



Head Gesture Recognition using a 6DOF Inertial IMU

I. C. Severin, D. M. Dobrea, M. C. Dobrea

Ionut-Cristian Severin *

Department of Applied Electronics and Intelligent Systems
Faculty of Electronics Telecommunications and Information Technology,
“Gheorghe Asachi” Technical University of Iași
Bd. Carol I, No. 11 A, Ia si 700506, România
*Corresponding author: ionut.severin@etti.tuiasi.ro

Dan-Marius Dobrea

Department of Applied Electronics and Intelligent Systems
Faculty of Electronics Telecommunications and Information Technology,
“Gheorghe Asachi” Technical University of Iași
Bd. Carol I, No. 11 A, Ia si 700506, România
mdobrea@etti.tuiasi.ro

Monica-Claudia Dobrea

Department of Applied Electronics and Intelligent Systems
Faculty of Electronics Telecommunications and Information Technology,
“Gheorghe Asachi” Technical University of Iași
Bd. Carol I, No. 11 A, Iași 700506, România
mcdobrea@etti.tuiasi.ro

Abstract

The recognition of the head movements is a challenging task in the Human-Computer Interface domain. The medical, automotive, or computer games domains are only several fields where this task can find practical applicabilities. Currently, the head movement recognition is performed using complex systems based on video information or using an IMU sensor with nine freedom degrees. In this paper, we describe a new approach for recognizing head movements using a new type of IMU sensor with six freedom degrees placed on top of a headphone pair. The system aims to provide an easy control method for people suffering from tetraplegia for a specific set of activities. The system collects data from the inertial sensor placed on top of the headphone to analyze and then extract the features for head movement recognition. We did construct and evaluated eight predictive models of classifying head movements activity to determine which one is the best fit for the proposed head movement recognition system. The comparison and performance evaluation provided by each predictive model lead to demonstrate the performances delivered by our new system for head movements recognition.

Keywords: head gesture, IMU, Head gesture recognition, intelligent system, embedded systems

1 Introduction

Detecting the position of the human body in order to control various devices or to offer the possibility to manage certain information that may affect the way of interaction between humans and computers is a common topic today [3], [17], [9], [11], [13], [4]. At this moment, the possibility of determining the orientation and classification of movements, in most cases, is used in systems that use the management and analysis of the video signal acquired by a video camera [2]. The topic regarding head movement identification becomes an essential topic in the field of human-computer interaction (HCI). It may have applicability in areas such as medicine, automotive, robot control, etc. The necessity to develop wearable systems for people with disabilities is another area of interest in our days. This kind of application can be used to control a wheelchair with help from hand gesture [10] with information that is coming from an accelerometer or using head gesture recognition with help from 4 capacitive sensors placed around the neck [4]. Other common technics used are that which perform classification of the head movements with the help of video signal acquired with a smart video camera [2] or a simple web camera [14] where the accuracy of gesture classification was equal to 87%. Also, in the technical literature exist the possibility to get head classification using 9-axis microelectromechanical system (MEMS) motion sensor such in [15], where data were acquired from 5 subject people. For each subject, a number of 5 repetitions were done, for this situation, the average classification rate was equal with 87.56%.

