Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4822 - 4832
Received 05 March 2021; Accepted 01 April 2021.

A Comprehensive Study of Real-Time Hand GestureDetectionand Recognition

Pratyay Mukherjee
ComputerScienceand Engineering Department
SRMInstituteofScienceandTechnologyKattankulathur,TamilNadu,Indiapratyaymukherjee9@gmail.com

Nimisha Khaitan
ComputerScienceandEngineering Department
SRM Institute of Science and TechnologyKattankulathur,TamilNadu,Indiakhaitannimisha@gmail.com

T. Balachander
Computer Science and Engineering Department
SRM Institute of Science and Technology
Kattankulathur, Tamil Nadu, India
balachat2@srmist.edu.in

ABSTRACT

Hand gesture is an important form of communication used to convey instructions, give commands, and is an importantmeans for mute people to express themselves. The vast majority of the world population do not know about sign language thatthemutepeopleusetocommunicatewitheachotherandhencefinditdifficulttocomprehendwhatthedisabledpersonistryingto convey. With the growth in the field of human-computer interaction, a significant amount of research work has been devotedto bridging this communication gap.Breakthroughs in the field of computer vision, deep learning, Internet-of-Things, andcomputational power or capability has motivated a significant amount of work towards real-time hand gesture detection andrecognition. The primary reason for the popularity of this field of research is the challenge of generalization of the system tomakeitaccessibletoallthepeoplewhoneeditandtheproblems associatedwithhanddetection.Inthispaper,weexploresomenovel approaches used to solve this problem and compare the results of these systems to determine their compatibility underdifferentscenariosandtheirimpactonsociety.

Keywords – hearing impairment; speech impairment; assistive devices; hand gesture detection; hand gesture recognition;convolutionalneuralnetworks;imageprocessing;deeplearning

I. INTRODUCTION

Withtheadvancementoftechnologyandthegrowthofresearchinthefieldofhuman-computerinteraction,ithasbecomethe responsibility of computer scientists to find ways to improve the lifestyle of the section of the population that is disabled. Acommunication gap existsbetweenthedisabledpeopleandthenormal(abled)peoplethatneed tobebridgedandcomputerscanplayamajorroleinthiscasetocreateasocio-economicimpactontheworld.Aglobalpictureoftheconditionofdisabledpeoplecan be found in the World Health Organization’s World Report on Disability [1] where the importance of assistive devices andaccess to technology has been given primary emphasis along with legal policies to protect their rights and opportunities. Anestimated 1 billion people in the world which accounts for 15% of the world population suffer from some kind of disability. Toimprove the accessibility and quality of the assistive devices standardization of technology is required as well as educationcampaignstopromotethesestandardsandfamiliarizethepeoplewiththeassistivedevices[2].Theeffect oftheseeffortson theimprovementinthelifestyleofthedisabledpopulation issignificantandcanleadtothegrowth ofsocietyinthedevelopinganddevelopednations,

As stated in [3], *speech disability* is the inability of a person to speak properly. This includes speech disability rangingfrom difficulty in uttering words to complete muteness. As stated in [3], *hearing disability* is the inability of a person to hearproperly and the degree of disability is estimated based on the better of the 2 ears. Since the ability to speak mostly comes fromtheabilitytohear,childrenwhohavehearingimpairmentfrombirthusuallyhavespeechimpairmentaswell. Severalreportsandcasestudiessuggestthatthesepeoplearedeniedequalopportunitiesandhumanrightsandsubjectedtosocialinjustice.Thefactsheetonpersonswithdisabilit

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4822 - 4832
Received 05 March 2021; Accepted 01 April 2021.

iescompiledandpublishedbyUN [4]givesstatisticsthatsupportthisstatement.

1. "Disability rates are significantly higher among groups with lowereducational attainment in the countries of theOrganisationforEconomicCo-operationandDevelopment(OECD),saystheOECDSecretariat.Onaverage,19percentofless-educatedpeoplehavedisabilities,comparedto11percentamongthebettereducated."
2. "The World Bank estimates that 20 percent of the world's poorest people have some kind of disability, and tend to beregardedintheircommunitiesasthemostdisadvantaged."
3. "Ninetypercentofchildren withdisabilitiesin developingcountriesdonotattend school,saysUNESCO."
4. "IntheOECDcountries,studentswithdisabilitiesinhighereducationremainunder-represented,althoughtheirnumbersareontheincrease,saystheOECD."
5. "Anestimated386million oftheworld'sworking-agepeoplehavesomekindofdisability,saystheInternationalLabourOrganization (ILO). Unemployment among persons with disabilities is as high as 80 percent in some countries. Oftenemployers assumethatpersonswithdisabilitiesareunabletowork."
6. "Research indicates that violence against children with disabilities occurs at annual rates at least 1.7 times greater thanfortheirpeerswithoutdisabilities."

