Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 26 (2023), No. 1, pp. 133–145
DOI : 10.47974/JSMS-953

MEMS based eating and drinking gesture spotting using machine learning techniques

Sivakannan Subramani *
Department of Advanced Computing
St. Joseph's University
Bangalore
India

Paavana Sathish †
Department of Advanced Optical Technologies
Friedrich Alexander University
Erlangen
Germany

Pushpalatha Subramani §
Department of Medical Electronics
MVJ College of Engineering
Bangalore
India

Abstract

An effective human-computer interaction for collaborative, collective and brilliant computing is provided by the Gesture recognition. Our study uses a single 3-axis accelerometer and 3-axis gyroscope for acquiring the data. Dynamic Time Warping (DTW), time and frequency domain and gesture discrepancy features are extracted from the raw data and these features are subjected to supervised machine learning algorithms such as KNN, SVM, RF and ANN. The system is evaluated depending on the data base of more than 10,000 traces obtained from five subsets. Through this study, we intend to identify eating and drinking gestures and more broadly short activities on real time environment. Many approaches for gesture recognition in specific settings have already been explored and studied using the combined interaction of several sensors. Despite these powerful hypotheses, gesture interpretation is still fragile and often depends on the individual's positioning relative to the cameras not with the sensors.

Subject Classification: *68T07.*
Keywords: *Gesture recognition, Dynamic time warping, Machine learning, Time and frequency domain features, MEMS.*

E-mail:* **sivakannan87@gmail.com (Corresponding Author)
† E-mail: **paavanasathish@gmail.com**
§ E-mail: **lpushpa199@gmail.com**

1. Introduction

Various unprompted continuous motions such as the head motions, body, hand, or the fingers are used to communicate particular facts in human-computer interactions. Therefore, these gestures or the motions can be considered as a completely habitual means of information conveying tool with several other features that is used in computer interaction. Till date, almost all of our computer system interactions are handled using conventional devices like keyboards, mouse's and other gadgets that are controlled with a remote mainly designed for immobile interaction. Utilizing the strong technical innovation, gesture-based interfaces can serve as an alternative method of manipulating systems, such as navigation in office software or playing some PC games [1].

Expressive and sensible body movements comprising physical mobility of the face, hands or the arms can be quite incredibly useful for comprehending realistic information, or for inter-relating with the surroundings. This includes a posture which means a stationary position without motion in any part of the body and a gesture a dynamic motion of any of the body parts. Multiple-to-one mappings generally exist from concepts to movements and vice versa. On the whole they can be Arm and hand gestures, Face and head gestures and Body gestures [2].

Wearing a bulky device which bears many cables which form the linkage between the device and the system is necessary in the instrumental based gesture recognition. This forms an obstacle to the user's contact with the machine in its ease and naturalness. Vision-based approaches will have to deal with trouble issues associated with blockage of parts of the individual's body when addressing this issue. Much of the literature's research on gesture recognition is focused on computer-vision techniques [3-7].

The very promising replacement to the vision-based recognition system is resorting to other recognizing technologies, such as techniques which deploy on acceleration or on electromyogram (EMG). Control of gestures which works on acceleration is best applicable to differentiate the apparent, intensive gestures with dissimilar hand movement paths. The size of discriminable hand gesture range is limited to 4-8 gestures due to certain inherent problems with EMG measurements including separability and reproducibility of measurements [8]. As a result, after examining all the sensing devices and techniques in literature, a 3-axis accelerometer is the sensing device utilized to acquire the data pertaining to gestures in this research.

Following the rapid development of MEMS technology, gesture spotting and recognition carried out based on the database obtained from a 3-axes accelerometer is a newly arising technique for gesture-based interaction. Accelerometers are incorporated in almost all the up-coming age group's personal thermionic gadgets such as smartphones [8-9] offers new interaction probabilities over an extended range of administration, for example, house-hold machines, workplace gadgets and virtual gaming techniques.

With reference to the complexes involved in the computation of the device and the exactness of the identification process, the developed device advances to give high competition to the previously designed gesture spotting and recognition devices and the other mathematical models.

