LITERATURE SURVEY

1. **TITLE : Indian Sign Language Gesture Recognition using Image Processing and Deep Learning**

   AUTHOR : Neel Kamal Bhagat, Vishnusai Y, Rathna G N

   YEAR : 2019

   The data captured by the Microsoft Kinect RGBD camera. Given that there is no one-to-one mapping between the pixels of the depth and the RGB camera, they used computer vision techniques like 3D construction and affine transformation. Convolutional Neural Networks (CNNs) were utilized for training 36 static gestures relating to Indian Sign Language (ISL) alphabets and numbers. Their model achieved an accuracy of 98.81% on training using 45,000 RGB images and 45,000 depth images. Convolutional LSTMs were used for training 10 ISL dynamic word gestures and an accuracy of 99.08% was obtained by training 1080 videos. The model also showed competitive adaptability to American Sign Language (ASL) gestures when the ISL model's weights were transferred learned to ASL and it resulted in giving 97.71% accuracy. Effective real-time background subtraction was done using depth perception techniques. Computer vision techniques were used to achieve one-to-one mapping between the depth and the RGB pixels. The depth and segmented static model achieves an accuracy of 98.81 % and the dynamic model achieves 99.08% on the training set. Further research can be focused on the real-time prediction of more words related to ISL and also on sentence formation.

2. **TITLE : Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues**

   AUTHOR : Muhammad Al-Qurishi , Thariq Khalid, And Riad Souissi

   YEAR : 2021

   This paper researcher conducted a comprehensive review of automated sign language recognition based on machine/deep learning methods and techniques. Their study also indicates that input modalities bear great significance in this field; it appears that recognition based on a combination of data sources, including vision-based and sensor-based channels, is superior to a unimodal analysis. In addition, recent advances have allowed researchers to move from simple recognition of sign language characters and words towards the capacity to translate continuous sign language communication with minimal delay. This paper proposes, the convolutional neural network (CNN) model was one of the first to gain major attention. In addition to CNN's, other architectures such as hidden Markov models (HMMs) and recurrent neural networks (RNNs) are frequently applied. The support vector machine (SVM) model is frequently used for this purpose as well while random forest (RF) and K-nearest neighbor (k-NN) are sometimes chosen for the classification task. Some hybrid models are emerging that combine the best characteristics of several types of neural networks, and solutions of this type may represent the most logical path forward with respect to advanced SLR applications. It is reasonable to expect breakthroughs in this field in the future, and many of the research studies may include key elements that eventually become a part of the final solution to automated sign language recognition. Even at this stage, many SLR tools can be practically used to some extent and can

provide immediate relief to disabled people as well as point to future directions of research.

3. **TITLE : Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language.**

   AUTHOR : Farman Shah, Muhammad Saqlain Shah, Waseem Akram, Awais Manzoor, Rasha Orba Mahmoud, And Diaa Salama Abdelminaam

   YEAR : 2021

   All over the world, deaf people use sign language as the only reliable source of communication with each other as well as with normal people. These communicating signs are made up of the shape of the hand and movement. In Pakistan, deaf people use the Pakistan sign language (PSL) as a means of communication with people. In scientific literature, many studies have been done on PSL recognition and classification. Most of these works focused on colored-based hands while some others are sensors and Kinect-based approaches. These techniques are costly and also avoid user-friendliness. In this paper, a technique is proposed for the recognition of thirty-six static alphabets of PSL using bare hands. The dataset is obtained from the sign language videos. At a later step, four vision-based features are extracted i.e. local binary patterns, a histogram of oriented gradients, an edge-oriented histogram, and speeded-up robust features. The extracted features are individually classified using Multiple kernel learning (MKL) in a support vector machine (SVM). They employed a one-to-all approach for the implementation of basic binary SVM into the multi-class SVM. A voting scheme is adopted for the final recognition of PSL. The performance of the proposed technique is measured in terms of accuracy, precision, recall, and F-score. The simulation results are promising as

compared with existing approaches. They suggest the use of other features e.g. maximally stable extremal regions (MSER) features instead of SURF. Also, there is a lot of room to further improve the dataset. The segmentation phase of our methodology follows some constraints which are also needed to be user-friendlier.

