Vol.5 No.3 July 2024

Multimedia Research

Received 01 April, Revised 02 June, Accepted 9 July

Controlling the Computer using Hand Gestures

Pradnya Kedari
Department of Computer Science and Engineering, MIT School of Engineering, MIT Art, Design and Technology, Pune, Maharashtra, India

Shubhangi Kadam
Department of Computer Science and Engineering, MIT School of Engineering, MIT Art, Design and Technology, Pune, Maharashtra, India

Rajesh Prasad
Department of Computer Science and Engineering, MIT School of Engineering, MIT Art, Design and Technology, Pune, Maharashtra, India

Abstract: Human-computer interaction platforms have many ways to implement as webcams and other devices like sensors are inexpensive and can get easily be available in the market. The most powerful way of communicating between humans and machines is through gestures. For higher conveyance between the human and machine/computer to convey information, the hand gesture system is very useful. Hand gestures are a sort of nonverbal type of communication that may be employed in several fields. Research and survey papers included hand gesture applications that have acquired different alternative techniques, including those supported by sensor technology and computer vision. In this system, we aimed to build a real-time gesture recognition system using hand gestures. Particularly, we will use the convolutional neural network (CNN) throughout the process. This application presents a hand gesture-based system to control a computer that is performing different operations using a neural network. Our application is defined in five phases, Image frame acquisition, Hand tracking, Features extraction, Recognition of gestures, and Classification (performing the desired operation). An image from the webcam will be captured, and so hand detection, hand shape feature extraction, and hand gesture recognition are done.

Keywords: Computer Vision, Convolutional Neural Network, Deep Learning, Hand Gestures, Python, OpenCV

1. Introduction

Gesture recognition is a popular and in-demand analysis field in Human-Computer Interaction technology. It has several employments in virtual environment management, medical applications, sign language translation, robot control, music creation, and home automation. There has been a special significance recently in the HCI study. The hand is the most helpful communication tool in several body parts, because of its expertise. The word gesture is employed for several cases involving human motion, particularly of the hands, arms, and face; just some of these are informative.

Convolutional neural networks are the most popular employed technique for the image classification task. An image classifier takes an input image, or input sequence of images, and categorizes them into one among the possible classes that it was trained to classify. They have applications in different fields such as the medical domain, self-driving cars, educational domain, fraud detection, defense, etc. There are several techniques and algorithms for image classification tasks and also there are some challenges like data overfitting. During this project Controlling Computer using Hand Gestures, we aimed to make a real-time application using OpenCV and Python. OpenCV is a real-time open-source computer vision and image-processing library. We'll use it with the help of the OpenCV python package. Fig 1 demonstrates the methodology of the proposed system.

https://doi.org/10.46253/j.mr.v5i3.a2

Fig.1. *Methodology of Proposed System*

1.1 Market Survey

Over the traditional mechanical communication technologies, the gesture recognition system has become known as the most popular technology. The domain market is divided on different bases like Technology, Type, Practice, Product, Use, and Geography. Additionally, Assistive robotics, Sign language detection, Immersive gaming technology, smart TV, virtual controllers, Virtual mouse, etc is divided.

1.2 Research Gap

- Most of the methods used Arduino and sensors, direct device webcam is used in very few methods.
- There might be miss-recognitions of gestures in case the background environment has elements that appear like human skin.
- The hand should be within the range limit.
- Dataset overfitting is the main concern.

2. Literature Survey

In the computer science and technology era, gesture recognition is a crucial field that can use to translate human gestures using different computer vision techniques and algorithms. Numerous human body motions can create gestures but the most common type of gesture generation stands up from the face and hands. The complete policy of tracking gestures to their representation and changing them to some useful command is referred to as gesture recognition. Different techniques and methods have been employed for the design and development of such kind of task.