Sign language [10] makes use of hand gestures to enable a mute person to communicate with normal people and otherpeople with hearing and/or speech disabilities. However, this requires that the other person knows the sign language. Hence, itis important to come up with efficient, convenient, and cost-effective solutions to bridge this communication gap. In this paper,wediscusssomeofthesesolutionsandcorrespondingexamplesin detail.Thesolutionshavebeenclassifiedinto2broadgroupsbasedonhowtheydetectthehandgestureasfollows.

1. Hardware-basedapproaches:Thesemethods employsensors andphysicaltrackers tomaphandgestures.
2. Software-based approaches: These methods employ computer vision, image processing techniques, machine learningalgorithmstomapandclassifyhandgestures.

Thevariousmethodsarecomparedandanalyzedtodeterminewhichoftheseperformsbestundercertaincircumstances.

II. LITERATURESURVEY

The solutions to hand gesture detection and recognition have been classified based on the methods or tools they use inthe detection phase of the process. By detection of hand gestures, it means to find out the existence of hand in the perceivedenvironment from the given input of image or video frames. By recognition of hand gestures, it means the classification of thehanddetectedintheinputintooneofthepre-definedclasseswhichinthecaseofsignlanguagearethecharactersofthelanguage[5].

A. Hardware-based Approaches

These methods of hand gesture detection and recognition were more common before the advent of powerful computersthatledtothemassivedevelopmentofdeeplearningandcomputervision.Hardware-basedapproachesinvolvetheuseofsensors,radars, or tools that have an ensemble of these to perceive the gesture of the hand. The hand detection phase is implicit in thesemethods asthehardwarecomponentispresentonthehandsofthe userwhoistryingtocommunicatewithsomeoneelsei.e. thehand is already detected and there is no ambiguity as to what the output of the sensors are trying to describe. The recognitionpart is carried out by some machine-learning, statistical or rule-based method that takes the description of gesture as input fromthesensorsareprocessesit.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4822 - 4832
Received 05 March 2021; Accepted 01 April 2021.

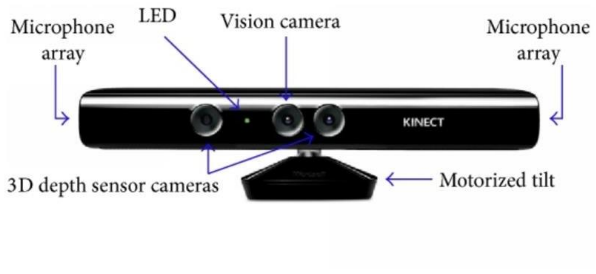

Figure 1. A Microsoft Kinect camera with 3D depthsensorand vision cameras.(Fig.1.[6])

Figure 2. Features extracted(left) from gesture described byparticipant(right) built from Kinect 3D depth image (Fig. 6.[7])

Oneofthemost commonlyusedsensorsisMicrosoft Kinect[6]thatgivesa3Dviewoftheactionsperformedbytheuserin focus. The depth-view gives a better understanding of the environment that has encouraged researchers to extend its utility tomake efficient human-computer interaction systems. A good use of the Kinect sensor for hand gesture recognition can be foundin [7]. It gives a generalized approach that is mostly invariant to skin color and lighting conditions as it is based on the depth-view of the hand gesture. Kinect sensors are used to get an RGB image of 1920x1080 and depth image of 512x424. The depthimageisobtainedaftercalculatingthespecklesspreadrandomly.Thedepthdistanceisspatiallystratified.To obtaintheROIofthe hand gesture, there are two steps. First is to discover the location of the gesture and detect whether it is left or right hand.Secondstepistodetecttheskincolorlayerbylayerintheframes. Thepossibilityofnoisehastoberuledoutbeforeproceedingfurther. The binary image of the gesture is extracted and then from ROI the whole hand is carved out. The fingertip detectionmodule was used to acquire data to build a classifier model. The position of gestures was at random distance from the sensoracquiredusing5differentparticipants.Theresultswereestablishedandtheaccuracyofmorethan95%wasa chievedforvariousdistances for the gestures. The methodology can be extended easily to train the classifier to detect and classify sign languages.There are systems that have been built using Kinect that are dedicated to sign language hand gesture detection. CopyCat[8] isaninteractive-educationalgamedevelopedfordeafchildrentodeveloptheirlanguageandmemoryskills.Intheprototypephaseit used colored gloves fitted with accelerometers that was tracked with a single Firewire camera to detect and recognize handgesturesandtracktheheadpositiontoassisthimorhertoplaythegame. AnotherprototypemodelbasedonthiswasmadeusingKinect [9] with the hope of making it more convenient, robust and easy-to-deploy. The Kinect depth image was used to extract20featuresfrombothstanding and sitting posturesof theparticipants (adultsand children).TheAmericanSignLanguage[10]

dataset was used to train and recognize the characters being shown by the candidate. Although results were good, they were notas goodastheprevious approachto CopyCat[8]inrecognizingbothwordsandsentences withgoodaccuracy.

