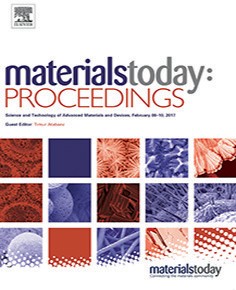
[Materials Today: Proceedings xxx (xxxx) xxx](https://doi.org/10.1016/j.matpr.2019.06.748)



Contents lists available at [ScienceDirect](http://www.sciencedirect.com/science/journal/22147853)

Materials Today: Proceedings

journal homepage: [www.elsevier.com/locate/matpr](http://www.elsevier.com/locate/matpr)

Advanced gesture recognition system using long-term recurrent convolution network

C. Bhuvaneshwari [⇑](#_bookmark0), A. Manjunathan

*Department of ECE, K.Ramakrishnan College of Technology, Trichy 621112 India*

# a r t i c l e i n f o

*Article history:*

Received 13 June 2019

Accepted 30 June 2019 Available online xxxx

*Keywords:* Raspberry pi LRCN

Sensor Gesture Deaf-dumb

# a b s t r a c t

The reliable communication between deaf-dumb and normal people is difficult, because deaf and dumb people use sign language for their communication which is unknown to the normal person. In this research work, to provides solution for problem faced by deaf and dumb people. It deals with an idea of gesture image recognition. In this the concept of embedded system and image processing is combined together which is very effective. This device convert gesture image to voice. The conversion is done using Raspberry pi. Gesture image recognition is processed by implementing LRCN (Long-term Recurrent Convolution Network) algorithm.

© 2019 Elsevier Ltd. All rights reserved.

Peer-review under responsibility of the scientific committee of the International Conference on Recent Trends in Nanomaterials for Energy, Environmental and Engineering Applications.

1. Introduction

Deaf-dumb people face struggle in expressing their thoughts. Their mode of communication is Sign Language. Since there is no common mode of communication between deaf-mute and normal people they suffer lot. They feel discomfort to speak in a public place or in emergency case. Hence a common mode of communica- tion bridge is needed. This system facilitate translation of gesture image to sound. The translation process is done using raspberry pi. In that, gesture image are recognized by implementing LRCN (Long-term Recurrent Convolution Network) algorithm. The input is taken as image and is processed. Finally the output is produced in loudspeaker. Each image has a appropriate voice which is prede- fined and it is stored in voice module.

1. Related work

Balakrishnan reveals that flex sensors are used to get input from various gestures positions through gloves and microcontroller is used to for A/D conversion and finally Bluetooth is used to send data [[1]](#_bookmark6). Jadhav [[2]](#_bookmark7) depicts that flex sensors and accelerometer are used as sensor mounted on the gloves to predict the movement of angle tilt, rotation and direction change. Kaur [[3]](#_bookmark8) explains that Gesture Recognition Using PCA by three steps as image acquisition,

⇑ Corresponding author.

*E-mail address:* [mecbhuvaneshwari@gmail.com](mailto:mecbhuvaneshwari@gmail.com) (C. Bhuvaneshwari).

Hand Segmentation and finally Conversion from RGB to YCbCr. Sethi [[4]](#_bookmark10) elucidates that gesture extraction using Skin color seg- mentation and Region Growing and gesture matching using feature point matching using SIFTS Correlation matching. Learning visual behavior for gesture analysis [[5]](#_bookmark11) deals on gestures are modal in the space of possible human motion, and gestures are viewpoint- dependent. Thomas [[6]](#_bookmark12) explains gesture recognition model using feature extraction and dynamic gesture models. Surachai [[7]](#_bookmark13) depicts pixel wise classification and k-means algorithm to differen- tiate left and right hand gestures. TuKhoa [[8]](#_bookmark14) uses skin colour for hand gesture tracking and recognition.

1. Materials and methods

Components used in proposed method are raspberry pi, camera, voice module, speaker, power supply. [Fig. 1](#_bookmark1) shows the steps involved in conversion of gesture image to speech are input stage, processing stage and output stage. The input is in the form of image.

