International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 4, July-Aug-2021, ISSN (Online): 2395-566X

Gesture Recognition for User Interaction

Srija Nagabhyru, Anshul Joshi
Dept of Information Technology,
SRMIST
SRM University, Chennai, India
nn1354@srmist.edu.in, aa1531@srmist.edu.in

***Abstract*- With the massive influx and advancement of technologies there is a scope for us to interact with our systems in the best possible way. One such technology would be Gesture based Human Computer Interaction. So our system makes use of HCI which would help us interact without touching the screen. It is a well known fact that two dimensional user interfaces are everywhere, but with the increasing popularity of Extended Reality(XR) we require a better, more sophisticated three dimensional user interface.**

***Keywords*- Human Computer Interaction, Gesture recognition, Dynamic environment, Event handling.**

I. INTRODUCTION

Using gestures to control user interfaces could enable interaction in situations where hardware controls are missing, for instance, wall-sized displays and could support disabled people to interact with a computer where other controls fail. Different to common hardware supported interaction controls like mouse and keyboard setups or joysticks for instance, gesture interaction does not suffer from a predefined and limited command setup.

The amount of possible gestures is only limited by the users' creativity. Gesture recognition has been practiced for a long time driven by e.g. software that detects coloured gloves or even the bare hands using video cameras. Since we are interested in comparing different gesture based interface navigation controls, we are confronted with the problem of quickly implementing different variants of interactions with the same application.

Gesture-based interaction offers a rich set of possibilities by combining hand movements and postures to control an interface. In the existing system, we do not have any other way except using Touch screens, Physical buttons or voice commands to interact with the kiosks/laptops/desktops. There are many applications like media player, MS-Office, Windows picture manager etc which require a natural and intuitive interface. Nowadays most of the users use keyboards, mouse, pen, joysticks to interact with computers.With the increasing fear of covid and potentially other diseases, we are afraid to interact with public technologies like kiosks in malls and fast food chains.

II. LITERATURE SURVEY

Kinect: is a line of motion sensing input devices produced by Microsoft. It is a combination of hardware and software contained within the Kinect sensor accessory that can be added to any existing Xbox 360. Aside from its gaming benefits, some of the Kinect's other key features have made it useful in the health industry. For example, it has a voice recognition system, which allows users to search for things on the Internet or movies through Bing—Microsoft's client search engine—by speaking commands. The console also has a facial recognition system that uses a person's physical features for security purposes. But Kinect has privacy and security issues as the system files our resources automatically outside of the hard drive and onto a cloud. Additionally, the system can easily be hacked into as well.

Orbecc: Orbbec Astra is a device that brings depth sensing to the connected computer. Orbbec prepared a body-tracking SDK for Astra with Unity support. Orbbec Persee is a standalone sensor with an integrated Operating System. Orbbec Persee already includes a body-tracking SDK. All of the heavy-lifting is done on the device itself. But, chances are we will need to rewrite most of our code, though.

Project Soli: Project Soli, driven by Google's Advanced Technology and Projects (ATAP) team, was first showcased back in 2015. The idea is that a radar chip can be used to detect hand movements and gestures to interpret what they could mean. Soli is a dedicated radar chip to collect raw data of hand gestures and then interpret them correctly for the right commands. Google says the miniature radar understands human motions at various scales, from the tap of a finger to the movements of the body.

It is always sensing for movement while maintaining a low footprint — keep in mind Soli is not a camera and doesn't capture any visual images. Soli relies on a custom-built machine learning (ML) model to understand a large range of possible movements. The Soli radar chip emits electromagnetic waves in a broad beam and when a human hand interacts with this, some of these waves are reflected back to the antenna. The ML-model quickly interprets the properties of the reflected signal to carry out the required command.

III. PROPOSED SYSTEM

A gesture-based user interface and interaction system that includes media capturing devices, a processor, and a display device. The media-capturing device captures media associated with a user and their surrounding environment. An advanced way to get a more sophisticated three dimensional user interface without a need of any extra peripherals. Which works across different platforms.