A flex sensor [11] finds its application in various digital goniometric systems and naturally finds in place in a number ofhuman-computer interaction systems. It measures the amount of deflection from the bend in the sensor and it is generally usedin rehabilitation centers for goniometric applications. A goniometer glove is a device that uses flex sensors to measure the bendandhence, theposturemadebythe

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4822 - 4832
Received 05 March 2021; Accepted 01 April 2021.

Figure 3. A goniometric glove with sensor array containing flex sensors to givethemeasureofbendsin fingerand wrist. (Fig.4. [11])

hand. A communication application based on Universal Sign Language is proposed in [12] that detects and recognizes handgestureandconvertsitintospeechandtext.ThedevicemakesuseofArduinoUnoboard,flexsensorsandanAndroidapplicationthatexecutestheprogramfordetectionandrecognition(ArtificialNeuralNetworkClassifier[13])ofhandgesturefromtheinputof the sensors. The 4 flex sensors used in this system send the input of deflection based on the bend in the resistor that changesits resistance into the neural network with 4 inputs. The neural network model is embedded into the Arduino Uno board thatsends the recognized gesture to the Android application. This application then converts it into speech and text and provides theinterfacetotheusers.Thetrainedneuralnetperformsphenomenalonunseendatagivingameansquarederrorofjust1.0015e-09. However, the system is not particularly convenient due to the dependence on too much hardware making it inconvenient,non-portableandcommerciallyinviable.

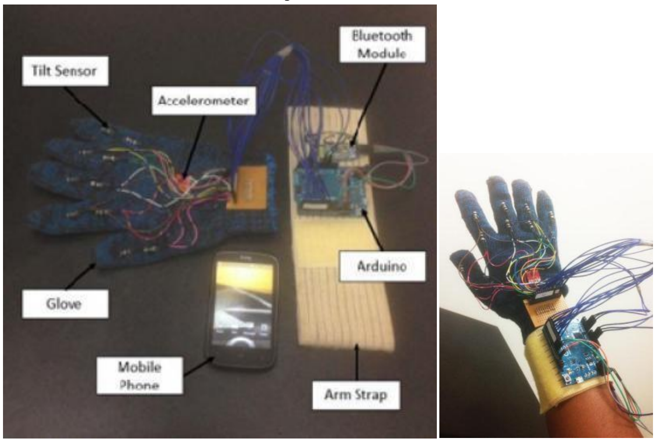

Figure 4. Environmental setup of hand gesture detection and recognition usingsensorglovescontaining tiltsensorsand accelerometer.(Fig.1. [14])

Glove-based approach used to popular before vision-based approach and micro-controllers that could perform the sametask more efficiently became more popular and economic. Gloves are fitted with electronic sensors to measure the tilt, stretchand bend of the hand that gives an accurate estimate of the posture of the hand. In [14] an efficient glove-based approach forMalaysian Sign Language (MSL) hand gesture detection has been proposed. The system makes use of the following hardwaredevices attachedonthegloveonhand.

1. Tiltsensor:Flexsensorsattachedtoeachofthefingers ofthegloves todetecttheflexingofthefingers ofthehand.
2. Accelerometer: To detect the movement of the

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4822 - 4832
Received 05 March 2021; Accepted 01 April 2021.

hand and wrist i.e. roll, pitch and yaw motions. The accelerometertogetherwiththetiltsensorareusedfordetectionofhand.

3. Arduino board: A rule-based classification algorithm is embedded on the board that is attached to a strap on the hand.This has memory of the orientation of the hand and the associated character of MSL. The classification algorithm isrule-basedandachievesanaccuracyrangingfrom78.33%to95%ontesting.

The result from the Arduino is given as input to an Android mobile application that displays the characters conveyed by handgestures.