The starting approach of interaction with a computer using hand gestures was first projected by Myron W. Krueger in 1970 [1]. The aim of the perspective was achieved and also the mouse cursor control was performed with the help of an external webcam (Genius FaceCam 320), a software package that would interpret hand gestures and then turn the recognized gestures into OS commands that controlled the mouse actions on the display screen of the computer device [2]. Choosing hand gestures as a communication tool in HCI will allow the development of a good variety of applications in the absence of physical contact with computing devices [3]. At present, most of the HCI depends on the devices such as a keyboard, or mouse, however, growing importance in a category of methods and techniques based on computer vision has been popular because of the skill to recognize human gestures simply [4].

Detection and recognition of a particular human body gesture and carrying information to the computer is the main objective of gesture recognition. The overall objective of this system was to create human gestures which can be admitted by computer devices to control good sorts of devices that are at distance using different hand poses [5]. Hand gesture recognition is based on robotic computer vision to handle the devices such as digital TV, play stations, etc. Hand gesture recognition for sign language was considered a weighty research area lately. But because of different issues, like skin tone color difference, the complex and disturbing environments, and also the different static and dynamic hand gestures, the common problem of that system was raised. Hand gestures recognition for the management of TV is recommended by [6]. In this, only one hand moment is used to control the TV. A hand picture looks like an icon that follows the hand movements of the user appearing on the screen display on the TV. In this

paper [7], the actual human-computer interface which is the HCI model which is based on hand gestures and uniquely accepts gestures to operate using a monocular camera and assist the system in the HRI case has been developed. The evolved system relies on a classifier based on Convolution Neural Network to extract features and recognize particular hand gestures.

The HMM is Hidden Markov Model is considered a crucial tool for dynamic gesture recognition in actuality. The system employed by HMM operates at present and the general aim to build this system is to operate in static environments. The proposed methodology was for training, to employ the topology named LRB of HMM with the Baum Welch Algorithm and also for testing, the Forward and Viterbi Algorithms and checking the input sequences, and building the maximum productive achievable pattern recognition state sequence [8]. In this paper [9], even the developed model seems to be easy to handle as compared to the newest available system or command-based system but the drawback is developed system is less powerful in spotting the gestures and recognizing the same. So despite the complex environmental background and a normal lighting environment background, an existing system need to improve and requires to build further a good network for gesture recognition. This system is built for a total of six classes. However, this existing model can be used to control operations such as PowerPoint presentations, windows picture managers, media players, games, etc. In this paper [10], using an Arduino Uno and Ultrasonic sensors, operations such as handling the media player, and volume increase/decrease are performed on the laptop. Arduino, Ultrasonic sensors, and Python are used for serial connection. For interactive and effective learning, such type of system can be used in teaching classrooms.

Hand Gesture recognition systems based on devices like Arduino UNO and several ultrasonic sensors manage a device where they can control VLC by involving operations like playing and pausing videos and also for page scrolling up and down [11]. This paper [12] suggests a convenient hand gesture monitoring system based on ultrasonic sensors, which is built using Arduino microcontroller ATMEGA32. It is claimed that extra hardware is not required to classify hand gestures and also claimed that simple low-cost ultrasonic sensors can be used to notice different range limits to identify hand gestures. In this paper [13], the hand gesture system relied on Arduino UNO and python programming with the wired ultrasonic sensor developed to manage a device and they included operations like zoom in/out and image rotation, etc. This trial is the successful trial of working a hand motion sensing system using sensors and Arduino kits in wireless mode radio frequency. A hand gesture recognition system for Microsoft Office and media players with their own dataset is developed. [14].

3. Implementation and Working

3.1 Dataset

We tried to use the available dataset, but we faced an overfitting problem. Thus we create our own dataset for training the model. We took a total of 10 different hand gestures to perform activities like opening WhatsApp, PowerPoint presentation, Microsoft Word, Microsoft Edge, Google Chrome, Video Player, Xbox, Paint, etc. We took a total of 3000 images for training, 2000 images for testing, and 500 images for validation. Fig 2 exhibits the sample dataset images.