2. Proposed Work

The first phase of data collection in designing the gesture recognition system includes constructing a data base of thousands of gestures on which the system can work. So, we collect M traces to create a database of the given gestures to train the system with N gestures. The data acquisition is carried out using BNO055 sensor. The gesture data trace is basically obtained due to the hand movement in terms of acceleration in the 3 directions, namely x, y and z axes. Later, the database is divided into training dataset and test data set. In our work, 80% of the data is considered to be the training data and 20% of it is for the testing procedure. The training dataset is also provided with the label class containing the variable 1, 2 and 3 indicating eating, drinking or null respectively. The data classification is carried out depending on the data readings obtained from the sensor BNO055.

After receiving the data from the sensor, it is given to microcontroller ESP32, from where data is sent to the system via Bluetooth module. The data from the sensor is transferred to the microcontroller ESP32 by programming it using the Arduino IDE with the ESP32 Arduino Core platform. The data from the ESP32 microcontroller is later sent to the system via the Bluetooth. The collected data is subjected to CAST (Crossings based Adaptive Segmentation Technique). CAST investigates with the varying lengths of gesticulations outstandingly and also adjusts to in coming new data by modifying the moving-averages approximately. CAST also avoids the undesired impact of the gyroscope long-term drift.
To classify the data obtained from the sensor, a predictive model is being developed which is priory trained with the dataset of eating, drinking and null (gestures apart from eating and drinking) gestures, using the BNO055

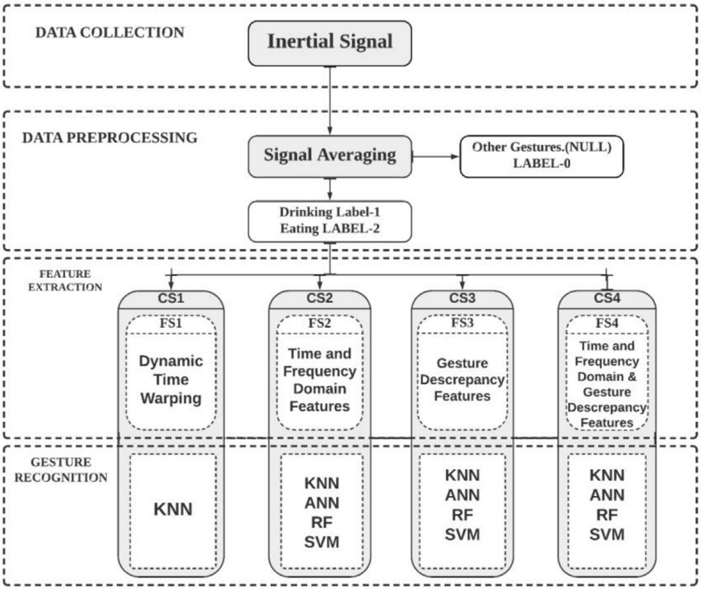

Figure 1

Block diagram representing computational solutions and the algorithms used

sensor data. The dataset is divided into training set and testing set in 4:1 ratio. The below figure shows the graph of the training dataset that was used to train the model. The complexity of the Gesture recognition occupies a wide area starting from the simpler gestures such as moving the hand either to right or to left or up and down to the complex gestures which represent an alphabet or a number.

The gesture data for the proposed system was obtained using a single 3-axis accelerometer at different time *t*. These hand gestures immensely suffer from variations in the time period and hence, traces of the same gesture with variation in lengths is the first major challenge in classifying the data and also for creating a novel system for gesture recognition.

In order to solve the above difficulty, we are using the machine learning algorithms that compute the similarity cost between each gesture and hence classify them either as eating, drinking or null gesture with high accuracy as shown in the Figure 1.

In this work, the features of the gestures are obtained by the four feature sets. Each feature set uses the appropriate machine learning algorithm for classification[10]. The four feature sets and their corresponding algorithms used are mentioned below in Table 1.

