International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 11, ISSN No. 2455-2143, Pages 88-93
Published Online March 2023 in IJEAST (http://www.ijeast.com)

HEAD TRACKING AND VOICE RECOGNITION USING OPEN CV IN PYTHON

Umang Jain, Isha Keshariya, Vrinda Arora, Atharv Misal
Dept. of CS
NMIMS University
Mumbai, Maharashtra, INDIA

Ratnesh Chaturvedi
Assistant professor
Dept. of CS
NMIMS University, Mumbai

Abstract— **It is common knowledge that computers are an organization's central nervous system in the modern world. They can store their information on computers and retrieve it whenever required. As a result, having a working grasp of computers is essential. People with impairments cannot perform the tasks that need to be done on the computer. This paper outlines the proposed system for the disabled, which primarily uses the head tracking element utilizing software that mainly consists of head detection and voice recognition. The document also briefly discusses the system's potential benefits for the disabled. Thanks to the head detection feature, they can move the pointer across the screen and operate any software by simply moving their head. Thanks to the implementation of speech recognition technology, they can use their voices to engage with the system. Actual individuals validated the system through testing to ensure it met their needs and effectively provided assistance. These devices are also incredibly inexpensive and disabled individuals who rely on the care and support of others for survival can purchase them.**

Keywords—**Disability, Wearable device, Speech Recognition, Head Recognition Facial Gesture Recognition, low-cost device,webcam, sensors**

I. INTRODUCTION

Many people in our society cannot perform daily activities due to physical limitations. In the modern world, computers are necessary to operate any and all businesses. A database contains all of the information that is relevant to the companies. Many of us may have the misconception that using computers is simple.

However, there are still a great number of people that struggle when using computers. Even though AI-based equipment such as artificial hands and legs are already on the market, these prosthetics are still prohibitively expensive.

We can create a line of goods that are accessible to people with a variety of various types of disabilities. They could assist in obtaining comfort while utilizing computers by developing specialized, lightweight devices. The designers created these items to assist users in carrying out fundamental tasks on their smart gadgets by wearing them. The operations may include anything from launching an application to keying in some information for a customer. This project's scope is vast since few products are in the market for specially-abled people. The products that are successful in their implementation are costly, making them unreachable for middle- and lower-class people.

Looking at the market survey: A few companies have applied this idea of aiding the disabled, but they are not very well-known. It doesn't matter how widespread the use of a given gadget may be; its price is likely to be high.

Somebody found that these kinds of wearable devices are available in India. However, it's for Rs. 15000. The device is shown Figure 1 [21]. Furthermore, it's very costly for people with disabilities who cannot earn for themselves to buy these products. Hence, we have decided to design the device to be easily reusable and affordable.

Figure1: Wearable devices

There is a good chance that there can be an increase in demand for devices of this kind in the not-too-distant future as soon as people become aware that such gadgets can acquire at rates within a reasonable range. There is a perfect chance that this can take place. Therefore, this is an appropriate time to implement the planned method.

II. LITERATURE REVIEW

Developers develop many multi-modal systems using head movement and speech recognition [1]. The designed system was put to use for hands-free operations with Graphical User Interface in a variety of tasks, including text editing in Microsoft Word and communications over the internet [2]. Research is still looking for methods that can accurately detect and track eye and head movements [3]. The head-tracking technique enables many disabled individuals to handle the computer mouse and virtual slides without using their hands or a webcam [4]. Even this kind of device developed uses eye tracking. But the devices are not that much reliable and are not useful for practical application [5]. The developers designed another multi-modal system with head tracking, speech recognition, and tongue motion.

And it was observed that repeated usage of these devices improved the task completion time and typing accuracy [6]. Two types of Head tracking with facial gesture recognition were designed and compared – a web camera-based and another using a Kinect sensor. The results suggested that the Kinect system performed best [7].

Even for the head tracking part, many researchers suggested using LCD, which was quite popular because of their lower prices. But, the only issue it had was with window management and application switching [8]. It is possible to integrate eye tracking with a mouse for a single monitor, which will cut down on the amount of mouse movement needed. Eye tracking is much more difficult when many displays are used since this causes the head to move through a variety of postures and angles [9]. When a person walks, talks, grins, or even touches the eye-tracking glasses or their face, the head may move slightly. This can lead to substantial inaccuracies in gaze-direction signals and tracking loss [10].