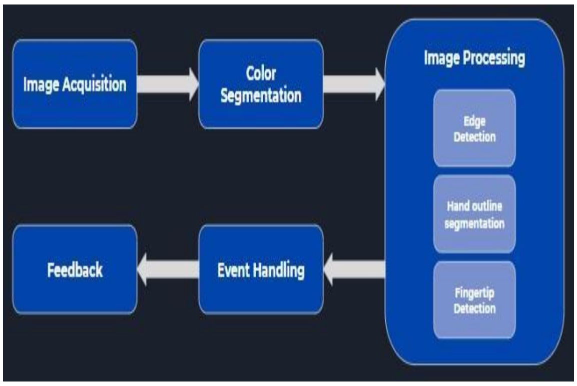

Fig 1. Proposed System.

IV. MATHEMATICAL MODELLING OF SUBMODULES

Gesture recognition has been a very interesting problem in the Computer Vision community for a long time. This is particularly due to the fact that segmentation of foreground objects from a cluttered background is a challenging problem in real-time. The most obvious reason is because of the semantic gap involved when a human looks at an image and a computer looking at the same image. Humans can easily figure out what’s in an image but for a computer, images are just 3-dimensional matrices.

It is because of this, computer vision problems remain a challenge. We are going to recognize hand gestures from a video sequence. To recognize these gestures from a live video sequence, we first need to take out the hand region alone, removing all the unwanted portions in the video sequence. After segmenting the hand region, we then count the fingers shown in the video sequence to instruct a robot based on the finger count.

1. Segment the Hand region:

The first step in hand gesture recognition is obviously to find the hand region by eliminating all the other unwanted portions in the video sequence. First, we need an efficient method to separate foreground from background. To do this, we use the concept of running averages. We make our system look over a particular scene for 30 frames. During this period, we compute the running average over the current frame and the previous frames. By doing this, we essentially tell our system that the video sequence that you got as input (running average of those 30 frames) is the background.

A robust background subtraction algorithm should be able to handle lighting changes, repetitive motions from clutter and long-term scene changes. The following analyses make use of the function of V(x,y,t) as a video sequence where t is the time dimension, x and y are the pixel location variables. e.g. V(1,2,3) is the pixel intensity at (1,2) pixel location of the image at t = 3 in the video sequence.

For calculating the image containing only the background, a series of preceding images are averaged. For calculating the background image at the instant t, where N is the number of preceding images taken for averaging. This averaging refers to averaging corresponding pixels in the given images. N would depend on the video speed (number of images per second in the video) and the amount of movement in the video.

After calculating the background B(x,y,t) we can then subtract it from the image V(x,y,t) t time t = t and threshold it. Thus the foreground is where Th is threshold. Similarly we can also use median instead of mean in the above calculation of B(x,y,t). Usage of global and time-independent thresholds (same Th value for all pixels in the image) may limit the accuracy of the above two approaches.

2. Motion Detection and Thresholding:

To detect the hand region from this difference image, we need to threshold the difference image, so that only our hand region becomes visible and all the other unwanted regions are painted as black. This is what Motion Detection is all about. Thresholding is the assignment of pixel intensities to 0's and 1's based on a particular threshold level so that our object of interest alone is captured from an image.

This is the simplest segmentation method. It separates out regions in an image corresponding to objects of interest that need to be analyzed. Based on the variation of intensity between the object pixels and the background pixels, separation is done. To differentiate the pixels of interest from the rest, a comparison of each pixel’s intensity value is carried out with a predetermined threshold (which can be set according to the area to be surveillanced). If the intensity value of a given pixel is greater than the predetermined threshold then it’s intensity value is set to 255 and if it is less than predetermined threshold then it is set to 0.

3. Gesture Recognition:
There are 4 steps involved to do gesture recognition, each of them having a different mathematical approach.

3.1 Contour Generation: The outline or the boundary of the object of interest. A contour is defined as the line that joins all the points along the boundary of an image that have the same intensity. cv2. findCont ours() function in OpenCV helps us find contours in a binary image. If you pass a binary image to this function, it returns a list of all the contours in the image. Each of the elements in this list is a numpy array that represents the (x, y) coordinate of the boundary points of the contour (or the object).

3.2 Bitwise-AND: A bitwise AND is a binary operation that takes two equal-length binary representations and performs the logical AND operation on each pair of the corresponding bits, which is equivalent to multiplying them. Thus, if both bits in the compared position are 1, the bit in the resulting binary representation is 1 (1 × 1 = 1); otherwise, the result is 0 (1 × 0 = 0 and 0 × 0 = 0).