Table 1

Computational solutions with their corresponding algorithms

FEATURE SETS	ALGORITHMS
Dynamic Time Warping	KNN
Time and Frequency domain features	RF, ANN, SVM, KNN
Gesture Discrepancy	RF, ANN, SVM, KNN
Time and Frequency domain features & Gesture Discrepancy	RF, ANN, SVM, KNN

2.1 *Computational Solutions*

2.1.1 CS1 - Dynamic Time Warping

Dynamic Time Warping (DTW) is one of the algorithms used in time series analysis to calculate similitude between two temporal sequences, which may differ in length. Similarities in walking, for example, could be observed using DTW, even if one person moved faster than the other or if there were accelerations and decelerations during an observation. DTW has been applied to temporary video, audio and graphics data sequences — indeed, any data that can be transformed into a linear sequence is analyzed with DTW. The distance obtained from DTW is used to interpret gestures. To do this, a classification model for *K*-Nearest Neighbors (KNN) is used, by which, the hidden portions are allocated to the extremely usual classifier of their closest neighbors.

2.1.2 CS2 -Time and Frequency Domain Features

The time domain analysis shows the variations occurred in amplitude of the signal over a period of time. The frequency domain analysis indicates the amount of signal which lies within the specific frequency band. This function offers a professional summary of the data concerning a broad variety of characteristics of the signal. Those include central tendency measures, periodicity, dispersion, direction shifts, distribution of frequencies and region of magnitude. The spectrum of characteristics proposed was estimated over the tri-axial an x, antero-posterior $a\ y$ and vertical $a\ z$ acceleration. The proposed range of characteristics was measured on the medio-lateral a_x, antero-posterior a_y and vertical a_z acceleration corresponding to the tri-axial accelerometer readings, as well as on the yaw g_x, roll g_y and pitch g_z corresponding to the tri-axial gyroscope readings around the potential region. These functions are carried out using RF, ANN, SVM and KNN algorithm. The classification output is obtained by running each algorithm at a time.

2.1.3 CS3 - Gesture Discrepancy

Gesture discrepancy measure is a mean of a signal descriptor. It is the measures the minute differences between the two very similar activities. This feature set uses the Soft-DTW differentiable loss function to calculate the barycentre of the gesture for each of the gestures in the dissimilar proposed gestures. The computation of the prediction is carried out by using the quadratic complexities. In addition, the DTW distances calculated to each of the barycentre are utilized to construct a set of features. These are calculated by the RF, ANN, SVM and KNN algorithms.

2.1.4 CS4 - Time and Frequency Domain Features & Gesture Discrepancy

This is a combination of both the time and frequency features and the gesture discrepancy feature sets for the prediction analysis. This was performed to estimate if the integration of a gesture-discrepancy measure to well-established feature vectors i.e., the time and frequency domain gestures upgrades the system's identification rate.

3. Results & Discussions

The primary objective of the computerized simulation is to come up with techniques which give habitual, effective and workable tools for humans-computer interaction. These requirements can be met by giving the gestures as input modality. The efficient and powerful modes of interaction were the natural human gestures. Gestures can be of various types such as the static gesture (where the user assumes one particular position or configuration) or dynamic gesture, defined by motion. The process by which the user's gestures are made familiar to the novel system is called Gesture Recognition. In order to support the recognition of gestures, the position and movement of the body parts must be tracked and interpreted for the recognition of significant movement.

By bringing together all the multidisciplinary sensors various systems for recognition of gestures have been developed. This leads for the improvement of the system's performance and also increases the exactness of the recognition. Implementation of accelerometers on personal electronic gadgets such as the Smartphone's, wrist watches etc, has led to the advancement of the work of developing and presenting a novel gesture recognition system based solely on data from a single tri-axial accelerometer and a tri-axial gyroscope.

The system uses dynamic time warping, time and frequency domain features and the gesture discrepancy features to train the system efficiently. Dynamic time warping is used to estimate the similarity between gestures that were performed after conventional methods have failed due to different gesture trace durations. During the testing stage, the comparison of the unknown traces to the gesture which was produced by the generation of affinity to choose the suitable subset from the co-ordinate traces.