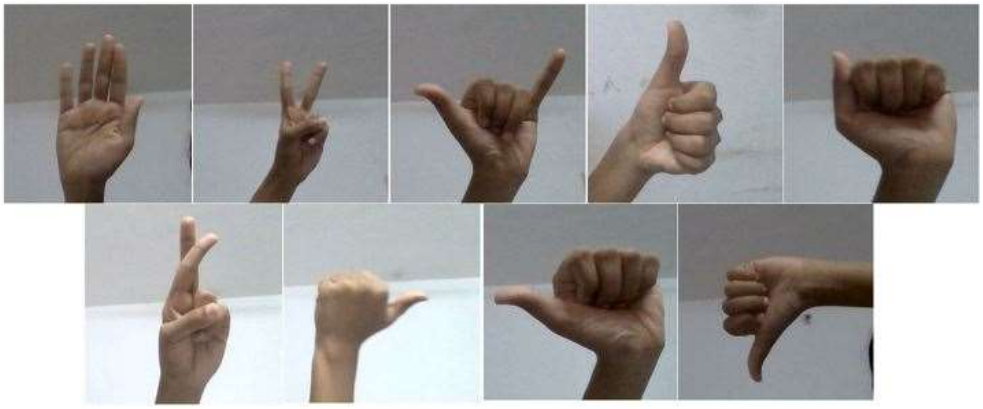

Fig.2. *Sample Dataset Images*

3.2 Implementation

(i) Train The Hand Gesture Recognizer Model

In this module, first glance at the dataset. We have a total of 3000 images for training, 2000 images for testing, and 500 images for validation. We have performed image data augmentation, and as somebody said "Keras ImageDataGenerator is a gem!" it lets us augment our images in real-time while the model is still training. We can apply any random transformations to each image from the dataset as it is passed to the model. This will not only make our model powerful but will also lay aside the overhead memory!

Now our next task is to train our hand gesture detector model. For that, we have used Convolution Neural Network (CNN) architecture revised from Squeezenet and VGGNet. Learned features of CNN remain hidden and thus it is used as a black box.