Latern [15] addresses the problems of real-time hand gesture detection and recognition system such as resilience tobackgroundnoise,generalizationtodifferentpeopleandminimization oflagbetweeninputtothesystemandthecorrespondingoutput.ItmakesuseofFMCW(Frequency-ModulatedContinuousWave)Radarsensortodetectmovementofhandandfeedinthe movement of the hand to the data acquisition module. The sensor is not affected by the lighting, visual background noise oratmospheric pressure. The data acquisition module converts this raw data to a time-series [16] 3D spectrogram dataset that isclassified into simple pre-defined classes of gesture such as pushing, pulling, sliding, etc. The spectrogram shows uniquetrajectoriesofthehandcorrespondingtoeachclassofgestureswhicharefedintoa3DRecurrentConvolutionalNeuralNetwork(commonly used fortime-series data analysis,classificationand regression) that is trained withConnectionist TemporalClassification(CTC)[17]algorithmtomakethemodelcapableofmakingpredictionsofthehandgestureastheinputisstreamedin. Latern with its fast, robust and accurate classification (96% accuracy) has an edge over other experimented conventionaltime-series classification algorithms used in hand gesture recognition such as Hidden Markov Models (72% accuracy) and 2D-CNN(90%).

TABLE 1. Comparative study of experimental architectures carriedoutin Laternwithradardatai.e.spectrogram. (Table1. [15])

	2D-CNN	3D-CNN
NoLSTM	78%	81%
LSTM	86%	92%
LSTM-CTC	91%	96%

B. Software-basedApproaches

Inthe recentyears,withthedevelopmentofefficientdeeplearningandcomputervisionalgorithmsandmorecomputational powers software-based approaches have become more popular than the hardware-based approaches. Thesemethods are mostly vision-based and hence are convenient, cost-effective and accessible. However, these set of solutions thatare vision-based and only depend on a video-stream input comes with its own set of challenges. Hand detection becomes asignificant part of the overall solution. The detection system needs to be robust, noise resilient and generalized. Preprocessingalso becomes a concern as it effects the performance of the recognition stage based on the presence of multiple noise objects inthe foreground and background, multiple desired objects (multiple hands in the same image or video frame), image quality,precision of transformation functions on the image, etc. The recognition phase usually employs classification models andalgorithms rangingfromsimple rule-basedorstatistical(probability-based)algorithms tocomplexdeeplearningmodelsinvolvingartificial, convolutional,recurrent neuralnetworksandtheirensembleorvariations.

Some extremely simple approaches based on empirical observation and threshold methods have been employed over theyears such as one proposed in

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4822 - 4832
Received 05 March 2021; Accepted 01 April 2021.

[32]. This method reveals the main challenges of hand gesture detection and recognition theprimary ones being generalization i.e. invariance to lighting condition, skin tone, gender, age, etc. A threshold is applied on thedistance between the features obtained from the hand region on the video frame and the ones in the database. The class with theminimumdistanceormaximumsimilarityis chosen.Thesimplemodelhastroubledistinguishingsimilargestures.

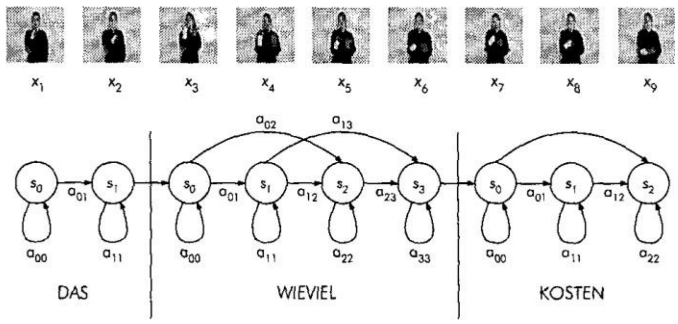

Figure 5. The HMM generating text (in German below) from input video sequence (above). A Bisak-HMMperwordinthesentence.(Fig. 1.[27])

Hidden Markov Models (HMM) [26] are probabilistic (statistical) models used for linear sequence 'labeling' problems. It contains hidden states and the transition from one state to the other is based on the state's transition probability distributionthatemitsaresiduebasedonthestate'semissionprobabilitydistribution.Thus,thesemodelscanbeuse dtogenerateasequencebased on another underlying sequence. An efficient use of HMM in real-time German Sign Language detection and recognitionis proposed in [27]. Although a glove is used in this method, it is solely used for data input of the hand i.e. hand detection butthefunctioningoftheapplicationissolelydependentonthetrainingandusageoftheHMM.That'swhythismeth odisincludedin this section. The method uses a video camera to capture the image of the hand gesture. The hand is detected with the help ofacolorfulglovewornbytheuserandfeaturescharacterizingthepositionofthehandrelativetootherhandandthea xisofthe

bodyareextracted.AnHMMisbuiltforeachsymbolinGSLandthetransitionandemissionprobabilitydistributioni sestimatedwith the features extracted from the segmented image of the hand. A language model is implemented to enhance the real-timeperformance of the system that gives a-priori knowledge of sequence of words. The HMM model gives a good recognitionaccuracy of 91.8% on 97 signs even without invoking a bigram language model. With the introduction of the bigram model theaccuracyis increasedto93.2%.