An algorithm for detecting head motion based on the Adaboost face detection algorithm can be proposed, which is not dependent on specific biometric identification and tracking. This algorithm enables head movements such as tilting up and down, moving left and right, and opening and closing the mouth [11]. With a camera, the Head-Trace Mouse software system for the disabled can monitor the user's head and mouth motions to replace the standard mouse. Some individuals with issues in their upper limbs have tried and appreciated the technique. [12].

The AdaBoost algorithm underpins the head movement detection and efficiently turns user input into on-screen mouse clicks and scrolls. Adding the AdaBoost algorithm significantly improves the approach [13]. It is essential to develop a system or assistive computer software that accurately depicts the basic functions of a computer mouse through the camera of a computer system, using the AdaBoost algorithm, so that people with disabilities may use it efficiently. The researchers tested the effectiveness of the proposed algorithm in a real-world setting with the highest attainable degree of precision. The suggested algorithm's primary goal is to enable hands-free human-computer interaction and mouse motion control using head movements; the system can provide outcomes as anticipated [14].

In the proposed method, a disabled individual or user must sit in front of a computer camera. The camera detects the location of the user's head. Utilizing AdaBoost, the mouse cursor goes from one location on the desktop to another. No extra hardware is attached, and head movement is the only requirement.

These Head movements are outlined below:
∑= {M0, M1, M2, M3, M4} Where;

M0: Standard Head,
M1: Head Left,
M2: Head Right,
M3: Head Up,
M4: Head Down

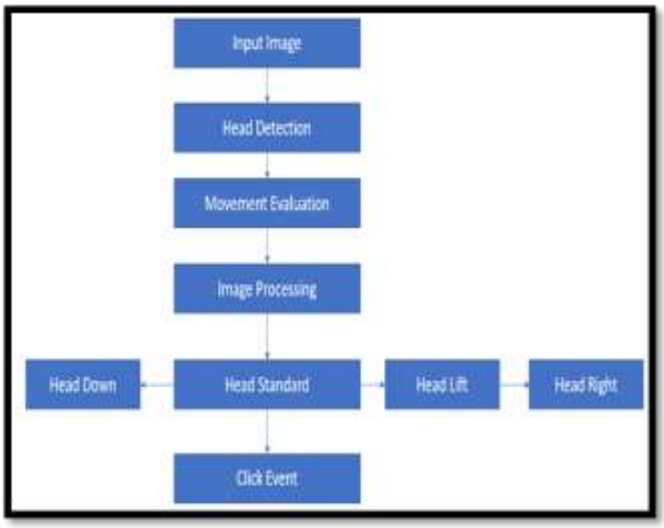

Figure 2: Flowchart of the Proposed Algorithm

Users or those with disabilities are able to conduct all Mouse operations. Figure 2 depicts the procedure of the algorithm [14].

Therefore, a method for head gesture recognition that is based on machine learning and uses a wearable device can identify six common head gestures that we make in our day-to-day lives. These head gestures include nodding, shaking, raising our heads, tilting our heads, turning our heads to the left, and turning our heads to the right. We recommend HGR-DT (Head Gesture Recognition using Decision Trees) and HGR-RF (Head Gesture Recognition using Decision Trees using Random Forest), respectively, based on the classifiers that were utilized in the proposed techniques [15].

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 11, ISSN No. 2455-2143, Pages 88-93
Published Online March 2023 in IJEAST (http://www.ijeast.com)

"hMouse" is the name given to the innovative head-tracking and camera mouse technology that is currently under development for use with hand-free perceptual user interfaces [16].
hMouse calculates the user's head roll, tilt, yaw, scaling, horizontal, and vertical movements in order to provide mouse control. This calculation is based on the findings of precise tracking. In order to browse and fine-tune the position of the pointer, it is necessary to calculate the relative position of the tracking window in image space in relation to the user's head tilt or yaw rotation. After the movement of the mouse pointer to the desired location, activating the virtual mouse buttons is accomplished by rotating the head.

The results of the experiments indicate that hMouse is effective under the following conditions:
1. user jumping,
2. excessive movement,
3. big degree rotation,
4. turning around,
5. hand/object occlusion,
6. portion face out of camera shooting region, and multi-userocclusion.