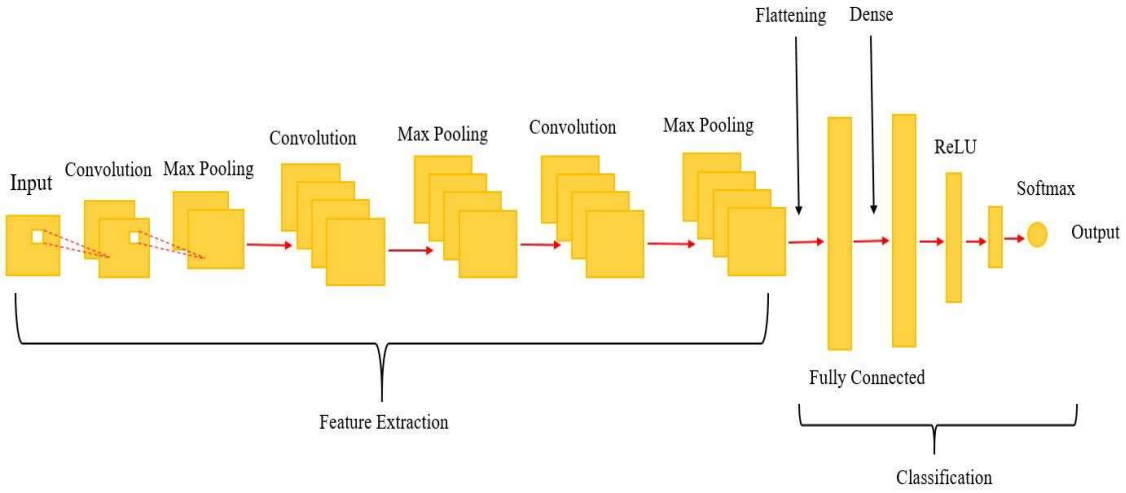

Fig.3. *CNN Architecture*

Fig 3 illustrates the CNN architecture. In this, for creating the model, we have used VGG19 (inception learning). VGG is a successor of AlexNet. VGG19 is a deep convolution neural network having a convolution layer, pooling layer (max pool), fully connected layer, ReLU, and softmax in its architecture. The first step is the convolution operation. In this step, feature detectors are mapped, which basically serve as the neural network's filters. The second step is pooling, pooling layers are used to minimize the dimensions of the feature maps. However, will use a specific type of pooling, max pooling. Max pooling selects the maximum features for the next layer. The third step will involve flattening, flatten is the function that converts the pooled feature map to a single column that is passed to the fully connected layer. Fully Connected layers, in this part, everything that we covered throughout the section will be merged together. Dense adds the fully connected layer to the neural network. In the dense layer, neurons are supposed to connect deeply. Rectified Linear Unit or ReLU is a linear function that will output the input directly if it is positive, otherwise, it will output zero. Mathematically, it is defined as y = max (0, x). The softmax function transforms input values into values between 0 and 1 so that they can be interpreted as probabilities. The final output of this layer will always remain between 0 and 1. For this reason, it is usual to append a softmax function as the final layer of the neural network. Softmax is often used as the final layer in the network, for a classification task.

We used the Adam optimizer because the accuracy rate of this optimizer is much better than others. The whole workflow of our proposed system is explained in the following image Fig 2,

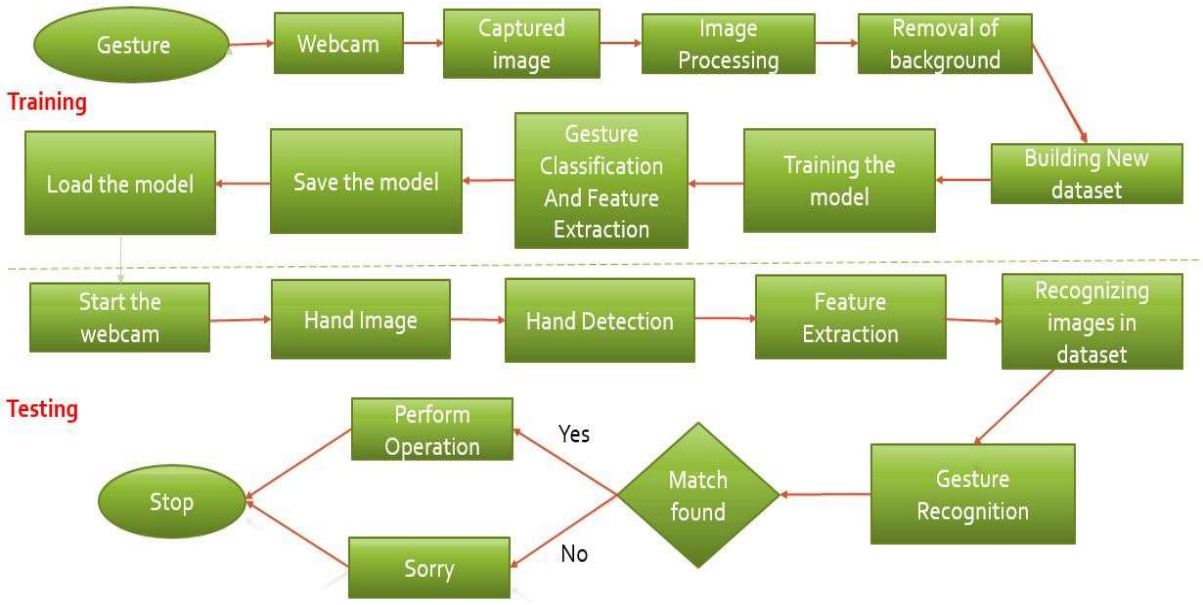

Fig.4. *Workflow of Proposed System*

Now after the creation of the model, we compiled and trained the model. Then after checking the accuracy, visualization of results by plotting graphs, and saving the model we finished with the training module. Here, the complete task is taking the image of the hand gesture as input using the webcam and then compressing the image by using the CNN algorithm to match the images in the dataset in order to detect the hand moment accurately. The captured image is preprocessed, and a hand detector tries to filter out the hand image from the captured image. A CNN classifier is used to recognize gestures from the processed image after feature extraction. Fig 4 shows the workflow of the proposed system.

(ii) ***Recognize The Hand Gesture and Perform Operations***

Now we have a trained model. This model will detect the hand, then features will be extracted, and if the gesture is recognized then a particular operation will be performed. We also added one more functionality in our implementation that tells the user which action is performed.