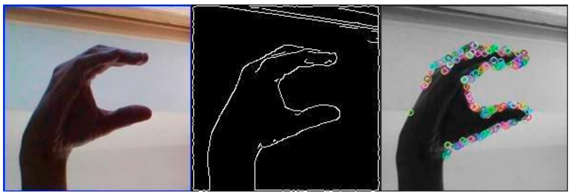

Figure 6. From left to right – original image, segmented edge image using Canny edge detection, ORBfeatures.(Fig.3,4,5 [18])

Apureimageprocessing-basedapproachforASLhandgesturedetectionandrecognitionisproposedin[18].Thismethodhas 4 phases which are segmentation, feature extraction, generation of histogram and visualisation. Before the algorithm flowentersthese4stagesitisnecessarytoperformpreprocessingontheimagestoremovenoiseandenhancethequal ityoftheimage.Thisbecomesimperativemainlyifthesystemisimplementedforcommercialpurposesasthelaterp haseshavebeentrainedandtestedonhigh-

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4822 - 4832
Received 05 March 2021; Accepted 01 April 2021.

resolutionimagesofASLhandgestures.The4phasesareas follows.

1. Segmentation-Beforethealgorithm flowentersthelaterstages itisnecessarytoperformpreprocessingontheimagesto remove noise and enhance the quality of the image. This becomes imperative mainly if the system is implementedfor commercial purposes as the later phases have been trained and tested on high-resolution images of ASL handgestures.Cannyedgedetection[19]isusedtosegmentouttheedgesofthehandandremovebackgroundnoise.
2. Featureextraction–ORB[20]isusedforthispurposethatdetectsdistinctpatcheson theimageandcharacterizeseachimagewithavectorof32elements.
3. Histogram of visual vocabulary – The key descriptors for each class of image is identified from the feature extractionphase. K-means clustering [21] is used for the purpose of finding the clusters of key features. 150 key features areidentified.
4. Recognition – In this phase the feature vector describing the hand gesture is classified into one of the pre-definedcharacters of ASL. The system has been tested with different classifier models and Multi-Layer Perceptron (MLP) hasbeenfoundtooutperformtheothers withanaccuracyof96.96%.

TABLE 2. Accuracy comparison of experimented feature extraction techniques(top tobottom)andclassifiermodels(lefttoright)in [18].

Feature Extraction Technique	SVMAccuracy	MLPAccuracy
ORB	**85.25**	**96.96**
HOG	87.4	79.91
LBP	34.8	61.23
PCA	**92.87**	**98.31**

With the use of ORB almost all classifier models perform well but Support Vector Machines (SVM) and Multi-LayerPerceptron (MLP) stand out from the rest.Although PCA(Principal Component Analysis) gives better performance with most of the classification models, ORB extracts features that are concise and informative enough to be used to train models fast with almost equal accuracy. The method has been tested using static images only and real-time applications may find lags due to expensive computation techniques in phases 1 to 3. Another approach that performs verywell on static images is [22] but it also has the same drawback in terms of real-time applications. The Variance in the datasetwas introduced to add noise, so that model can classify better on test set and give the model resilience to noise. Irrespective oflightingconditionandbackground,thefeaturesareextractedafterconvertingimageintograyscale.Thewristislocatedandthenusing the width of arm the whole hand is outlined. The image is converted into a binary image where there are black and whitepixels.Thewhitepixelscontainthewholehandgesture.Thedatasetissplitintotestandtraindatainaratioof17:3.Thedatasetis fed to SVM and parameters are optimised. K-nearest neighbors helped in finding optimal dimensions for classification. 10-fold crossvalidationhelped todeterminethemarginalandgaussianparameters.Themethod gavemorethan 98%accuracy.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4822 - 4832
Received 05 March 2021; Accepted 01 April 2021.

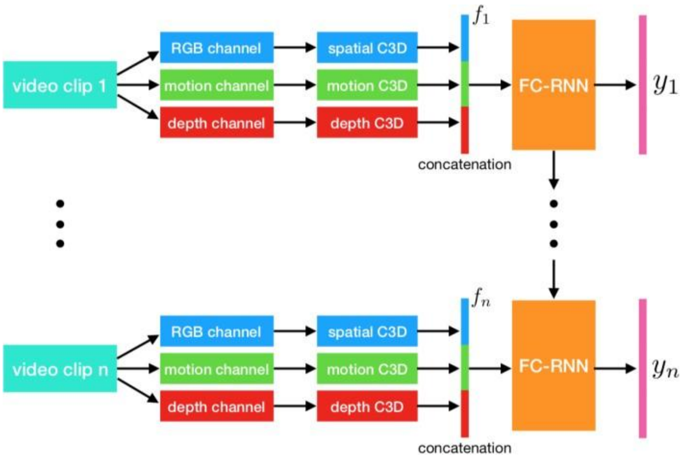

Figure 7. The proposed 3DCNN architecture using 3 channels RGB, Motion and Depth and fully-connected RNNstorecognizeseriesofgestures(yi)inreal-time. (Fig. 2.[23])