The proposed wearable system, presented in this paper, comes with some advantages than the system from [15]. The first advantage is the low cost, the second one regards complexity, and the last one is related to the number of head gestures classified and the number of repetitions. Also, the obtained results in this paper can be compared to those obtained by us in a previously published work [4]. In that published paper, we used to determine the position of the head four capacitive sensors placed around the neck. The accuracy of classification obtained in that publication was higher than 80%, but, in that paper, we had a small database with data acquisition from 4 subjects, and 30 recordings were done for each subject. Also, the obtained results in [4] were influenced by the characteristics of the neck circumference for each test subject, which may represent a limitation. In order to classify the movements provided by each subject, in this paper, the information provided by the pitch, yaw, and roll angles will be used instead of signals supplied by 4 capacitive sensors placed around of neck. This thing can be an advantage because the human characteristics do not affect the performances of the classification system. In this paper, we propose another classification system for head movements, which has as a primary objective obtaining a good classification rate with a low price of implementation. Also, the main focus is to obtain a new approach of head movement classification with help from a cheaper IMU sensor with the possibility to integrate into a more complex system such as a robotic wheelchair system or other HCI complex systems. Also, we have as the primary focus to prove the possibility of classifying the human head movement without using the classical method like in [2] or [14]. Validation of the new proposed classification system will be made with the help of our previous publication paper [4] and with the help of comparing it to the already existing systems [14] and [15]. Our proposed classification system is structured in 3 main parts: signal acquisition part, signal processing part, and movement classification part. The signal acquisition part for this paper is performed by the IMU 6DOF sensor and MEGA2560 development board. After this step, the head position estimation, from the acquired signals, can be done using mathematical methods such as (a) the Kalman filter, (b) low pass filter method, (c) high pass filter method, or by using (d) the complementary filter method [8]. Besides the mentioned techniques, in the technical literature, there are other methods of 3D space position identification based on an IMU sensor. One of the most used methods is also known as the Quaternions technique [5].

In this article, we will use the Kalman method and the complementary filtering method for fusion all six output signals from IMU sensors. The signal processing part represents the following step of our solution. This step will start from the moment when all tree signals yaw, pitch, and roll are valid until then all measurements were performed and saved it with the help of Visual Basic application (VBA). The acquired signals will be put inside of a Comma Separated Variables (CSV) file from where will take it and proceed offline according to the technique specified in this paper. The third part of the proposed system is composed of movement classification. In this case, several computational intelligence methods were used. The classification methods used in this paper are represented by: Support-vector machine, Random Forest, k-Nearest Neighbors, Decision Tree Classifier, Extra Tree Classifier, Gaussian Naive Bayes, Logistic Regression, Quadratic Discriminant Analysis, and Adaptive Boosting Classifier. The previously mentioned computational methods will be used for classifying 8 different commands. In the database creation process, we got the measurements from 7 peoples who did repeat for 40 times each specified command. In this paper, one of the main objectives will be to evaluate and determine the best classifier that can accurately recognize the head movements using the proposed system. Also, another focus will be put on establishing the best classification system using this new approach without using classical methods to identify the body movement. The performance analysis will be done offline, and for the next research, based on this idea, the system will be implemented and evaluated in real-time.

2 System layout

The system proposed in this paper, for movements classification determined by the position of the head, can be seen as described in Fig. 1 and is composed of three parts. The first part is represented by a 6DOF IMU sensor, which provides as output six signals, three from the accelerometer sensor and three from the gyroscope sensor. These six signals represent the main texture signals, which will be used for getting the yaw, pitch, and roll angle value. As is presented in Fig. 1, the second main part of our classification system is represented by the processing and filtering of all raw data value which coming from the IMU sensor. At this level, we used Kalman and complementary filter methods for fusion all six signals provided by accelerometer and gyroscope sensors to only 3 signals represented by values of the texture signals Roll, Pitch, and Yaw. The fusion method has done on the development board MEGA2560. The data acquisition represents the next part of our solution. This step was done with the help of a VBA application that took all of the data using de Component Object Model (COM) protocol and placed it inside of a CSV file. The VBA application uses the COM protocol to get the value transmitted by the development board MEGA 2560. For evaluation of the classification rate value provided by our proposed system, in the next part was performed an offline analysis. At this step, we prepared the acquired signals, removing the garbage part of the received signals. The length of each desired piece was selected in a range of 167 samples to 310 samples. In the last stage, the acquired and filtered command information was put together to perform an offline analysis with help of different estimation algorithms. We used Support Vector machine (SVM), Random Forest (RF), k-Nearest Neighbors (KNN), Decision Tree Classifier (DTC), Extra Tree Classifier (ETC), Gaussian Naive Bayes (NB), Logistic Regression (LR), Quadratic Discriminant Analysis (QDA) and Adaptive Boosting Classifier (ABC). Fig. 1 presents an overview of the proposed classification system.

