York, NY: Springer,1995.
[197] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition,"*Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 955–974,1998.
[198] X. Wang and K. K. Paliwal, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition,"*Pattern Recognition*, vol. 36, pp. 2429–2439, 2003.
[199] B. Scholkopf, C. Burges, and V. Vapnik, "Extracting support data for a given task," in *Proceediings, First International Conference on Knowledge Discovery & Data Mining*, Menlo Park, CA: AAAI Press,1995.
[200] K. E. Clarkson and P. J. Moreno, "On the use of support vector machines for phonetic classification," in Proceeding *of International Conference, Acoustics, Speech and Signal Processing*, 1999, pp.585–588.

[201] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multi-class support vector machines,"*IEEE Transactions on Neural Networks*, vol. 13, pp. 415–425,2002.
[202] K. Crammer and Y. Singer, "On the Algorithmic Implementation of Multi-classs Kernel-based Vector Machines,"*Journal of Machine Learning Research*,vol.2,pp.265–293,2001.
[203] T. G. Dietterich and G. Bakiri, "Solving multi-class learning problems via error-correcting output codes,"*Journal of Artificial Intelligence Research*, vol. 2, pp. 263–286,1995.
[204] E.Witten and I.H.Frank, *Data Mining:Practical Machine Learning Tools and Techniques*. San Francisco, CA: Morgan Kaufmann,2005.
[205] E. Kong and T. G. Dietterich, "Error-correcting output coding correct bias and variance," in *Proceedings of the 12th International Conference on Machine Learning*, 1995, pp. 313– 321.
[206] L. I. Kucheva, "Using diversity measures for generatin error-correcting output codes in classifier ensembles,"*Pattern Recognition Letters*, vol. 26, pp. 83–90,2005.
[207] J. Cai and A. Goshtasby, "Detecting human faces in color images,"*Image and Vision Computing*, vol. 18, pp. 63–75,1999.
[208] K.Wong, K.Lam, and W.Siu,"A robust scheme for live detection of human faces in color images," in *Proceedings of Science et Technologie des Systemes pi-Conjugues*, 2003, pp. 103–114.
[209] D. Chai and K. N. Ngan, "Face segmentation using skin color map in videophone applications,"*IEEE Transactions on Circuits and Systems for Video Technology*,vol.9,no. 4, pp. 551–564, 1999.
[210] G. Kukharev and A. Novosielski, "Visitor identification elaborating real time face recognition system," in *Proceedings of the 12th Winter School on Computing Graphics(WSCG)*, 2004, pp.157–164.
[211] N.Otsu,"A threshold selection method from gray-level histograms,"*IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66,1979.
[212] X. Yang, X. Shen, J. Long, and H. Chen, "An Improved median-based Otsu image thresholding Algorithm,"*AASRI Procedia*, vol. 3, pp. 468–473, 2012.
[213] P.-S. Liao, T.-S. Chen, and P.-C. Chung, "A Fast Algorithm for Multilevel Thresholding," *Journal of Information Science and Engineering*, vol. 17, pp. 713–727, 2001.
[214] D.Zhang and G.Lu,"A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval,"*Visual Communication and Image Representation*, vol. 14, no. 1, pp. 41–60, 2003.
[215] F. J. Valverde-Albacete, J. Carrillo-de-Albornoz, and C. Pelaez-Moreno, "A Proposal for New Evaluation Metrics and Result Visualization Technique for Sentiment Analysis Tasks," in *Information Access Evaluation meets Multilinguality, Multimodality and Visualization*, P.Forner, H.Muller, R.Paredes, P.Rosso, and B.Stein, Eds.Berlin: Springer Verlag, 2013.
[216] N. Japkowicz and M. Shah,*Evaluating Learning Algorithms: A Classification Perspective*. Cambridge: Cambridge University Press, 2011.
[217] J. Davis and M. Goadrich, "The Relationship Between Precision-Recall and ROC Curves," in *Proceedings of 23rd International Conference in Machine Learning*, 2006, pp.1–8.
[218] C. Goutte and E. Gaussier, "A Probabilities Interpretation of Precision, Recall and F-score with Implication for Evaluation," in *Proceedings of European Colloquium on IR Research*, Berlin,2005.
[219] M. Sokolova and G. Lapalme, "A systematic analyses of performance measures for classification tasks,"*Information Processing and Management* ,vol.45,pp.427–437,2009.

[220] G.D.Finlayson, M.S.Drew, and C.Lu,“Entropy minimization for shadow removal,”
International Journal of Computer Vision, vol. 85, no. 1, pp. 35–57, 2009.
[221] K. Deb and A. H. Suny, “Shadow Detection and Removal Based on YCbCr Color Space,” *Smart Computing Review*, vol. 4, no. 1, pp. 23–33, 2014.
[222] K. E. E. Rani and G. Jemilda, “Shadow detection in single still image using TAM based multi-step algorithm,”*International Journal of Computer Science and Technology*, vol. 2, no. 4, pp. 496–500, 2011.
[223] D. Miyazaki, Y. Matsushita, and K. Ikeuchi, “Interactive shadow removal from a single image using hierarchical graph cut,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp.223–230.
[224] L. Xu, F. Qi, and R. Jiang, “Shadow removal from a single image,” in *Proceedings of the6th International Conference on Intelligent Systems Design and Applications*, 2006, vol. 2, pp. 1049–1054.
[225] S. A. V. Brandon Garcia, “Real-time American Sign Language Recognition with Convolutional Neural Networks,” Stanford University, Stanford, CA, 2016.
[226] Sahoo, Ashok & Mishra, Gouri & Ravulakollu, Kiran. (2014). Sign language recognition: State of the art. ARPN Journal of Engineering and Applied Sciences. 9. 116-134.
[227] Shraddha Taunk,Dipesh Kumar Sharma & Ram Nivas Giri. (2014). Static Gesture Recognition of Devnagari Sign Language using Feed-Forward Neural Network. International Journal of Advanced Research in Computer Engineering &Technology (IJARCET) Volume 3 Issue 10.
[228] Noda, Kuniaki & Yamaguchi, Y & Nakadai, Kazuhiro & Okuno, Hiroshi & Ogata, Tetsuya. (2014). Lipreading using convolutional neural network. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH. 1149-1153.