Convolutional Neural Networks [13] are one of the most popular variants of Artificial Neural Networks that find itsapplications in the field of image data analysis, natural language processing, computer vision, etc. With the advent of morecomputational power CNNs have become fast, efficient and cost-effective. Sign language detection and recognition systemsbasedonCNNsmayormayhaveseparatedetectionandrecognitionphases.ThisisbecauseCNNsarecapableofidentifyingthefeatures from the input image themselves and classify the input accordingly. Naturally, most state-of-art hand gesture detectionand recognition systems use CNNs for their functioning. [23] makes use of 3D recurrent convolutional neural networks fordetection and recognition of American Sign Language. A new ASL dataset including multiple modalities was created. Thisdataset consists of multiple channels. The proposed model is multi-channel end to end network structure which is used to detectand identify American Sign Language using continuous videos. The model is trained to recognize words and sentences apartfrom ASL characters. The 3DRCNN model is integrated with C3D to acquire sequential and temporal information of clips. TheC3D cuts the video clips into clips of fixed length and then proceeds for finetuning and classification. The features are fed asinputtoRNNmodel forfinalrepresentation.Thefeaturesextractedfrommotion,depthC3DandRGBnetworksformthefeaturerepresentation. The dataset is acquired using Kinect 2.0 sensor. For RGB channel, the accuracy of 61.8% was achieved. Formotionanddepthchannels,usedforsentenceformationinASLitwasconsiderablylessandwasapproximately50%.Apartfrommodels that are heavily dependent on the CNN to detect the features, there are proposed systems as in [24] that make the CNNarchitecture simple by feeding in a pre-processed image of hand thereby eliminating the need for the CNN to detect the hand.This method performs hand region segmentation using mask images from the open access dataset [25] that are convolved overthe input image to generate. This generates a region consisting of the hand gesture. The fingertip of the hand is detected usingconnected-component analysis. It is fed as input to the CNN model. The CNN model is trained on the dataset containing theorientationofthefingertipsandthecorrespondinglabels.Theaccuracyscoreof96.2%isequivalenttothe3DRCNNmodelusedpreviously. However,inreal-timethesimplemodelcangiveanoverheadforthepreprocessingthatneedstobeperformedbeforetheinputcanbefedtotheCNN.

Boosted Classifiers [28] are machine learning training algorithms that are based on the idea that a number of rough rulesof classification are easier to find and works better in terms of accuracy and generalization when they are merged together toform a single classification rule with high accuracy. An application of boosted classifiers for highly accurate hand gesturerecognition is proposed in [29]. The pre-processing stage of this method is an adaptive skin segmentation process that performsface detection based on AdaBoost [28] boosted classifier to build a skin

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4822 - 4832
Received 05 March 2021; Accepted 01 April 2021.

model. This adaptive model can be used to identify thepixels of the image that have skin color. Now, the segmented image is again used to train a boosted classifier to identify theregion of the image that contains the hand and put a marker on it so that the detection algorithm need not be executed when thehandchangesposition.Thehandregionisusedtotrainacascadeofclassifiersusingthebootstrapmethod[30]tom akethemodelcapable of distinguishing between object and non-object regions without the need to introduce a lot of non-object examples inthedataset.Activelearning[31]isaprocedureforconstructingthetrainingexamplessemi-automatically.Activelearningisusedin this application to construct the positive training example data set which is used in the final stage to classify the gestures.Although the individual boosted classifiers are accurate on their own data domain, the hand region segmentation failed to begeneric enough to give a good accuracy and led to an overall accuracy that was much lower (63.1% to for fist gesture to 84.3%forfive-fingergestures)thanexpectedwithboostedclassifiers.

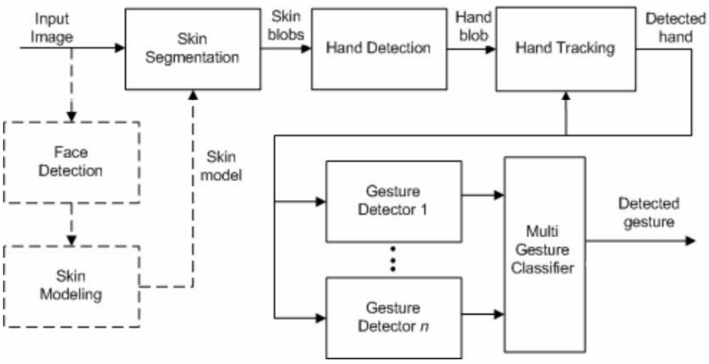

Figure 9. Architecture(top) of a cascaded hand detection and tracking system(upper portion) and thegesturerecognition(bottomportion).(Fig. 1[29])

III. CONCLUSION

A good number of solutions have been provided in this paper regarding the hand gesture detection and recognition withthe aim of helping the disabled people of the society that gives a picture of the progress and ongoing research in the field.However,mostofthemethodsthathavebeenproposedusuallylackthereliability, commercialviabilityandthefundingrequiredtoturnthemintoabusinessproduct.

