A Review on Hand Gesture Recognition

Ramanuja M A[1], Pradyumna Ramesh[2], Shikha Yadav[3], Rohan B J[4], Yashpal Gupta S[5]

Department of Information Science and Engineering Vidyavardhaka College of Engineering, Mysuru, India

Abstract: User Interface (UI) and Human Computer Interaction (HCI) have come a long way from the time computers went personal. The constant work and research done to explore HCI allows for new ways and improvements. Hand gesture recognition is part of HCI that quickly picked up after the advancements in computer hardware and machine learning, particularly computer vision. Hand gesture recognition is a natural alternative to interact with computer as compared to mechanical devices(mouse, keyboard, etc),just as we interact with other humans through hand gestures. We review the existing tools and techniques that make hand gesture possible today, along with some common pitfalls and potential fixes.
Keywords: Hand Gesture Recognition, Human Computer Interaction (HCI), Depth Imaging, ComputerVision.

I. INDRODUCTION

HCI or Human Computer Interaction is one of the core considerations when we design systems for pleasant human experience. HCI and UI are tightly coupled today, add to that the recent advances in machine learning and we just cannot ignore the opportunities and importance that HCI provides. There are many tools and methodologies that existing HCI models use such as data gloves to obtain data and HMM(hidden markovnikov model) to make sense of the obtained data.

We can think of data collection, prediction and testing that form the core of any HCI model. Existing HCI models collect data in different ways for example, some detect through the finger positions, some use depth imaging to understand the gesture and some use additional hardware like data gloves. All these techniques are reviewed and compared later in the paper and a potentially better alternative is discussed.

Prediction algorithms also prove to be critical in the performance of the HCI model used. There are two major paradigms used for prediction,1.Rule based approach,2.machine learning based approach. We see that rule based approach is ideal for cases where the classification rules are simple, i.e, the number of classes for the prediction to fall in are minimal in number. But humans have varied ways of gestures and the possibilities are large enough that, in recent times we have seen a shift of prediction towards machine learning based approaches, Hidden Markov Model(HMM) being popular. We review these models and their efficiency later in the paper.

Testing is a way to get the HCI model to perform accurately and continuously improve in case of machine learning based approach Later in the paper we review how certain models require high quality data for testing which is not always possible to obtain.

In this paper, we review the existing tools and techniques surrounding HCI and discussalternative methodologies.

II. LITERATURE REVIEW

A. Hand Gesture Data Collection

Hand Gesture Data Collection is the process in which data about the gesture can be obtained.

In [1] G.R.S Murthy, et al. talks about two principle approaches commonly used to interpret human gestures.1.Methods which use data gloves and 2.Methods which are vision based. Data gloves or wired gloves are according to wikipedia,"A wired glove (also called a dataglove [Oxford English Dictionary] or cyberglove) is an input device for human–computer interaction worn like a glove. Various sensor technologies are used to capture physical data such as bending of fingers. Often a motion tracker, such as a magnetic tracking device or inertial tracking device, is attached to capture the global position/rotation data of the glove". Data gloves introduce problems such as restricted finger movement, sensor accuracy, etc. Vision based approach ,the authors mention ,are natural and allow us to interact with the system using only our hands. However the authors conclude that it is difficult todesign vision based interface for generic usage. In [2] Yanmin Zhu,et al. the authors discuss about the detection and segmentation on the basis of skin color, shape and background subtraction. Skin colors, the authors elaborate, can help in detecting the gesture because the color of our hands and palms usually vary. This difference gives us a hint aboutthe orientation of the hand gesture. The authors go on to explain how the contour of hands can be obtained by applying edge detection operators such as roberts and laplace operator. According to the authors, background subtraction can significantly reduce the input size of the data and allow for faster and efficient computation. However, optimum distance from the computer and ample lighting are necessary for the above approach.

In [3] Fakhreddine Karray, et al. the authors mention the three ways in which data can be obtained from the gesture.1.Visual driven,2.Audio driven,3.Sensor driven. Data gloves are primarily sensor driven and introduce problems like restricted finger movement. Audio driven gestures are popular with AI assistants where you need an informational reply. Visual driven gestures are still the best suited to enter input into the computer system.