4. **TITLE : RNN-Transducer-based Chinese Sign Language Recognition**
   AUTHOR : Liqing Gao, Haibo Li, Zhijian Liu, Zekang Liu, Liang Wan, Wei Feng
   YEAR : 2020

   Sign Language Recognition (SLR) targets interpreting sign language video into natural language, which largely facilitates mutual communication between the deaf and the general public. SLR is usually formulated as a sequence alignment problem, wherein connectionist temporal classification (CTC) plays an important role in building effective alignment between video sequence and sentence-level labels. However, CTC-based SLR methods tend to fail if the output label sequence is longer than the input video sequence. Besides, they ignore the interdependencies between output predictions. This paper addresses these two issues and proposes a new RNN-Transducer-based SLR framework, i.e., visual hierarchy to lexical sequence alignment network (H2SNet). In the framework, they design a visual hierarchy transcription network to capture the spatial appearance and temporal motion cues of sign video on multiple levels. Meanwhile, they utilize a lexical prediction network to extract effective contextual information from output predictions. RNN-Transducer is applied to learn the mapping between sequential video features and sentence-level labels. Extensive experiments validate the effectiveness and superiority of our approach over state-of-the-art methods. They

recommended it would be interesting to study sign language translation constrained by the dialogue context, to facilitate instant communication between the deaf using sign language to a person using speech.

5. **TITLE : Dataset of Pakistan Sign Language and Automatic Recognition of Hand Configuration of Urdu Alphabet through Machine Learning**

   AUTHOR : Ali Imrana, Abdul Razzaq, Irfan Ahmad Baig , Aamir Hussaina, Sharaiz Shahida, Tausif-ur Rehmana

   YEAR : 2021

   Social correspondence is one of the most significant columns that the public is dependent on. Notably, language is the best way to communicate and associate with one another both verbally and nonverbally. There is a persistent communication gap among deaf and non-deaf communities because non-deaf people have less understanding of sign languages. Every region/country has its sign language. In Pakistan, the sign language of Urdu is a visual gesture language that is being used for communication among deaf people. However, the dataset of Pakistan Sign Language (PSL) is not available publicly. The dataset of PSL has been generated by acquiring images of different hand configurations through a webcam. In this work, 40 images of each hand configuration with multiple orientations have been captured. In addition, they developed, an interactive android mobile application based on machine learning that minimized the communication barrier between the deaf and non-deaf communities by using the PSL

dataset. The android application recognizes the Urdu alphabet from input hand configuration. This paper proposed a dataset of sign language and android-based communication systems which is a hand configuration acknowledgment framework for Pakistan Sign Language (PSL). The dataset of the PSL is available publicly on the cloud and could be used by anyone and the framework interprets the communication through sign-to-text translation by using a mobile camera. The system accomplished 80–90% accuracy on different tests in various light conditions. They tested their model on different hand configurations.

6. **TITLE : Helping Hearing-Impaired in Emergency Situations: A Deep Learning-Based Approach**

   AUTHOR : Qazi Mohammad Areeb, Maryam , Mohammad Nadeem, Roobaea Alroobaea, and Faisal Anwer

   YEAR : 2021

   Hearing-impaired people use sign language to express their thoughts and emotions and reinforce information delivered in daily conversations. Though they make up a significant percentage of any population, the majority of people can’t interact with them due to limited or no knowledge of sign languages. Sign language recognition aims to detect the significant motions of the human body, especially the hands, analyze them and understand them. Such systems may become life-saving when hearing-challenged people are in desperate situations like heart attacks, accidents, etc. Deep learning-based hand gesture recognition models are developed to accurately predict the emergency signs of Indian Sign Language (ISL). The dataset used contains the videos for eight different emergencies. Several frames were extracted from the videos and are fed to three different models. Two models are designed for classification, while one is an object

detection model, applied after annotating the frames. The first model consists of a three-dimensional convolutional neural network (3D CNN), while the second comprises a pre-trained VGG-16 and a recurrent neural network with a long short-term memory (RNN-LSTM) scheme. The last model is based on YOLO (You Only Look Once) v5, an advanced object detection algorithm. The prediction accuracies of the classification models were 82% and 98%, respectively. The YOLO-based model outperformed the rest and achieved an impressive mean average precision of 99.6%.

7. **TITLE : A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition**

   AUTHOR : Nikolas Adaloglou, Theocharis Chatzis, Ilias Papastratis, Andreas Stergioulas, Georgios Th. Papadopoulos, Vassia Zacharopoulou, George J. Xydopoulos, Klimnis Atzakas, Dimitris Papazachariou, and Petros Daras

   YEAR : 2021

   In this paper, a comparative experimental assessment of computer vision-based methods for sign language recognition was conducted. By implementing the most recent deep neural network methods in this field, a thorough evaluation of multiple publicly available datasets is performed. Their present study provides insights into sign language recognition, focusing on mapping non-segmented video streams to glosses. For their work, two new sequence training criteria, known from the fields of speech and scene text recognition, are introduced. Furthermore, a plethora of pretraining schemes are thoroughly discussed. Finally, a new RGBD dataset for the Greek sign language was created. To the best of their knowledge, this is the first sign language dataset where three annotation

levels are provided (individual gloss, sentence, and spoken language) for the same set of video captures. In their future work, efficient ways for integrating depth information that would guide the feature extraction training phase can be devised. Moreover, another promising direction is to investigate the incorporation of more sequence learning modules, like attention-based approaches, to adequately model inter-gloss dependencies. Future SLR architectures may be enhanced by fusing highly semantic representations that correspond to the manual and non-manual features of SL, similar to humans. Finally, it would be of great importance for Deaf-non Deaf communication to bridge the gap between SLR and SL translation. Advancements in this domain will drive research into SL translation as well as SL-to-SL translation, which has not yet been thoroughly studied.