The image is captured with the help of camera. The hand ges- ture sign is the input image. The input sign language used can be American or Indian sign language. [Fig. 2](#_bookmark3) depicts the initial stage of the camera when no gesture sign is detected and [Fig. 3](#_bookmark4) shows different hand gesture signs to provide information.

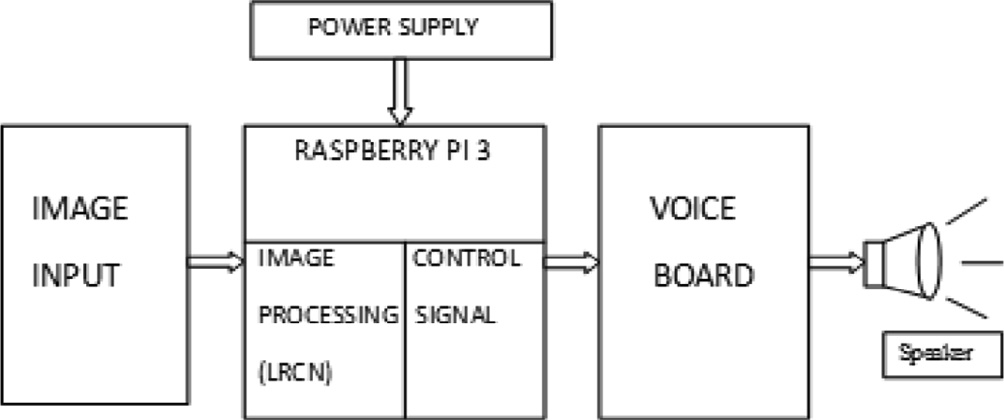
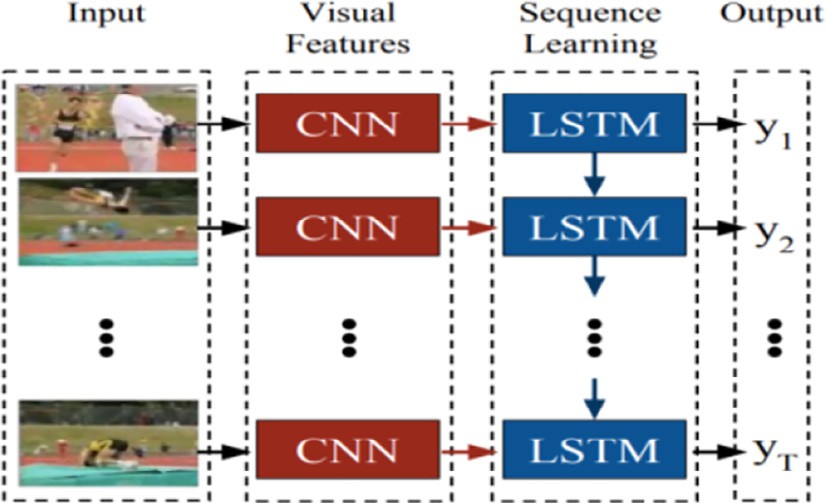
Raspberry pi 3 is the processing platform. The code is written using Python programming language. The image processing is done by implementing LRCN (Long-term Recurrent Convolution

<https://doi.org/10.1016/j.matpr.2019.06.748>

2214-7853/© 2019 Elsevier Ltd. All rights reserved.

Peer-review under responsibility of the scientific committee of the International Conference on Recent Trends in Nanomaterials for Energy, Environmental and Engineering Applications.

2 *C. Bhuvaneshwari, A. Manjunathan / Materials Today: Proceedings xxx (xxxx) xxx*

Fig. 1. Block diagram of proposed method. 