In [8] Jesus Suarez,et al. the authors talk about using depth imaging for detecting hand gesture recognition. Depth cameras are used to achieve these tasks, Microsoft Kinect and ASUS Xtion being some of them. In this paper, the authors answer questions like "what methods are being used to achieve hand localization and gesture recognition with depth cameras?". Hand localization, which is a process in which we detect hands in an image, is also discussed. The authors split hand localization into two steps,1.Hand segmentation and 2.Hand tracking. The paper also explains about the advantages of depth cameras over color cameras. However, ample lighting and heavy compute power is still an issue.

The above citings imply that optimum distance from the computer and ample lighting are some of the persistent issues that prevail across all these different HCI models. Our model, which captures data through an accelerometer mounted on an arduino, tackles these issues effectively.

B. Prediction Algorithms

In [1] G.R.S Murthy, et al. the authors classify the prediction algorithms into two categories,1.Rule based approaches and 2.Machine learning based approaches. Rule based approaches are a set of manually encoded rules written for feature inputs. The authors rightly point out that the major problem with rule based approaches is that they rely on the ability of a human to encode rules. HMM or Hidden Markov Models is a type of machine learning based approach which treats a gesture as an output of a stochastic process. The is the most popular model in this category. However, its computing is questionable at times.

In [2] Yanmin Zhu,et al. the authors provide data about some experiments performed using the HMM. The training set was composed of 60 different gestures performed by 20 different people, amounting to 1200 images. With stationary background and normal lighting, the system performed at an accuracy of 90 percent. The input given to the HMM was a combination of spatial and temporal feature vectors. However, optimum lighting is necessary in the above algorithm.

In [4] Mokthar M. Hassan,et al the authors have applied an appearance based model for hand detection and recognition. The authors have used the grayscale color model that converts from the input color images. Local computing of brightness was used as the recognition system. This model is not compute-heavy but because brightness is an important factor in recognition, lighting becomes an issue. It also introduces the problem of optimal distance from the computer.

In [7] Simei G. Wysocki,et al. the authors study about the multilayer perceptron and its role in predicting the gesture. The authors point out that the above algorithm qualifies three main requirements,1.Computational time,2.Memory requirement,3.Classification accuracy. However, trainingthese models is cumbersome as require huge data and are difficult to build. They also require quality test case data. Later in the paper, we discuss how our model allows the use of SVM algorithm and thus simplifies the process of prediction.In [8] Jesus Suarez,et al. the authors talk about the availability of Kinect and available libraries and their impact on gesture recognition. Wikipedia describes Kinect as "Kinect is a line of motion sensing input devices produced by Microsoft and first released in 2010. The devices generally contain RGB cameras, and infrared projectors and detectors that map depth through either structured light or time of flight calculations, which can in turn be used to perform real-time gesture recognition and body skeletal detection, among other capabilities. They also contain microphones that can be used for speech recognition and voice control. "The authors focus on depth cameras and their ability to capture more as compared to normal image recognition. However, this method requires high compute power and ample lighting. In [9] Ying Wu,et al. the authors classify gestures in terms of what they convey. For eg, communicative gestures, conversational gestures and controlling gestures etc. These act as additional data points for the prediction of the gestures and help increase accuracy. However, it is reallydifficult to capture them precisely.

In [10] H.Lu,et al. the authors describe a four step framework for gesture recognition which consists of hand detection, hand tracking, hand segmentation and gesture recognition. The only disadvantage of this framework is that it is a long process and can be an overkill for simple gesture recognition.

C. Testing

In [5] Xingyan Li,et al. the authors describe about the testing done on the FCM algorithm(Fuzzy C-Means).The authors reveal that FCM can be reliable and fast but also conclude that good quality training data is a hard requirement in this case. This makes the FCM little impractical as high quality training data is not always guaranteed.

In [12] S Ghosh,et al the authors talk about the SVM(Support Vector Machine) algorithm and its use cases. The authors point out that SVM is great when the number of classes to be classified is relatively small Proving that in this case, SVM algorithm is a great fit.

In [11] A. Singh,et al. the authors discuss about various supervised machine learning algorithms and their applications. SVM can be a nice choice in the case of hand gesture recognition as it is reliable and fast and requires relatively less data to test, with data quality allowed to be compromised to a certain extent.

Table-1 and Figure-1 depict the comparisons of our model versus the various working models surveyed in this paper.