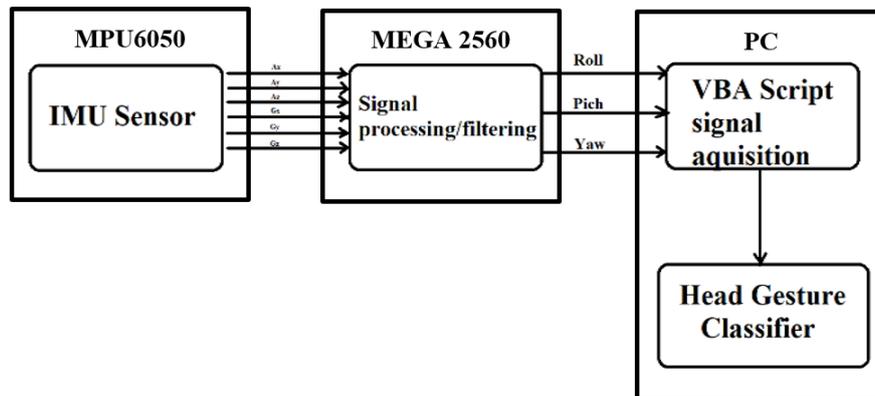


Figure 1: Block diagram of the system

3 Signal processing

3.1 Sensors calibration

As we mentioned in the previous chapter, in this paper, we use an IMU sensor for getting the orientation of the head in the 3D space. The measurement of the head movements, with the help of the IMU sensor, is possible to get an unwanted high frequency noise generated by the operating principle of the sensor and by the unwanted subject movements. To remove this issue, in this paper, we used a correction method based on storing a sample of data obtained from the sensor and applying the correction factor after calculating the average on the chosen sample. To be able to get the most accurate head position, it will be necessary to perform the calibration process for both sensors. The calibration procedure of the accelerometric sensor used in this paper comprises two components: the offset value and the standard deviation. In this paper, the standard deviation has as the main scope to eliminate the “noise” from high frequency, which can appear on each measurement axis. Equations (1) and (2) describe the calibration technique used for accelerometer sensor [7]:

$$n_{Axyz} = \frac{1}{N} \sum_{i=0}^{N-1} a_i \quad (1)$$

$$\delta_{xyz} = \frac{1}{N-1} \sum_{i=0}^{N-1} (a_i - n_{Axyz})^2 \quad (2)$$

In Equation (1) and (2), we make the following notations:

- n_{Axyz} : the value of correction for accelerometric sensor;
- a_i : the raw data from the accelerometer sensor;
- N : the maximum number of data sample;
- δ_{xyz} : the variation of the signal on $x, y,$ and z axes.

In Equation (1), the value of the maximum sample number may affect the output response rate. The higher the value of N , reduce the noises that appeared on the measuring channel to 0. Equation (2) represents the standard deviation, which is described as a consequence of the distribution of the measured signal against its average. Because the inertial system used in this work uses two sensors, for the calibration step, it is also necessary to perform the calibration on the gyroscopic sensor according to the method mentioned above. Mathematical equations adapted for the gyroscopic sensor are given in (3) and (4):

$$n_{Gxyz} = \frac{1}{N} \sum_{i=0}^{N-1} g_i \tag{3}$$

$$\delta_{xyz} = \frac{1}{N-1} \sum_{i=0}^{N-1} (g_i - n_{Gxyz})^2 \tag{4}$$

In Equation (3) and (4) we make the following notations:

- n_{Gxyz} : the value of correction for gyroscopic sensor;
- g_i : raw data from the gyroscopic sensor;
- N : maximum number of the data sample;
- δ_{xyz} : variation of the signal on $x, y,$ and z axes.

Based on the previous relations, the sensors offset was removed, and the standard deviation becomes a unity, and, now, the signals are scale-invariant. As a result, the accelerometer and the gyroscope readings are invariant to the translation and to scale.