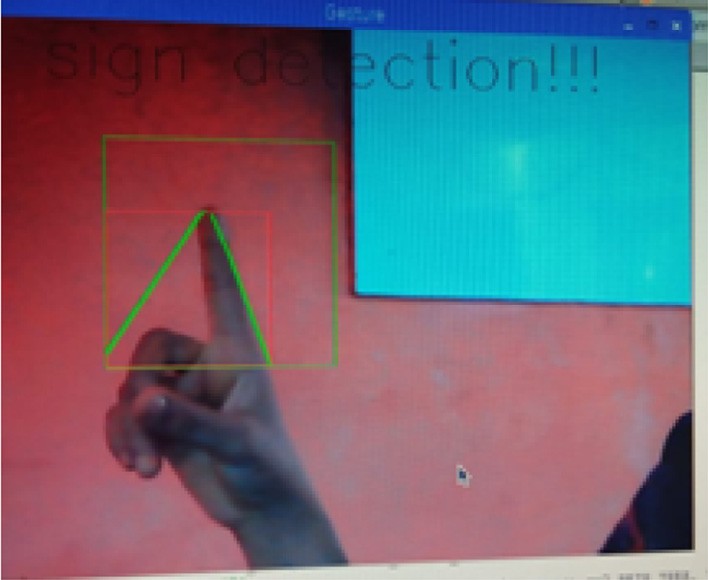
Fig. 4. LRCN works by combing CNN (Convolution Neural Network) and LSTM (Long short-term memory).

Fig. 2. This image depicts the initial stage of the camera when no gesture sign is detected.

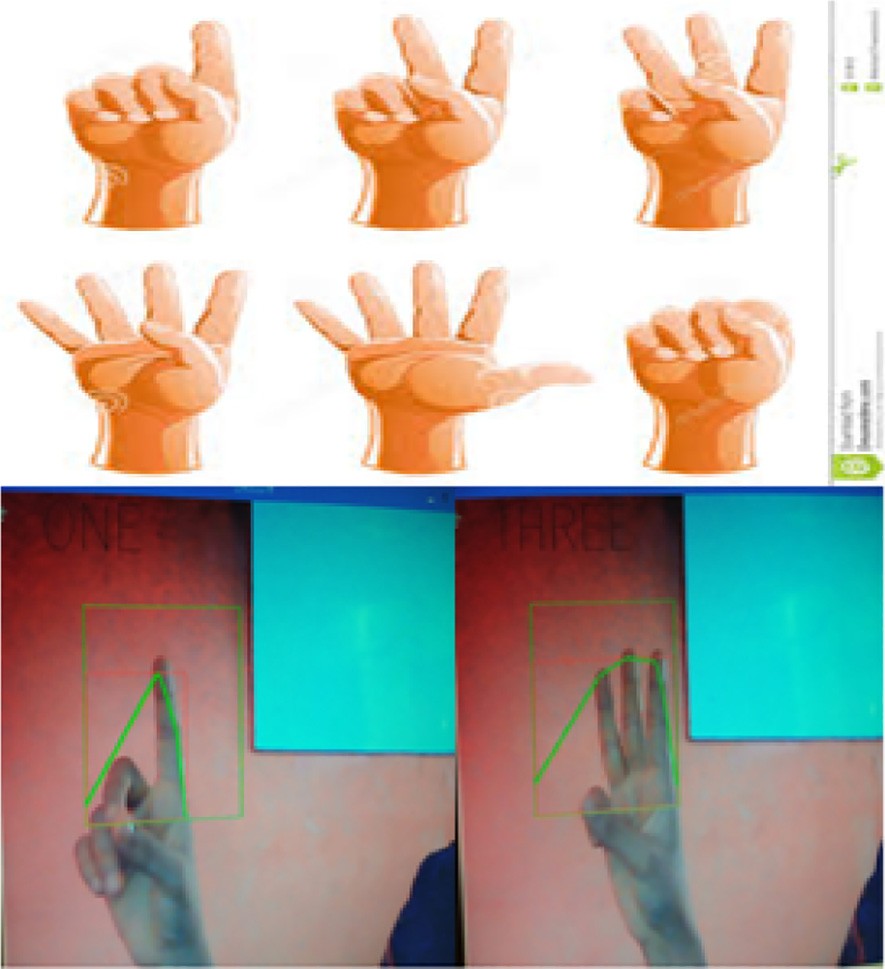


Fig. 3. It represents hand gesture sign. Each sign provide some message or information.

Network). LRCN is a class of architectures for both visual descrip- tion and recognition as shown in [Fig. 4](#_bookmark2). This combines long- range temporal recursion and convolution layer.