3.3 Pairwise Euclidean Distance: This is the distance between two points. the euclidean distance between a pair of row vector x and y is computed as:

dist(x, y) = sqrt(dot(x, x) - 2 * dot(x, y) + dot(y, y))

3.4 Convex Hull: Convex hull is a dynamic, stretchable envelope that wraps around the object of interest.

4.Event Handling and Feedback:
Once the gesture is identified the appropriate command for it will be executed. These commands will call the events for controlling the Home appliances Or Machine Operations. Feedback would be given by the system based upon some predefined gesture.

5. Advantages:

- The current system gives best results only in plain black background. This can be overcome through our proposed technique.
- No external camera setup is required for this model to work. The internal camera of a laptop can do the gesture recognition properly.
- Through the use of gesture recognition, remote control with the wave of a hand of various devices is possible.

V. EXPECTED OUTPUT

As the expected output of the program, the gesture which is made in front of the camera will be displayed. Like in this case, 1 is shown in the image, so the output will come as “This is - 1”. Further, corresponding actions assigned to that gesture will be performed Ex.- 1 can be for playing a video, 5 can be for pausing a video etc.

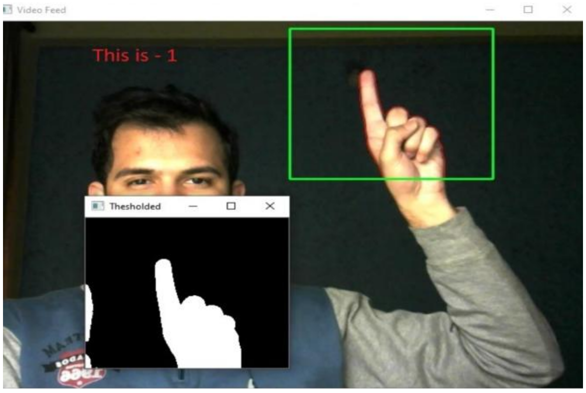

Fig.2 Expected output This is - 1.

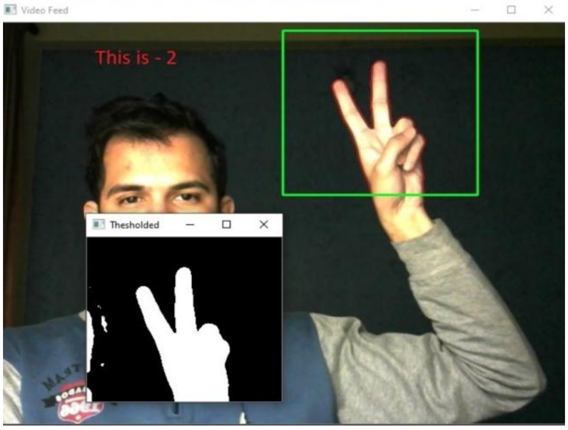

Fig.3 Expected output This is -2.

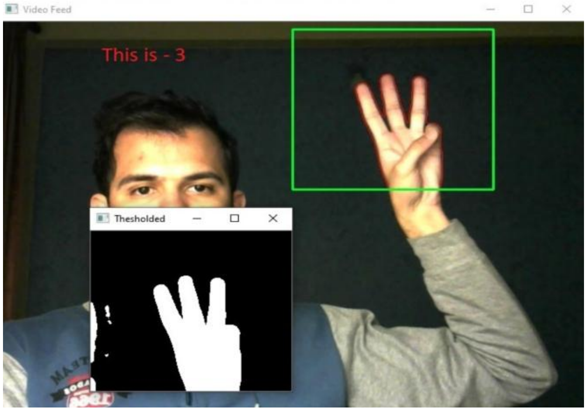

Fig 4. Text Expected output This is - 3.

VI. CONCLUSION

This gesture-based user interface and interaction system includes media capturing devices, a processor, and a display device. The media-capturing device captures media associated with a user and their surrounding environment. An advanced way to get a more sophisticated three dimensional user interface without a need of any extra peripherals. Which works across different platforms. Also this system will overcome various limitations faced by the existing systems.