It provides several alternatives for operating devices in a convenient manner, so boosting interactive computer games, machine directing, robot control, and access to machines for people who are disabled or elderly [17].
The Figure 3 shows that a 3D head posture includes a rotation in the tilt, yaw, and roll axes. They determine the relative location of the tracking window in the image space and map that information to the position of the mouse in the screen space. In addition to that, they used the yaw rotation and the head tilt as means of fine-tuning. The user must first position the pointer in the appropriate location before retrieving the head roll rotation in order for the mouse buttons to function. Once the cursor is in the appropriate location, the user can proceed to use the mouse. In addition, the threshold of the tracking window's size is utilized by them in order to define the hold and release mouse event. The hMouse system is a prototype camera mouse that is driven by 2D head tracking. It has recently been developed [17].

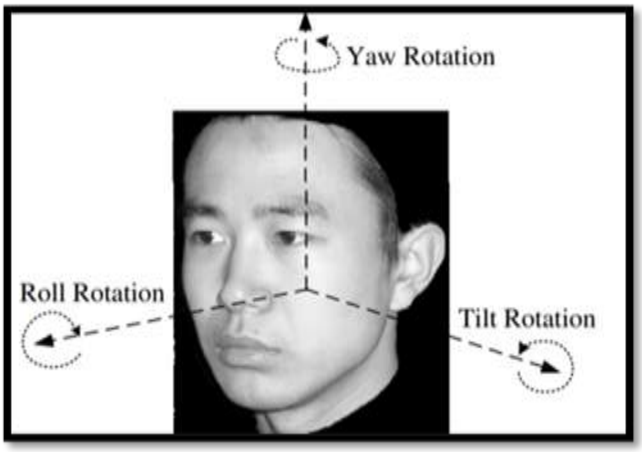

Figure 3: 3-D head pose and parameters

Using a computer without using your hands through head movements is already challenging. Nevertheless, when individuals with limited head control are required to communicate with a computer using only a single head gesture, such as a nod or a tilt, the task becomes considerably more challenging for them. Although if the goal of universal access is to remove barriers for people who have impairments, the high cost of specialised equipment employed by the majority of head-operated HCI solutions presents a new challenge.
Alternative computer access methods can help people with motor limitations communicate and learn in all aspects of life. Innovative approaches to computer interface that need only slight head motions. With less influence over people's lives, better their quality of life and help them become more independent. Thus, two innovative interaction methods, HeadGyro and HeadCam, use a smartphone's gyroscope sensor and a regular camera.
Software switches like HeadGyro and HeadCam do not require specialised devices and can be configured to function using switch-accessible interfaces instead. Both software switches have the ability to imitate the operation of physical switches by tracking head motions using the camera or gyroscope sensor of a smartphone. As can be seen in Figure 4, each of the six different head motions (also known as rotating head movements) will lead to one of the six different possible intersections. The movements of pitch (Fig. 4a), yaw (Fig. 4b), and roll (Fig. 4c) control the position of the earth, while yaw (Fig. 4b) and pitch (Fig. c) define the location of the red border lines on the right and left.
Once the interaction is complete, the system restarts the gravity function (i.e., the switch is released). In this manner, if the user maintains head stability for some time when there is no intersection, gravity gradually drags the earth back to its initial position (i.e., to the centre).

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 11, ISSN No. 2455-2143, Pages 88-93
Published Online March 2023 in IJEAST (http://www.ijeast.com)

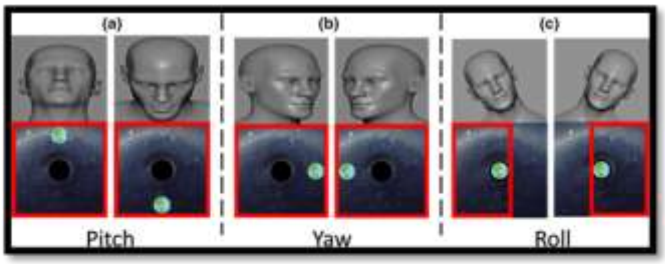

Figure 4: Six distinct junction states of the interface based on head rotational movement.

People with mobility impairments can benefit from gesture-based, hands-free human-computer interaction enabled by head movement. The most popular method for recognizing head movements now is a motion-based sensor. Detecting and identifying head movement is necessary to operate a robotic manipulator in medical aid properly. Researchers have yet to develop the best methods for evaluating head angular movements. In this study, we combine the algorithms for the vision sensor with the gyro sensor to accurately detect changes in head orientation. When the meal-assistance robot manipulator was employed while the user was seated, the system used data distribution to determine the user's head orientation.