CNN is a neural network with some convolution layer. A convo- lution layer has number of filter that does convolution operation. The important function of convolution layer is to extract features. It consists of convolution layer followed by fully connected layer. The pixel which undergoes sliding is called stride. All these result

are concatenated to form feature map. After each convolution pro- cess, the size of the image decreases. Therefore, it is a practice to add zeroes on the boundary of input layer so that the output is same as input. This is called as padding. Next is polling layer. It is used to reduce the processing time by reducing the size of the feature map. The purpose of pooling layer is to reduce the spatial size (height, width). Therefore the number of parameter are reduced and the number of computations is also reduced. Pooling is of three types. They are max pooling, average pooling, min pool- ing, max pooling is most commonly used one. The image size is reduced into half. A flattening layer is used in order to convert mul- tidimensional into a single dimensional. The classification of the image is done by fully connected layer, based on the features extracted by the previous convolution layers.

LSTM (Long Short Term Memory) is a special kind of recurrent neural network. It works well in many problem and most widely used nowadays. LSTM is explicitly designed to avoid the long- term dependency problem. LSTM networks perform better than RNN. This is achieved by multiplicative gate units that learn to open and close access to the constant error flow. LSTM networks introduce a new structure called a memory cell as shown in [Fig. 5](#_bookmark5). Each memory cell contains four main elements they are, input gate, forget gate, output gate, neuron with a self-recurrent. These gates allow the cells to keep and access information over long periods of time.

The LSTM architecture is different from recurrent neural net- work architecture. This enhancement is done to overcome the van- ishing gradient problem. The influence of a given input on the hidden layer, and therefore on the network output, either decays or grows exponentially as it propagates through an RNN.

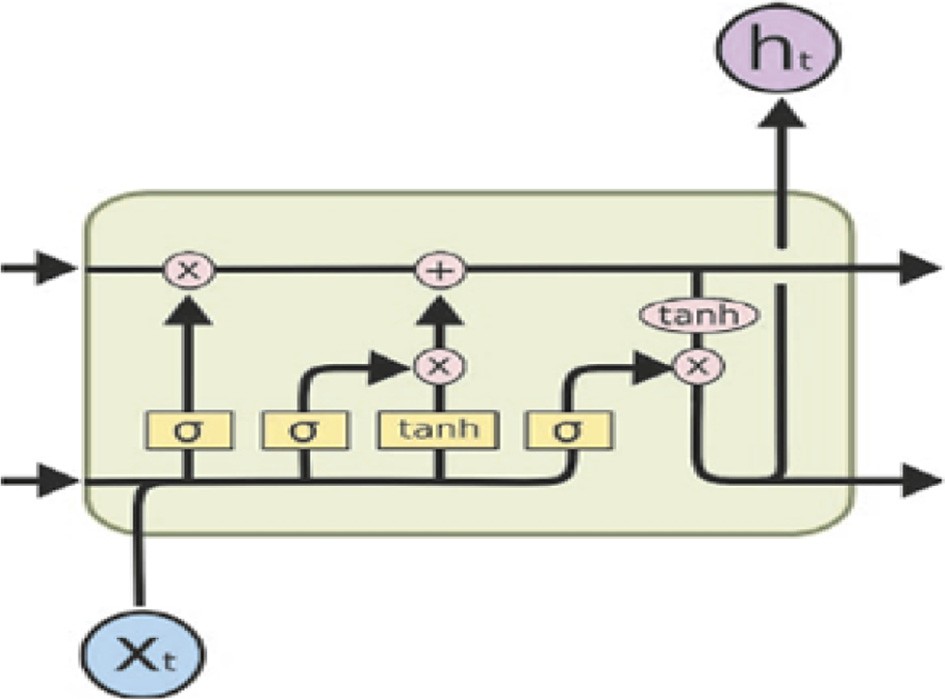
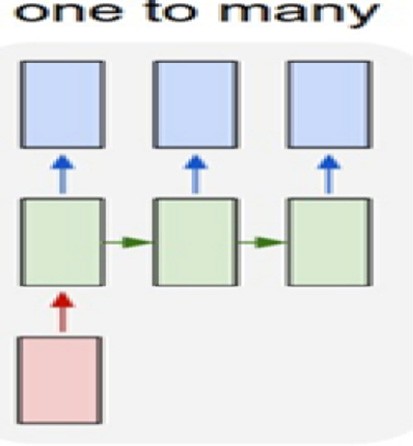


Fig. 5. LSTM memory cell. It consists of input gate, forget gate and output gate.