4. Result and Analysis

Fig 5 indicates the plots of train-Val accuracy. We got 80.40% validation accuracy. Fig 6 demonstrates the train vs-Val loss.

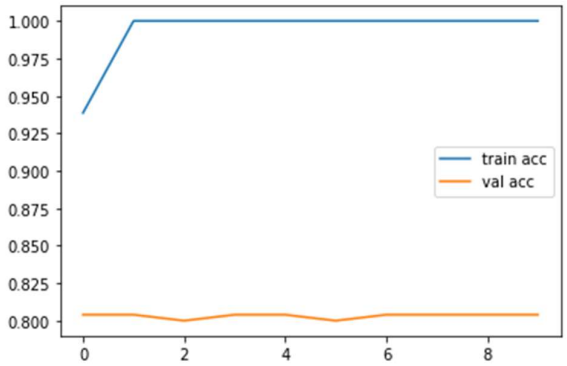

Fig.5. *Train vs Val accuracy*

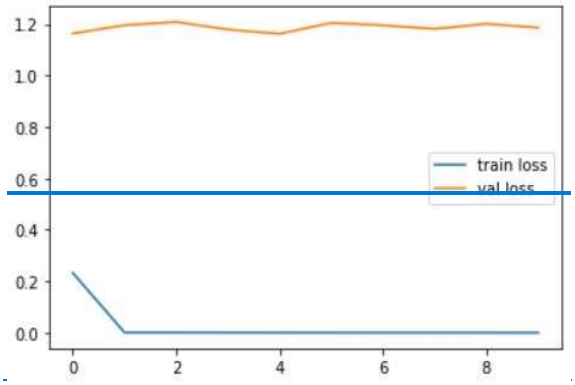

Fig.6. *Train vs Val loss*

We achieved the following result of evaluation metrics. Our model is confused in some of the gestures like a cross, scissors, up, etc. We achieved 85.90% accuracy at the time of testing our model. Following are the classification reports we got. Fig 7 demonstrates the overall result of the model. Fig 8 illustrates the confusion matrix.

	precision	recall	f1-score	support
0	0.30	0.73	0.43	84
1	1.00	1.00	1.00	200
2	1.00	0.87	0.93	229
3	0.99	0.96	0.98	206
4	1.00	1.00	1.00	200
5	0.86	1.00	0.93	173
6	1.00	1.00	1.00	200
7	1.00	0.47	0.64	422
8	0.47	1.00	0.64	94
9	0.96	1.00	0.98	192
accuracy			0.86	2000
macro avg	0.86	0.90	0.85	2000
weighted avg	0.93	0.86	0.87	2000

Fig.7. *Overall Result of Model*

```
[[ 61   0   0   0   0   0   0   0  23   0]
 [  0 200   0   0   0   0   0   0   0   0]
 [  0   0 200   2   0  27   0   0   0   0]
 [  0   0   0 198   0   0   0   0   0   8]
 [  0   0   0   0 200   0   0   0   0   0]
 [  0   0   0   0   0 173   0   0   0   0]
 [  0   0   0   0   0   0 200   0   0   0]
 [139   0   0   0   0   0   0 200  83   0]
 [  0   0   0   0   0   0   0   0  94   0]
 [  0   0   0   0   0   0   0   0   0 192]]
```

Fig.8. *Confusion Matrix*

Fig 9 illustrates the gesture-detecting frame. Below are some demons of the developed system. We assigned a total of 10 hand gestures to perform different operations. Fig 10 demonstrates the opening the Whatsapp. Then, Fig 11 illustrates the opening the google chrome, and Fig 12 demonstrates the opening of the PowerPoint presentation.