Boththehardwareandsoftware-basedapproacheshavetheirprosandconsandatradeoffneedstobeconsideredtochoosean efficient, reliable, convenient, generalized and accessible solution. Hardware-based approaches makes the detection andtracking phase easier and accurate but need funding for testing and then additional investment for producing in large numbersandsellingintheworldmarket.Theyalsomaybeconsideredinconvenientinsomecases.Software-basedapproachesarecheap,convenient and if implemented properly very efficient as well. This however depends on the algorithm used for hand trackingand detection only using video stream input and the performance of the classifier on recognizing hand gestures. The model usedforrecognitionandthelearningalgorithmdeterminetheaccuracyoftheclassifier.

Further research in the field will lead to breakthroughs and efficient solutions. The world as a whole will be benefittedwhenitsdisabledpopulationwillbeabletocontributemoresignificantlytothesocio-economicprogress.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4822 - 4832
Received 05 March 2021; Accepted 01 April 2021.

REFERENCES

[1] WorldHealthOrganization,"WorldReportonDisability",WorldHealthOrganization,2011.
[2] J. Lane, *Delivering on the 'D' in R&D: Recommendations for Increasing Transfer Outcomes from Developmental Projects*. Assistive TechnologyIndustryAssociation(ATIA),2008,pp.1-6.
[3] N.Advani, S.Bora, A.BhatandS.Yerolkar, "ASurveyonCommunicationGapbetweenHearingandSpeechImpairedPersonsandNormalPersons",*InternationalJournalofComputerScienceand Network*,vol.2,no.6,2013.
[4] "Factsheet on Persons with Disabilities", *un.org*, 2020. [Online]. Available: https://www.un.org/development/desa/disabilities/resources/factsheet-on-persons-with-disabilities.html. [Accessed:27-Nov-2020].
[5] Y.Amit,*2Dobjectdetectionandrecognition*.Cambridge,MA:MIT Press, 2002,pp.5-12,18-26.
[6] Z.Zhang,"Microsoft KinectSensorandItsEffect",*IEEE Multimedia*,vol.19,no.2,pp.4-10,2012.Available: 10.1109/mmul.2012.24.
[7] X. Ma and J. Peng, "Kinect Sensor-Based Long-Distance Hand Gesture Recognition and Fingertip Detection with Depth Information", *Journal ofSensors*,vol.2018,pp.1-9,2018.Available:10.1155/2018/5809769[Accessed 29 November2020].
[8] Z. Zafrulla, H. Brashear, P. Yin, P. Presti, T. Starner and H. Hamilton, "American Sign Language Phrase Verification in an Educational Game for DeafChildren",*201020thInternationalConferenceonPatternRecognition*,2010.Available:10.1109/icpr.2010.937[Accessed24 November2020].
[9] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton and P. Presti, "American sign language recognition with the kinect", *Proceedings of the 13thinternationalconferenceonmultimodalinterfaces-ICMI'11*,2011.Available:10.1145/2070481.2070532[Accessed24November2020].
[10] W.SandlerandD. Lillo-Martin,*Signedlanguageandlinguistictheory*.Cambridge: CambridgeUniversityPress,2005,pp.144-173,196-215.
[11] A. Sreejan and Y. Narayan, "A Review on Applications of Flex Sensors", *International Journal of Emerging Technology and Advanced Engineering*,vol.7,no.7,pp.95-100,2017.Available: https://www.ijetae.com/files/Volume7Issue7/IJETAE_0717_14.pdf.[Accessed25November2020].
[12] A. Abraham and V. Rohini, "Real time conversion of sign language to speech and prediction of gestures using Artificial Neural Network", *ProcediaComputerScience*,vol.143,pp.587-594,2018.Available:10.1016/j.procs.2018.10.435.
[13] C. Aggarwal,*NeuralNetworksandDeepLearning*.Springer,2018,pp. 1-52.
[14] A. Shukor, M. Miskon, M. Jamaluddin, F. Ali Ibrahim, M. Asyraf and M. Bahar, "A New Data Glove Approach for Malaysian Sign LanguageDetection", *Procedia Computer Science*, vol. 76, pp. 60-67, 2015. Available: https://www.sciencedirect.com/science/article/pii/S1877050915037771.[Accessed18 November2020].
[15] Z. Zhang, Z. Tian and M. Zhou, "Latern: Dynamic Continuous Hand Gesture Recognition Using FMCW Radar Sensor", *IEEE Sensors Journal*, vol. 18,no.8,pp.3278-3289,2018.Available:10.1109/jsen.2018.2808688.
[16] M.Ullah,"TimeSeriesAnalysisandForecasting",*BasicStatisticsandDataAnalysis*,2013.
[17] A. Graves, S. Fernández, F. Gomez and J. Schmidhuber, "Connectionist temporal classification", *Proceedings of the 23rd international conference onMachinelearning-ICML'06*,pp.369-376,2006.Available:10.1145/1143844.1143891 [Accessed20November2020].
[18] A. Sharma, A. Mittal, S. Singh and V. Awatramani, "Hand Gesture Recognition using Image Processing and Feature Extraction Techniques", *ProcediaComputerScience*,vol.173, pp.181-190,2020.Available:10.1016/j.procs.2020.06.022 [Accessed16November2020].
[19] J. Canny, "A Computational Approach to Edge Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. -8, no. 6, pp. 679-698,1986.Available:10.1109/tpami.1986.4767851.
[20] C. Luo, W. Yang, P. Huang and J. Zhou, "Overview of Image Matching Based on ORB Algorithm", *Journal of Physics: Conference Series*, vol. 1237, p.032020,2019.Available:https://iopscience.iop.org/artic