The sporadic feature of the gesture path is made used to project the subject's known gesture paths and the undetermined gesture paths back to the same subspace of lower dimension. Projection is carried out using four different KNN, ANN, SVM and RF classifiers. The analysis of the result is done based on the accuracy, precision, recall and F1 score. A confusion matrix is a table which is used on a set of test data for which the true values are known to describe the performance of a classification model [11].

3.1 *CS1 Result Analysis*

The below Figure 2 shows the feature analysis carried out for the dynamically warped features in x, y and z directions. This feature analysis determines the covariance, separability and instance distribution of features respectively.

As seen in the above Table 2, the results for the dynamically warped features subjected to the KNN algorithm produced the classification into

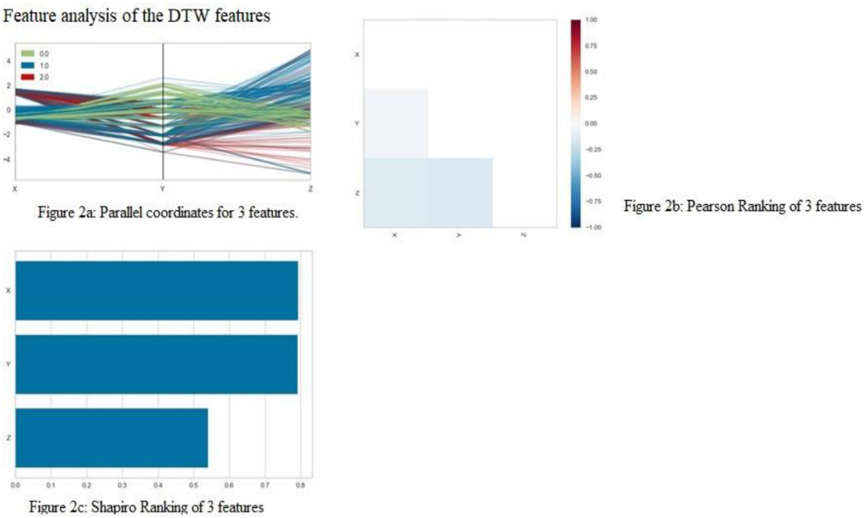

Figure 2a: Parallel coordinates for 3 features.

Figure 2b: Pearson Ranking of 3 features

Figure 2c: Shapiro Ranking of 3 features

Figure 2

Feature analysis of DTW features

Table 2

Results obtained in CS1

Computational Solution	Features	Algorithms	Accuracy	Precision	Recall	F1 Score
CS1	DTW	KNN	0.8215	0.8800	0.8251	0.8517

three classes respectively with 0.8215 accuracy, 0.8800 precision. The recall and F1 score are 0.8251 and 0.8517 respectively.

3.2 *CS2 Result Analysis*

The below Figure 3 shows the feature analysis carried out for the time domain and frequency domain features in t_x, t_y, t_z and f_x, f_y, f_z directions. This feature analysis determines the covariance, separability and instance distribution of features respectively. PCA and RadViz do the multivariate data analysis.

The Table 3 talks about the results obtained when the time and frequency domain features were subjected to different machine learning algorithms respectively. The ANN algorithm gives the best result with higher accuracy, precision and recall.