Table I
Comparison With Several Surveyed Hand Gesture Recognition Systems

| Reference	Methodology	Gesture set	Device	Operating environment	Accuracy(%)	Speed(ms)
Licsa´r and Szira´nyi [16]	User adaptive recognition with interactive training	Palm with varying fingers	Projector and camera	Only hand allowed toappear	Over 98	NA
Kim et al.[17]	Active shape model for gait recognition	Human gaits	Infra-red camera	Works across illumination changes	Over 90	NA
Kao et al.[18]	Face and hand gesture recognition by PCA	Palm with varying fingers	Color camera	Elevator	Over 94	NA
Xie[19]	Fuzzy neural network for mode classification	Four hand postures	Stereo vision	Hand profile in meeting room	NA	Real time
Stergiopoulou and Papamarkos [20]	SGO neuralgas network	No. of raised fingers	RGB camera	Hands in simple background	90.45	NA
Van den Bergh and Van Gool [21]	Classifier using traditional 2D Haarlets	Six hand postures with varying fingers	ToF (depth) plus RGB	Allow other persons in background	99.54	33.4
Qing etal.[22]	Traditional 2D Haarlets classifier with SCFG	Four hand postures with varying fingers	RGB camera	Hands in simple background	95.65	3.04
Ciprian etal.[23]	Dynamic hand gesture using tensor voting filter	Three/four static/ dynamic hand gestures	RGB camera	Hands in simple background	Simulation only	3.93
ours	SVM Classifier	Any alphabetical gesture	Accelero-meter mountedon an Arduino	Hands in any degree of lighting and any distance from the system	Over 90	Real Time
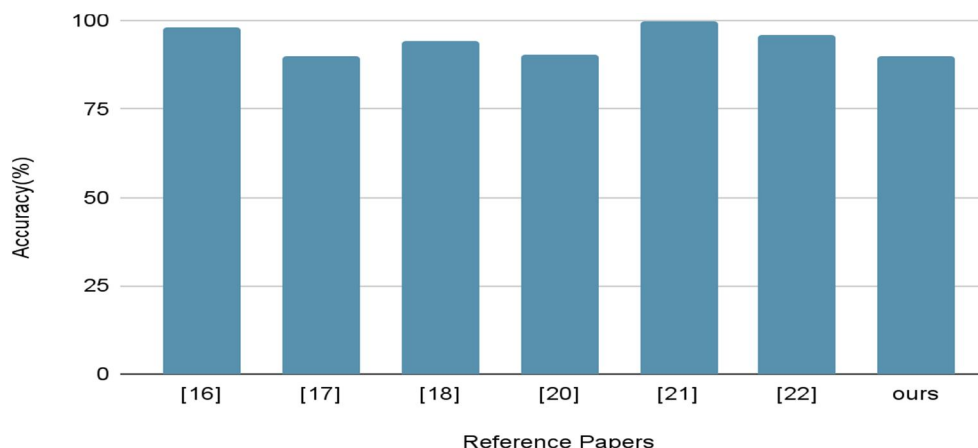
Fig.1 Bar chart comparison of accuracy(%) of the surveyed models

III.   CONCLUSION

The above citations discuss hand gesture recognition along the lines of data collection, gesture prediction and testing the model. The setbacks or shortcomings of the models above usually revolve around the issues of ample lighting, optimum distance from the computer, high memory usage, exceedingly long run time algorithms, requirements of high quality test data. Our model, since it uses an accelerometer mounted on an arduino, accurately captures gestures and feeds it to the computer through the wire, it does not require ample lighting and distance from the computer is no more an issue.Our model uses SVM algorithm, in accordance with the above citations, achieving accuracy upto 98 percent.

IV.   FUTURE WORK

In the future,we plan to make the connection wireless and pass the data over a network.We would also like to deeply couple the accelerometer and the arduino and embed them into a single device and make it wearable like a watch to further ease the usage of our device.Exploring and keeping track of the upcoming classification algorithms would also be helpful for our model.

REFERENCES

[1]   G. R. S. Murthy, R. S. Jadon. "A Review of Vision Based Hand Gestures Recognition," International Journal of Information Technology and Knowledge Management, vol. 2

[2]   Yanmin Zhu, Zhibo Yang, & Bo Yuan. (2013). Vision Based Hand Gesture Recognition. 2013 International Conference on Service Sciences (ICSS).