8. **TITLE : American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion**

   AUTHOR : Wenjin Taoa, Ming C. Leua, Zhaozheng Yinb

   YEAR : 2021

   American Sign Language (ASL) alphabet recognition by computer vision is a challenging task due to the complexity of ASL signs, high interclass similarities, large intraclass variations, and constant occlusions. This paper describes a method for ASL alphabet recognition using Convolutional Neural Networks (CNN) with multiview augmentation and inference fusion, from depth images captured by Microsoft Kinect. Their approach augments the original data by generating more perspective views, which makes the training more effective and reduces the potential overfitting. During the inference step, their approach comprehends information from

multiple views for the final prediction to address the confusing cases caused by orientational variations and partial occlusions. On two public benchmark datasets, our method outperforms the state-of-the-art. The confusion matrix and the most confusing pairs of the leave-one-out evaluation on the ASL benchmark dataset were tested on the 5th subject. evaluated on two public datasets, the ASL benchmark dataset, and the NTUdigit dataset. The experimental results have demonstrated that our method makes significant improvement compared to the previous work, achieving recognition accuracies of 100% and 93% in the half-half and the leave-one-out experiments, respectively, on the ASL benchmark dataset, and achieving recognition accuracies of 100% for both the half-half and the leave-one-out experiments on the NTU digit dataset

9. **TITLE : Image Processing based on Deep Neural Networks for Detecting Quality Problems in Paper Bag Production**
   AUTHOR : Anna Syberfeldt and Fredric Vuolutera
   YEAR : 2021
   The use of deep neural networks to perform automatic quality inspections based on image processing to eliminate the current manual inspection. Manufacturers must identify quality issues in production and prevent defective products from being delivered to customers. They investigate deep neural network was implemented in a real-world industrial case study, and its ability to detect quality problems was evaluated through the use of deep neural networks to perform automatic quality inspections based on image processing to eliminate the current manual inspection and analysis.

Their results show that the network has an accuracy of 94.5%, which is considered good in comparison to the 70–80% accuracy of a deep neural network implemented in a real-world industrial case study, and its ability to detect quality problems was evaluated, and trained human inspector to analyzed. Future work could focus on improving the solution so that it can assess not only the geometry of bags but also faults in print, coloring, and other purely aesthetic defects that are important to customers even though such flaws do not affect the function of the bag. The personnel at the company did indicate that issues with print and color are somewhat predictable, often happening when refilling printing materials or switching between product variants. Although it might not be critical to automatically detect aesthetic defects, doing so would reduce the burden on the operators in the line. It is thus worth investigating adding this capability as an extra feature of the network

10. **TITLE : Hybrid deep learning for detecting lung diseases from an X-ray image**

    AUTHOR : Subrato Bharati , Prajoy Podder, M. Rubaiyat Hossain Mondal

    YEAR : 2020

    Different forms of existing deep learning techniques including convolutional neural network (CNN), vanilla neural network, visual geometry group-based neural network (VGG), and capsule network are applied for lung disease prediction. The basic CNN has poor performance for rotated, tilted, or other abnormal image orientations. Therefore, we propose a new hybrid deep learning framework by combining VGG, data augmentation, and spatial transformer network (STN) with CNN. This new

hybrid method is termed here as VGG Data STN with CNN (VDSNet). The new model is applied to the NIH chest X-ray image dataset collected from the Kaggle repository. For the case of the full dataset, VDSNet exhibits a validation accuracy of 73%, while vanilla gray, vanilla RGB, hybrid CNN and VGG, and modified capsule network have accuracy values of 67.8%, 69%, 69.5%, and 63.8%, respectively. They recommended for future work, would apply modified VGG or other new transfer learning algorithms to the sample and full datasets and then make a hybrid algorithm with the fusion of GoogLeNet, AlexNet, and ResNet-152 architecture and also prepare a dataset by combining two or more chest X-ray datasets and then apply hybrid algorithms on the combined dataset for detecting various lung diseases. Future research scopes will also include the implementation of image data augmentation techniques such as color space augmentations, kernel filters, feature space augmentation, etc., to increase the accuracy of the automated chest X-ray diagnosis system. In the future, the proposed new VDSNet method can be applied to X-ray images of suspected COVID-19 patients to predict whether those patients have COVID-19-related pneumonia, or not.