The system assesses its precision by using a vision sensor and a gyro sensor. According to the findings of the experiments, the average accuracy of a correct head movement is 82% [18].

The system block diagram can be broken down into two primary components. The first component is a webcam made by Logitech called the C270. Its primary purpose is to record facial movements. The second component is a gyroscope sensor with the model number MPU6050. Its function is to detect any movements made by the user's head. OpenCV is a piece of software for processing digital images that is installed on a personal computer and used to analyse the data obtained from a camera.

The system uses a microcontroller to manage the data obtained from the gyroscope sensor, and then transmits the processed data to a computer via a serial data link. Both sets of data utilize a map of the head motions, which allows the head movements to be classified according to the direction in which the head moves. Controls for the position command can be determined using this information [18].

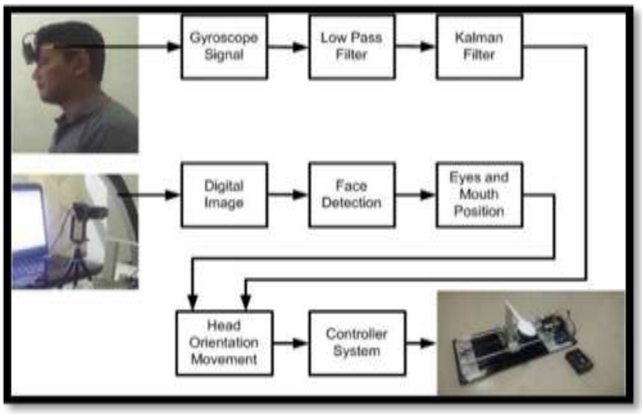

Figure 5: 3D model-based visual face-tracking technique

The researchers can introduce a novel camera mouse that is driven by a 3D model-based visual face-tracking technique shown in Figure 5. The mouse control can be used in three different modes: direct mode, joystick mode, and differential mode. These modes are determined by the calculated parameters for rigid and non-rigid face motion.

The researchers measured the accuracy of mouse navigation with the three mouse control modes using numerical data. The most effective approach to exercise control over a mouse is to employ a combination of the direct and joystick modes. With the direct mode, you have the ability to move the mouse cursor to a specific location on the screen. After that, the user can switch to the joystick mode to more precisely control where the pointer is located. If a non-rigid alteration is detected in the contour of the face, then any of the two modes can be activated or deactivated [19].

It uses standard infrared (IR) sensors to track head movements and a standard microcontroller (MCU) to map the sensor data to either relative movement (joystick mode) or an exact location on the screen (direct map-ping mode). Even though the way it works is pretty simple in theory, the system has to deal with several problems, such as optical noise from different sources, how people move their heads, accuracy, and power consumption [19].

In addition to that, a multi-modal human-computer interaction system was created and built from the ground up. EAR and MAR both include numerous modalities by opening both mouths and feature track eyes that move in parallel with one another. You can scroll up and down with the mouse. And mouse right-clicks, as well as other HCI multi-modal system features, can now be successfully implemented with the help of the proposed algorithm. In the future, it will be possible to extend the capabilities of existing systems so that certain actions and the functions that correspond to them can be configured.

The HCI methods can also control other computer interface devices. If the frame contains multiple faces, the model can be enhanced by considering only facial movements as shown in Figure 6 [20].