*C. Bhuvaneshwari, A. Manjunathan / Materials Today: Proceedings xxx (xxxx) xxx* 3

nition algorithm decreases when the distance is greater than 1.5 m between the user and the camera. System limitations restrict the applications such as the arm must be vertical; the palm is facing  the camera.

1. Conclusion and future work

*f* ¼ rð*W* :½*h*

*t*

*f*

Fig. 6. Image captioning.

*x* ] þ *b* Þ

*t*—1;

*t*

*f*

This designed system is very much helpful and provide a com- mon mode of communication between deaf-dumb and normal people. This method takes image as a input and gives speech as an output. This project is the combination of both embedded and image processing, which is implemented using raspberry pi. Hence our system reduces the barrier of communication for deaf and dumb people. The future work of the proposed system is to

increase more number of gesture images for gesture to speech

*it* ¼ rð *Wi* : ½*ht* —1; *xt*] þ *bi* Þ

*C*~*t* ¼ tanh ð*Wc* : ½*ht* 1; *x* ] þ *b* Þ

— *t c*

*Ct* ¼ *f t* m *Ct* —1 þ *it* m *C*~*t*

recognition and using different sign languages.

References

1. R. Balakrishnan, B.K. Santosh, H. Rahul, Shivkumar, Sunil Anthony, Embedded Based Hand Talk Assisting System for Dumb People on Android Platform, 4th National Conference on Design Innovations for 3Cs ‘‘Compute, Communicate, Control”.

*Ot* ¼ ð *W*0

: ½*h*

*t* —1

; *xt*] þ *b*0 Þ

1. Anaghaj Jadhav, Mandar P. Joshi, AVR based embedded system for speech impaired people, 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) International Institute of

*ht* ¼ *Ot* m tanh ð*Ct* Þ

LRCN model can be used for image description, activity recogni- tion and video description. In this we are using LRCN model for image recognition. For image description the input is static and output is sequential one. Both input and output are sequential for video description as shown in [Fig. 6](#_bookmark9). Input is sequential and output is static for activity recognition.

Simple, fast and easy to implement. Can be applied on real sys- tem and play games. No training is required. Speed and sufficient reliable for recognition system. Good performance system with complex background. The system successfully recognized static and dynamic gestures. Could be applied on mobile robot control. Irrelevant object overlap with the hand. Wrong object extraction appeared if the object is larger than the hand. Performance recog-

Information Technology (I2IT), Pune, 2016.

1. Mandeep Kaur, Ahuja Dr. Amardeep Singh, Hand Gesture Recognition Using PCA, IJCSET ([www.ijcset.net](http://www.ijcset.net/)) July 2015 vol. 5(7) pp. 267–271.
2. Ashish Sethi, S. Hemanth, Kuldeep Kumar, N. Bhaskara Rao, R. Krishnan,

SignPro-An Application Suite for Deaf and Dumb, (2) 5 (2012), 1203–1206

1. Andrew Wilson, Aaron Bobick, Learning visual behavior for gesture analysis, in: Proceedings of the IEEE Symposium on Computer Vision, Coral Gables, Florida, pp. 19–21, November 1995.
2. [Ginu Thomas, A review of various hand gesture recognition techniques, VSRD- IJEECE 1 (7) (2011) 374–383](http://refhub.elsevier.com/S2214-7853(19)32227-8/h0030).
3. Stewart Surachai, Ahmet, Two Hand Tracking using Color Statistical Model with the Kmeans Embedded Particle Filter for Hand Gesture Recognition”, 7th Computer Information Systems and Industrial Management Applications, 2008, pp. 201–205.
4. Pham, Nguyen, TuKhoa, A New Approach to Hand Tracking and Gesture Recognition by a New Feature Type and HMM, Sixth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE Computer Society, 2009, pp. 3–6.