le/10.1088/1742-6596/1237/3/032020.[Accessed17November2020].

[21] R. Xu and D. WunschII, "Survey of Clustering Algorithms", *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645-678, 2005. Available:10.1109/tnn.2005.845141[Accessed20 November2020].
[22] R. Kurdyumov, P. Ho and J. Ng, "Sign Language Classification Using Webcam Images", *Stanford CS229 Fall Projects*, 2011. Available:http://cs229.stanford.edu/proj2011/KurdyumovHoNg-SignLanguageClassificationUsingWebcamImages.pdf.[Accessed20November2020].
[23] Y. Ye, Y. Tian, M. Huenerfauth and J. Liu, "Recognizing American Sign Language Gestures from Within Continuous Videos", *2018 IEEE/CVFConference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. Available:https://openaccess.thecvf.com/content_cvpr_2018_workshops/papers/w41/Ye_Recognizing_American_Sign_CVPR_2018_paper.pdf.[Accessed21November2020].
[24] P.Neethu,R.SugunaandD.Sathish,"Anefficientmethodforhumanhandgesturedetectionandrecognition usingdeeplearningconvolutionalneuralnetworks",*SoftComputing*,vol.24,no. 20,pp.15239-15248,2020.Available:10.1007/s00500-020-04860-5 [Accessed21November2020].
[25] S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey", *Artificial Intelligence Review*, vol. 43,no.1,pp.1-54,2012.Available:10.1007/s10462-012-9356-9.[Accessed21 November2020].
[26] S. Eddy, "What is a hidden Markov model?", *Nature Biotechnology*, vol. 22, no. 10, pp. 1315-1316, 2004. Available:https://faculty.coe.drexel.edu/gailr/ECE-S690-501/Eddy-ATG4-reprint.pdf.[Accessed20November2020].
[27] B. Bauer, H. Hienz and K. Kraiss, "Video-based continuous sign language recognition using statistical methods", *Proceedings 15th InternationalConferenceonPatternRecognition.ICPR-2000*,vol.2, pp. 463-466.Available:10.1109/icpr.2000.906112[Accessed23November2020].
[28] R. Schapire, "The Boosting Approach to Machine Learning: An Overview", in *Nonlinear Estimation and Classification. Lecture Notes in Statistics*, D.Denison, M.Hansen,C.Holmes,B.Mallick andB.Yu,Ed.NewYork,NY:Springer,pp.149-171.
[29] H. Francke, J. Ruiz-del-Solar and R. Verschae, "Real-Time Hand Gesture Detection and Recognition Using Boosted Classifiers and ActiveLearning", *AdvancesinImageandVideoTechnology*,pp.533-547.Available:10.1007/978-3-540-77129-6_47[Accessed22November2020].
[30] Y. RAVIV and N. INTRATOR, "Bootstrapping with Noise: An Effective Regularization Technique", *Connection Science*, vol. 8, no. 3-4, pp. 355-372,1996.Available:10.1080/095400996116811 [Accessed 22 November2020].
[31] B. Settles, "Active Learning Literature Survey", *Digital.library.wisc.edu*, 2020. [Online]. Available: http://digital.library.wisc.edu/1793/60660.[Accessed:22-Nov-2020].