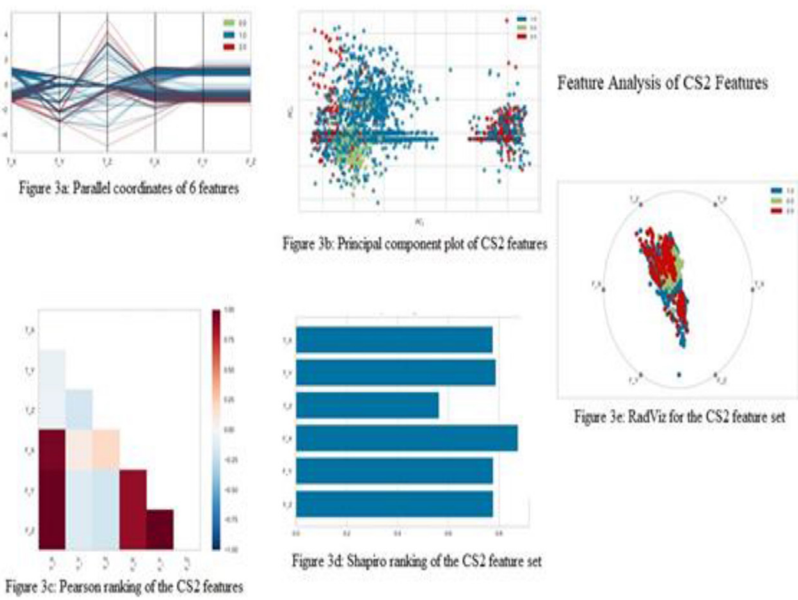

Figure 3

Feature analysis of CS2 feature set

Table 3

Results obtained in CS2 features set for respective algorithms

Computational Solution	Features	Algorithms	Accuracy	Precision	Recall	F1 Score
CS2	Time & Freq Domain Features	KNN	0.8350	0.8361	0.8916	0.8629
		SVM	0.8424	0.8571	0.8868	0.8717
		RF	0.8508	0.8606	0.8966	0.8783
		ANN	0.8647	0.803	0.8966	0.8884

3.3 *CS3 Result Analysis*

The below Figure 4 shows the feature analysis carried out for the gesture discrepancy features in x, y, z directions. This feature analysis determines the covariance, separability and instance distribution of features respectively. PCA and RadViz do the multivariate data analysis.

The Table 4 talks about the results obtained when the gesture discrepancy features were subjected to different machine learning algorithms respectively. The ANN algorithm gives the best result with higher accuracy, precision and recall.

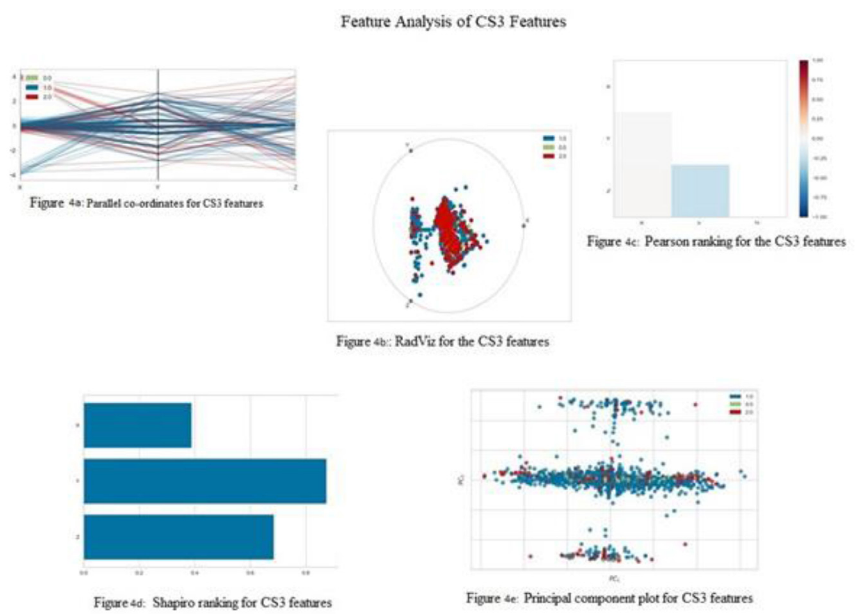

Figure 4

Feature analysis of CS3 feature set

Table 4

Results obtained in CS3 features set for respective algorithms

Computational Solution	Features	Algorithms	Accuracy	Precision	Recall	F1 Score
CS3	Gesture Discrepancy Features	KNN	0.8271	0.8311	0.8828	0.8561
		SVM	0.8480	0.8643	0.8863	0.8751
		RF	0.8452	0.8337	0.9159	0.8729
		ANN	0.8626	0.8614	0.9185	0.8890

3.4 *CS4 Result Analysis*

Figure 5 shows the feature analysis carried out for the combined feature set which includes both gesture discrepancy features in *x*, *y*, *z* directions and time and frequency domain features in t_x, t_y, t_z and f_x, f_y, f_z directions. This feature analysis determines the covariance, separability and instance distribution of features respectively. PCA and RadViz do the multivariate data analysis.