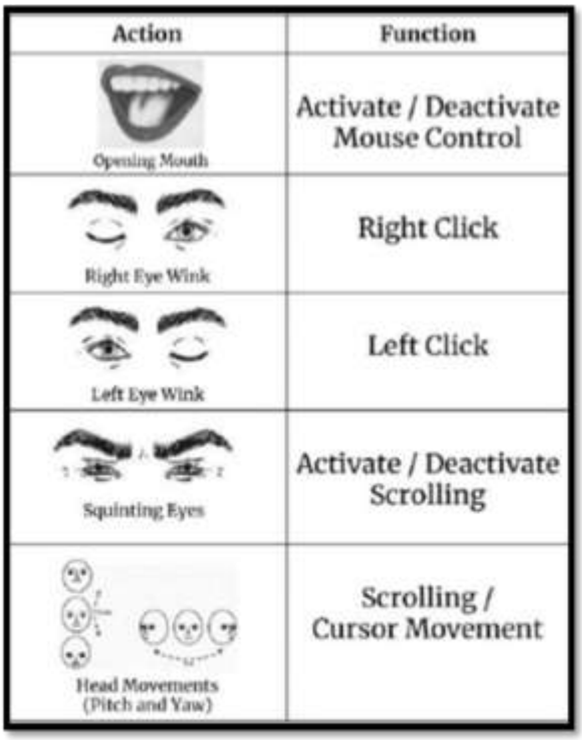

Figure 6: Usage

III. CONCLUSION

When we considered the difficulties faced by people without hands, we discovered that in addition to being unable to use computers efficiently, these individuals also experience feelings of social inferiority and isolation within society. One of the things we discovered when we considered the challenges faced by people without hands. People who have disabilities need support from other people for the rest of their lives. They are unable to provide for themselves in any capacity, including being able to generate money for themselves. Designing effective devices for disabled individuals has become essential, and it can now be done rapidly using machine learning and deep learning techniques. Consequently, it is essential to become an expert in designing effective devices for them.

The vast majority of the published works observed that while there was no need for human intervention in any of the activities, the process was substantially slower than it would have been otherwise. In the future, the researchers aim to increase the speed while maintaining the same level of precision in their work. They also conduct research on identifying other categories of human activities within the proposed system's framework. The researchers removed the major limitation of the eye tracking component by deciding to rely only on the head tracking component. Some problems associated with eye tracking included inaccurate readings, sluggish eye movements, and requiring the head or a finger to be used as a pointing device.

Persons with other forms of impairments, such as physical impairments, can also utilize the head tracking device. People who have disabilities are not the only people who can use this equipment. Victims of the disaster who have suffered amputations as a direct consequence of their injuries.

In addition, the eco-friendliness of these devices further demonstrates that users can interface with computers without actually touching the screens of the devices. It is common knowledge that the mouse, keyboard, and computer monitors all have the ability to act as vectors in the transmission of disease. Therefore, preventing the spread of diseases requires not touching other people.

IV. FUTURE WORK

In the future, the primary goal is to increase the accuracy as much as is practically possible and to make it accessible to the general public as quickly as possible so that it can help the masses. In addition, one of the primary goals is to differentiate between the natural blinking of the eyes and the intentional blinking of the subject. In addition, making it compatible with various operating systems, including Android, is on the list of things to do.

V. REFERENCE

[1] C. -W. Wu, H. -Z. Yang, Y. -A. Chen, B. Ensa, Y. Ren and Y. -C. Tseng, "Applying machine learning to head gesture recognition using wearables," 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), Taichung, Taiwan, 2017, pp. 436-440, doi: 10.1109/ICAwST.2017.8256495.

[2] A. Ronzhin and A. Karpov, "Assistive multimodal system based on speech recognition and head tracking," 2005 13th European Signal Processing Conference, Antalya, Turkey, 2005, pp. 1-4.

[3] A. Al-Rahayfeh and M. Faezipour, "Eye Tracking and Head Movement Detection: A State-of-Art Survey," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 1, pp. 2100212 -2100212 , 2013, Art no. 2100212, doi: 10.1109/JTEHM.2013.2289879.

[4] A. Arab, M. Fakoor, E. Arbabi and R. Daneshmand, "A low cost hands-free typing system for disabled people," 2012 19th Iranian Conference of Biomedical Engineering (ICBME), Tehran, Iran, 2012, pp. 75-78, doi: 10.1109/ICBME.2012.6519661.

[5] Eduardo Alcaraz-Mateosa, Iva Turicb, Andrés Nieto-Olivaresa,Miguel Pérez-Ramosa, Enrique Popletc, “Head-tracking as an interface device for image control in digital pathology: a comparative study” Volume 53, Issue 4, 2020 in Patologia ,Pages 213-217, ISSN 1699-8855, doi: https://doi.org/10.1016/j.patol.2020.05.007.

[6] M. N. Sahadat, A. Alreja and M. Ghovanloo, "Simultaneous Multimodal PC Access for People With

Disabilities by Integrating Head Tracking, Speech Recognition, and Tongue Motion," in IEEE Transactions on Biomedical Circuits and Systems, vol. 12, no. 1, pp. 192-201, Feb. 2018, doi: 10.1109/TBCAS.2017.2771235.