[3]   Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh, Mo Nours Arab, "HumanComputer Interaction: Overview on State of the Art", International Journal on Smart Sensing andIntelligent Systems.

[4]   Mokhtar M. Hasan, Pramoud K. Misra, "Brightness Factor Matching For Gesture Recognition System Using Scaled Normalization", International Journal of Computer Science & Information Technology.

[5]   Xingyan Li. "Gesture Recognition Based on Fuzzy C-Means Clustering Algorithm", Department of Computer Science. The University of Tennessee Knoxville.

[6]   S. Mitra, and T. Acharya. "Gesture Recognition: A Survey" IEEE Transactions on systems, Man and Cybernetics, Part C:

[7]   Applications and reviews, vol. 37

[8]   Simei G. Wysoski, Marcus V. Lamar, Susumu Kuroyanagi, Akira Iwata, "A Rotation Invariant Approach On Static-Gesture Recognition Using Boundary Histograms And Neural Networks,"IEEE Proceedings of the 9th International Conference on Neural Information Processing, Singapura.

[9]   Jesus Suarez,Robin R Murphy,"Hand gesture recognition with depth images:A review [9]Ying Wu and Thomas S. Huang,"Vision-Based Gesture Recognition-A Review

[10] Fang, K. Wang, J. Cheng and H. Lu, "A Real-Time Hand Gesture Recognition Method," 2007 IEEE International Conference on Multimedia and Expo, 2007, pp. 995-998, doi: 10.1109/ICME.2007.4284820.

[11] A. Singh, N. Thakur and A. Sharma, "A review of supervised machine learning algorithms," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 2016, pp. 1310-1315.

[12] S. Ghosh, A. Dasgupta and A. Swetapadma, "A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification," 2019 International Conference on Intelligent Sustainable Systems (ICISS), 2019, pp. 24-28, doi: 10.1109/ISS1.2019.8908018.

[13] D. L. Quam, "Gesture recognition with a DataGlove," IEEE Conference on Aerospace and Electronics, 1990, pp. 755-760 vol.2, doi: 10.1109/NAECON.1990.112862.

[14] Xia Liu and K. Fujimura, "Hand gesture recognition using depth data," Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings., 2004, pp. 529-534, doi: 10.1109/AFGR.2004.1301587.

[15] Yang, P. Premaratne and P. Vial, "Hand gesture recognition: An overview," 2013 5th IEEE International Conference on Broadband Network & Multimedia Technology, 2013, pp. 63-69, doi: 10.1109/ICBNMT.2013.6823916.

[16] Licsa´r, A., Szira´nyi, T.: User-adaptive hand gesture recognition system with interactive training. Image Vis. Comput. 23, 1102– 1114 (2005)

[17] Kim, D., Lee, S., Paik, J.: Active shape model-based gait recognition using infrared images. Intl J Signal Process Image Process Pattern Recogn 2, 1–13 (2009)

[18] Kao, Y.W., Gu, H.Z., Yuan, S.M.: Integration of face and hand gesture recognition. In: Proceedings of Third International Conference on Convergence and Hybrid Information Technology, vol. 1, pp. 330–335 (2008)

[19] Xie, W., Teoh, E.K., Venkateswarlu, R., Chen, X.: Hand as natural man–machine interface in smart environments. In: Proceedings of the 24th IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, pp. 117–122 (2006)

[20] Stergiopouloua, E., Papamarkos, N.: Hand gesture recognition using a neural network shape fitting technique. Eng. Appl.

[21] Artif. Intel. 22, 1141–1158 (2009)

[22] Bergh, M.V.D., Gool L.V.: Combining RGB and ToF cameras for real-time 3D hand gesture interaction. In: Proceedings of IEEE Workshop on Applications of Computer Vision, pp. 66–72 (2011)

[23] Qing, C., Georganas, N.D., Petriu, E.M.: Hand gesture recognition using Haar-like features and a stochastic context-free grammar. IEEE Trans Instrum Meas 57, 1562–1571 (2008)

[24] Ciprian, D., Vasile, G., Pekka, N., Veijo, K.: Dynamic hand gesture recognition for human-computer interactions. In: Proceedings of the 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, May 19–21, Romania, pp. 165–170 (2011)

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ◎ (24*7 Support on Whatsapp)