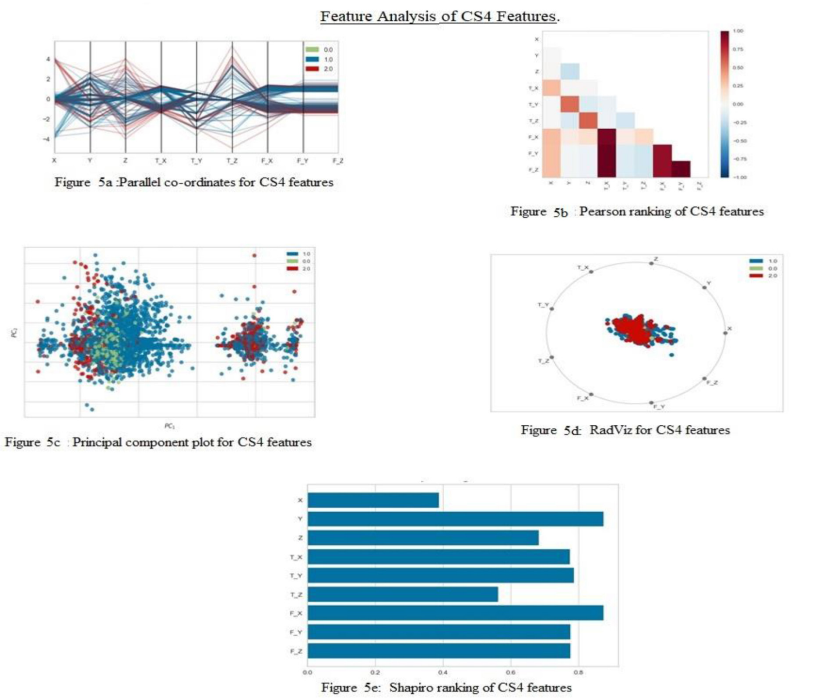

Figure 5

Feature analysis of CS4 feature set

Table 5

Results obtained in CS4 features set for respective algorithms

Computational Solution	Features	Algorithms	Accuracy	Precision	Recall	F1 Score
CS4	Combined Feature Set	KNN	0.8508	0.8446	0.9147	0.8783
		SVM	0.8605	0.8587	0.9153	0.8861
		RF	0.8745	0.8698	0.9271	0.8975
		ANN	0.8954	0.8918	0.9395	0.9151

The Table 5 talks about the results obtained when the gesture discrepancy features and the time and frequency domain features where together subjected to different machine learning algorithms respectively. The ANN algorithm gives the best result with higher accuracy, precision and recall.

Table 6 shows that out of all the four computational solutions. Out of all the four computational solutions, CS4 gives the best results due to the increased number of features which includes time domain features, frequency domain features and gesture discrepancy. The best result in whole experimental procedure is given by the ANN algorithm in the CS4 with about 89% Accuracy.

Table 6

Comparison table for results obtained in all the computational solutions

Computational Solution	Features	Algorithms	Accuracy	Precision	Recall	F1 Score
CS1	DTW	KNN	0.8215	0.8800	0.8251	0.8517
CS2	Time & Freq Domain Features	KNN	0.8350	0.8361	0.8916	0.8629
		SVM	0.8424	0.8571	0.8868	0.8717
		RF	0.8508	0.8606	0.8966	0.8783
		ANN	0.8647	0.803	0.8966	0.8884
CS3	Gesture Discrepancy Features	KNN	0.8271	0.8311	0.8828	0.8561
		SVM	0.8480	0.8643	0.8863	0.8751
		RF	0.8452	0.8337	0.9159	0.8729
		ANN	0.8626	0.8614	0.9185	0.8890
CS4	Combined Feature Set	KNN	0.8508	0.8446	0.9147	0.8783
		SVM	0.8605	0.8587	0.9153	0.8861
		RF	0.8745	0.8698	0.9271	0.8975
		ANN	0.8954	0.8918	0.9395	0.9151