Fig.9. *Gesture Detecting Frame*

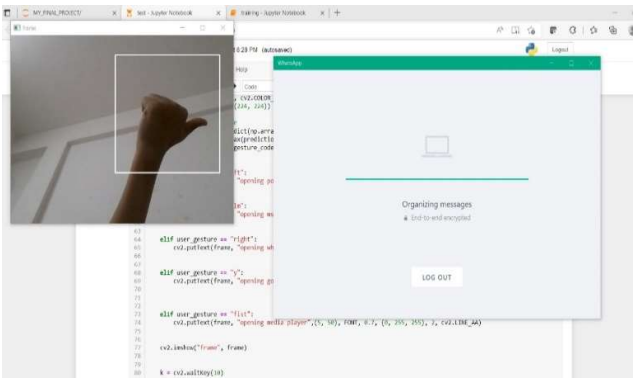

Fig.10. *Opening the WhatsApp*

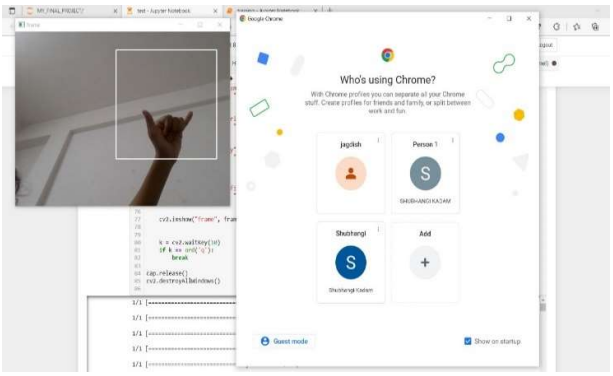

Fig.11. *Opening the Google Chrome*

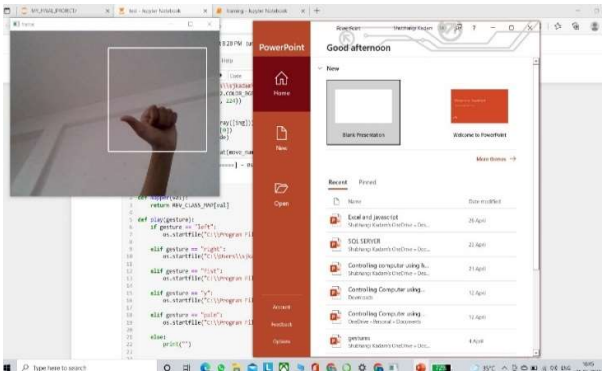

Fig.12. *Opening the PowerPoint Presentation*

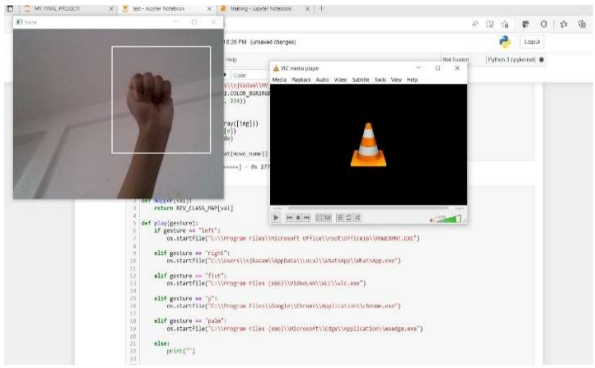

Fig.13. *Opening the VLC*

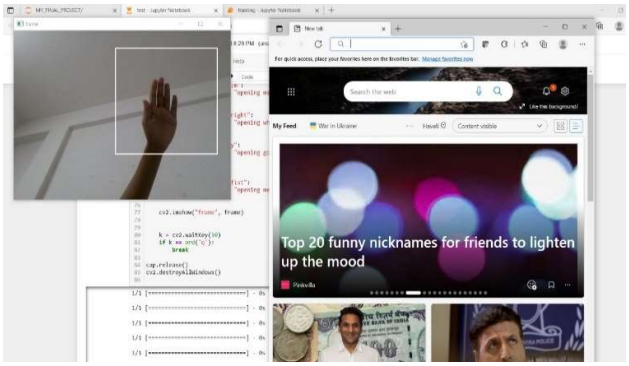

Fig.14. *Opening the Microsoft Edge*

Fig 13 illustrates the opening of the VLC and Fig 14 shows the opening of the Microsoft edge.