[7] M.-L.Machado,J.-Y.Guincestre, M.Aimé, M.-H.Drouet, C.Giry, F.LeDoze, G.Lamy Rousseau, P.Denise, S.Besnard, “Head Pilot: A new webcam-based Head Tracking System tested in permanently disabled patients” in IRBM, Volume 34, Issue 2,2013, Pages 124-130, ISSN 1959-0318, doi: https://doi.org/10.1016/j.irbm.2013.02.001.

[8] Shivanand P. Guness, Farzin Deravi, Konstantinos Sirlantzis, Matthew G. Pepper & Mohammed Sakel, “Development and Evaluation of a Mouse Emulator Using Multi-modal Real-Time Head Tracking Systems with Combining head tracking and mouse input for a GUI on Facial Gesture Recognition as a Switching Mechanism multiple monitors” in International Conference on Pattern Recognition, 2021, pp 218–232, doi:10.1007/978-3-030-68790-8_18.

[9] B. Coffen and M. S. Mahmud, "TinyDL: Edge Computing and Deep Learning Based Real-time Hand Gesture Recognition Using Wearable Sensor," 2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM), Shenzhen, China, 2021, pp. 1-6, doi: 10.1109/HEALTHCOM49281.2021.9399005.

[10] Yeon Gu Kang, Phill Kyu Rhee ,“Head Gesture Recognition Using Feature Interpolation” 2006, in International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, pp. 582-589, doi: 10.1007/11892960_70.

[11] Walter Vanzella, Natalia Grion, Daniele Bertolini, Andrea Perissinotto, Marco Gigante, and Davide Zoccolan, “A passive, camera-based head-tracking system for real-time, three-dimensional estimation of head position and orientation in rodents” 2019, in Journal of Neurophysiology , vol. 122, doi: 10.1152/jn.00301.2019.

[12] Roy S. Hessels, Diederick C. Niehorster, Gijs A. Holleman, Jeroen S. Benjamins, Ignace T. C. Hooge, “Wearable Technology for “Real-World Research”: Realistic or Not” 2020, in SAGE Public Health Emergency Collection, pp: 611–615, doi: 10.1177/0301006620928324.

[13] K. Sancheti, S. K. K, S. A and S. P, "Hands-free Cursor Control using Intuitive Head Movements and Cheek Muscle Twitches," TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 0356-0361, doi: 10.1109/TENCON.2018.8650532.

[14] L. . -P. Morency, A. Rahimi, N. Checka and T. Darrell, "Fast stereo-based head tracking for interactive environments," 2002, Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition, Washington, DC, USA, pp. 390-395, doi: 10.1109/AFGR.2002.1004185.

[15] Jilin Tu, T. Huang and Hai Tao, "Face as mouse through visual face tracking," 2005, The 2nd Canadian Conference on Computer and Robot Vision (CRV'05), Victoria, BC, Canada, pp. 339-346, doi: 10.1109/CRV.2005.39.

[16] Y. Song, Y. Luo and J. Lin, "Detection of Movements of Head and Mouth to Provide Computer Access for Disabled," 2011 International Conference on Technologies and Applications of Artificial Intelligence, Chung Li, Taiwan, pp. 223-226, doi: 10.1109/TAAI.2011.46.

[17] Y. Fu and T. S. Huang, "hMouse: Head Tracking Driven Virtual Computer Mouse," 2007 IEEE Workshop on Applications of Computer Vision (WACV '07), 2007, pp. 30-30, doi: 10.1109/WACV.2007.29.

[18] C. M. Asad et al., "Removing Disabilities: Controlling Personal Computer Through Head Movements and Voice Command," 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), Almaty, Kazakhstan, 2018, pp. 1-4, doi: 10.1109/ICAICT.2018.8747123.

[19] M. N. Islam et al., "Developing a Novel Hands-Free Interaction Technique Based on Nose and Teeth Movements for Using Mobile Devices," 2021, in IEEE Access, vol. 9, pp. 58127-58141, doi: 10.1109/ACCESS.2021.3072195.

[20] Y. Song, Y. Luo and J. Lin, "Detection of Movements of Head and Mouth to Provide Computer Access for Disabled," 2011 International Conference on Technologies and Applications of Artificial Intelligence, Chung Li, Taiwan, pp. 223-226, doi: 10.1109/TAAI.2011.46.

[21] https://dextrowaredevices.com/