4. Conclusion

In this work, human-computer interaction for gesture recognition is implemented using the ML algorithms. The data obtained from the sensor is sent to the microcontroller by the Arduino IDE programming. The microcontroller performs signal averaging to the obtained data. The result of averaging is sent to the system via the Bluetooth. In the system, the gesture recognition is performed using the KNN, SVM, RF and ANN algorithms. The output of the program will be the categorized gesture. Comparison of the accuracy of the algorithms used is also carried out. The usage of various different kinds of features is making the system more efficient and robust. The best results were achieved from ANN algorithm in the fourth computational solution and hence this suits the best for spotting eating and drinking gestures. The future scope of this work is it can be implemented as a hand wearable device for those patients whose diet maintenance is very important. Implementation using various other machine learning classification algorithms likes XG Boost, Naïve Bayes and so on to increase the efficiency of the system. Increase the dataset with more number of activities and increase the experimental task from 3 class classification to more numbers.

References

[1] Pigford & Andrews (2010) Pigford T andrews AW. Feasibility and benefit of using the Nintendo Wii Fit for balance rehabilitation in an elderly patient experiencing recurrent falls. *Journal of Student Physical Therapy Research.* 2010; 2(1):12–20

[2] S. Mitra and T. Acharya, "Gesture recognition: A survey," IEEE Transactions on Systems, Man and Cybernetics - Part C, vol. 37, no. 3, pp. 311–324, 2007.

[3] L. Hyeon-Kyu and H. Kim-J, "An HMM-based threshold model approach for gesture recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Oct. 1999.

[4] J. Alon, V. Athitsos, Q. Yuan and S. Sclaroff, "A unified framework for gesture recognition and spatiotemporal gesture segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pp. 1685–1699, 2009.

[5] S. Zhao, W. Tan, S. Wen and Y. Liu, "An improved algorithm of hand gesture recognition under intricate background," in ICIRA '08: Proceedings of the First International Conference on Intelligent

Robotics and Applications. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 786–794.

[6] J. Alon, V. Athitsos and S. Sclaroff, "Accurate and efficient gesture spotting via pruning and sub gesture reasoning," in In Proc. IEEE ICCV Workshop on Human Computer Interaction, 2005, pp. 189–198.

[7] X. Zhang, X. Chen, W.-h. Wang, J.-h. Yang, V. Lantz and K.-q. Wang, "Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors," in IUI '09: Proceedings of the 13th international conference on Intelligent user interfaces. New York, NY, USA: ACM, 2009, pp. 401–406.

[8] C. A. Wingrave, B. Williamson, P. D. Varcholik, J. Rose, A. Miller, E. Charbonneau, J. Bott and J. J. L. Jr., "The wiimote and beyond: Spatially convenient devices for 3D user interfaces," IEEE Computer Graphics and Applications, vol. 30, pp. 71–85, 2010.

[9] Mahendra Kumar Jangir & Karan Singh (2019) HARGRURNN: Human activity recognition using inertial body sensor gated recurrent units recurrent neural network, *Journal of Discrete Mathematical Sciences and Cryptography*, 22:8, 1577-1587, DOI: 10.1080/09720529.2019.1696552.

[10] Jasleen Kaur Sethi & Mamta Mittal (2019) A new feature selection method based on machine learning technique for air quality dataset, *Journal of Statistics and Management Systems*, 22:4, 697-705, DOI: 10.1080/09720510.2019.1609726.

[11] Dharm Singh Jat, Poonam Dhaka & Anton Limbo (2018) Applications of statistical techniques and artificial neural networks: A review, *Journal of Statistics and Management Systems*, 21:4, 639-645, DOI: 10.1080/09720510.2018.1475073.