5. Conclusion

With the growth of present technology, and as humans generally make use of hand movements that is hand gestures in daily communication in order to make intentions more clear, hand gesture identification is treated to be a crucial portion of Human-Computer Interaction (HCI), which provides devices the capability of detecting and classifying hand gestures, and perform activities subsequently. Research and analysis in the field of hand gestures have become more popular and exciting. It also allows a way of natural and simple interaction. Standard interactive techniques are based on several tools like a mouse, keyboard/touchpad, or touch screen, a joystick for gaming, and consoles for system management. In this paper, we also have discussed the overall review of gesture acquisition methods, the feature extraction process, the classification of hand gestures, and the challenges that face researchers in the hand gesture recognition process. In this application, we developed a deep-learning model for controlling a computer using hand gestures with the help of Python and OpenCV. It is a cost-effective model as we are not using any extra devices and sensors. We can define a project as creating a suitable dataset, training the model, and testing this model in real time. This project has limited scope, we assigned a total of 10 hand gestures to perform different operations, but in the future, we can add more operations like volume up/down, scroll up/down, swipe left/right, and many more, and can be possible to make completely hand gestures controlling device. Hand gesture recognition is used in many different applications like robotics, sign language recognition, HCI, digit and alphanumeric value, home automation, medical applications, gaming, etc. Hand gesture recognition provides an interesting interaction field in several different computer science applications.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

[1] M. KRUEGER, Artificial reality II, Addison-Wesley Reading (Ma), 1991.

[2] Horatiu-Stefan Grif and Cornel Cristian Farcas, "Mouse Cursor Control System Based on Hand Gesture", 9th International Conference Interdisciplinarity in Engineering, INTER-ENG 2015, 8-9 October 2015, Tirgu Mures, Romania.

[3] H.A JALAB, "Static hand Gesture recognition for human computer interaction", Asian Network for Scientific Information technology journal, pp. 1-7, 2012.

[4] J C.MANRESA, R VARONA, R. MAS and F. PERALES, "Hand tracking and gesture recognition for humancomputer interaction", Electronic letters on computer vision and image analysis, vol. 5, pp. 96-104, 2005.

[5] H. HASAN and S. ABDUL-KAREEM, "Static hand gesture recognition using neural networks", Artificial Intelligence Review, vol. 41, pp. 147-181, 2014.

[6] W. FREEMAN and C. WEISSMAN, "Television control by hand gestures", Proc. of Intl. Workshop on Automatic Face and Gesture Recognition 1995, pp. 179-183, 1995.

[7] Pei Xu, "A Real-time Hand Gesture Recognition and Human-Computer Interaction System", arXiv:1704.07296v1 [cs.CV] 24 Apr 2017.

[8] J.R Pansare, Malvika Bansal, Shivin Saxena, Devendra Desale, "Gestuelle: A System to Recognize Dynamic Hand Gestures using Hidden Markov Model to Control Windows Applications", International Journal of Computer Applications (0975 – 8887) Volume 62– No.17, January 2013.

[9] Ram Pratap Sharma and Gyanendra K. Verma, "Human Computer Interaction using Hand Gesture", Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015).

[10] Rohit Mukherjee, Pradeep Swethen, Ruman Pasha, Sachin Singh Rawat, “Hand Gestured Controlled laptop using Arduino”, International Journal of Management, Technology And Engineering, October 2018.

[11] Ayushi Bhawal, Debaparna Dasgupta, Arka Ghosh, Koyena Mitra, Surajit Basak, “Arduino Based Hand Gesture Control of Computer”, IJESC, Volume 10, Issue No.6, June 2020. [

[12] Udit Kumar, Sanjana Kintali, Kolla Sai Latha, Asraf Ali, N. Suresh Kumar, “Hand Gesture Controlled Laptop Using Arduino”, April 2020.

[13] Sarita K., Gavale Yogesh, S. Jadhav, “HAND GESTURE DETECTION USING ARDUINO AND PYTHON FOR SCREEN CONTROL”, International Journal of Engineering Applied Sciences and Technology, 2020, Vol. 5, Issue 3, ISSN No. 2455-2143, July 2020.

[14] Ram Pratap sharma, Gyanendra Varma, “Human computer interaction using hand gesture”, 2015.