VII. ACKNOWLEDGMENT

This work was supported in part by the SRM Institute of Science and Technology.

REFERENCES

[1] Shah, Syed & Ahmed, Ali & Mahmood, Iftekhar & Khurshid, Khurram. (2011). Hand gesture based user interface for computer using a camera and projector. 2011 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2011. 168-173. 10.1109/ICSIPA.2011.6144111.

[2] T. Singh," Finger Mouse," COMS W4735 Project Report, 2007.

[3] L. Jin et al." A novel Vision based Finger-Writing Character Recognition System," in Proc.18th International Conference on Pattern Recognition, 2006, pp. 1104-1107.

[4] Y. Liu, X. Liu and Y. Jia," Hand-Gesture Text Input for Wearable Computers," in Proc.4th IEEE International Conference on Computer vision System, 2006, pp. 8-8.

[5] Sanghi et al." A fingertip detection and tracking system as a virtual mouse, a signature input device and an application selector," in Proc. Southeastcon IEEE, 2008, pp. 503-506.

[6] Wikipedia, Gesture Recognition, http://en.wikipedia.org/wiki/Gesture_recognition (accessed on March 2013).

[7] J. O. Kim, C. Park, J. S. Jeong, N. Baek and K. H. Yoo, "A Gesture Based Camera Controlling Method in the 3D Virtual Space", International Journal of Smart Home, vol. 6, no. 4, (2012), pp. 117-126.

[8] J. Wachs, M. Kolsch, H. Stern and Y. Edan, "Vision-Based Hand-Gesture Applications", Communication of the ACM, vol. 54, no. 2, (2011), pp. 60-71.

[9] Kim, J.O., Kim, M. and Yoo, K.H., 2013. Real-time hand gesture-based interaction with objects in 3D virtual environments. International Journal of Multimedia and Ubiquitous Engineering, 8(6), pp.339-348.

[10] An, J.H. and Hong, K.S., 2011, January. Finger gesture-based mobile user interface using a rear-facing camera. In 2011 IEEE international conference on consumer electronics (ICCE) (pp. 303-304). IEEE.

[11] Z. Zhang, "Microsoft Kinect Sensor and Its Effect," in IEEE MultiMedia, vol. 19, no. 2, pp. 4-10, Feb. 2012, doi: 10.1109/MMUL.2012.24.

[12] Baldwin, R. "Google's Project Soli to bring gesture control to wearables." Engadget. com (2015).

[13] A.D. C. A. Coroiu and A. Coroiu, "Interchangeability of Kinect and Orbbec Sensors for Gesture Recognition," 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, 2018, pp. 309-315, doi: 10.1109/ICCP.2018.8516586.

[14] Giancola S., Valenti M., Sala R. (2018) Metrological Qualification of the Orbbec Astra S™ Structured-Light Camera. In: A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies. SpringerBriefs in Computer Science. Springer, Cham. https://doi.org/10.1007/978-3-319-91761-0_5.

[15] Bradski, Gary, and Adrian Kaehler. Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.", 2008.

[16] Laganière, Robert. OpenCV Computer Vision Application Programming Cookbook Second Edition. Packt Publishing Ltd, 2014.

[17] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). Human-computer interaction. Addison-Wesley Longman Ltd..

[18] Vision Based Hand Gesture Recognition Pragati Garg, Naveen Aggarwal and Sanjeev Sofat

[19] Ying Wu, Thomas S Huang, " Vision based Gesture Recognition : A Review", Lecture Notes In Computer Science; Vol. 1739 , Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction, 1999.

[20] A. Mulder, "Hand gestures for HCI", Technical Report 96-1, vol. Simon Fraster University, 1996

[21] Richard Watson, "A Survey of Gesture Recognition Techniques", Technical Report TCD-CS-93-11, Department of Computer Science, Trinity College Dublin, 1993.

[22] A. J. Heap and D. C. Hogg, "Towards 3-D hand tracking using a deformable model", In 2nd International Face and Gesture Recognition Conference, pages 140–145, Killington, USA,Oct. 1996.

[23] Y. Wu, L. J. Y., and T. S. Huang. "Capturing natural hand Articulation". In Proc. 8th Int. Conf. on Computer Vision, volume II, pages 426–432, Vancouver, Canada, July 2001.