3.2 Head position estimation

Using an MPU-6050 sensor, we determine the head gestures based on the 3D spatial movement of the head. The MPU-6050 is a microelectromechanical system (MEMS) that contains in its structure an accelerometric sensor and a gyroscopic sensor. For the data fusion, which is applied to all 6 outputs of the IMU sensor, in the literature, there are several methods such as fusion using the complementary filter [18], the Kalman filter [12], or Mahony Filter [6]. It is well known that an accelerometric sensor performs better at low frequencies, while a gyroscopic sensor performs very well at high frequencies. Because of this behavior, the value filtered with a complementary filter will be performed at both low and high frequencies according to relation (5) [1]:

$$Ang = (1 - \alpha) * (Ang + Gyro * dt) + \alpha * Acc \tag{5}$$

Following the implementation, the filtering of the yaw, pitch, and roll signals was obtained, removing the noise from the high and low frequencies, as is presented in Fig. 2. According to equation (5), the best results for filtering were obtained to the values of 0.98 for the high pass filter component and 0.02 for the low pass filter component.

Besides the complementary filter above mentioned, in this paper, we used for data fusion another technique which is called Kalman filter. Using an additional method, of data fusion in addition to the complementary filter has resulted in excellent results in terms of Yaw, Pitch, and Roll signals. The Kalman filter method is more complicated than the complementary filter, and this one contains two stage. The first stage is represented by the prediction step, while the correction component represents the second stage. The equations described at the implementation of the Kalman method, which was used in this paper can be seen in the following part [16]:

$$\hat{X}_k = A\hat{X}_{k-1} + BU_k \tag{6}$$

$$P_{k|k-1} = AP_{k-1}A^T + Q \tag{7}$$

$$K_k = P_{k|k-1}H^T [HP_{k|k-1}H^T + R]^{-1} \tag{8}$$

$$\hat{X}_k = \hat{X}_{k|k-1} + K_k (y_k - H\hat{X}_{k|k-1}) \tag{9}$$

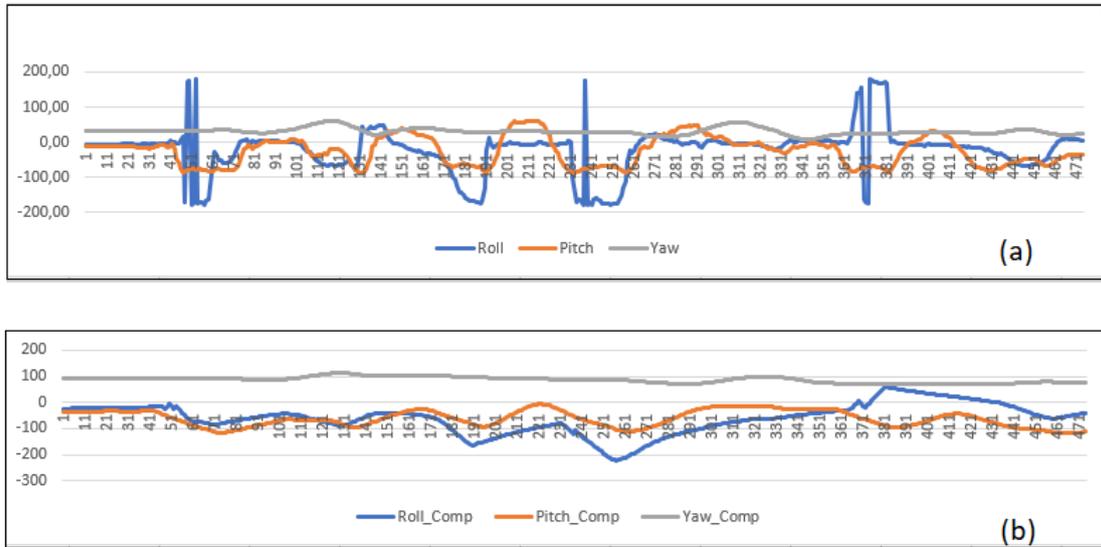


Figure 2: (a) Raw data, and (b) Complementary filter results

$$P_k = (I - K_k H) P_{k|k-1} \tag{10}$$

The previous equations were implemented and, base on them, we obtained excellent results regarding the data fusion from the accelerometer and gyroscope sensor without noises. The final pitch, roll, and yaw signals obtained by using the Kalman filter are exhibited in Fig. 3. From the figures, it can be observed that excellent filtering characteristics of the 3 signals can be observed depending on the different perturbations that appeared on the initial streams of data.

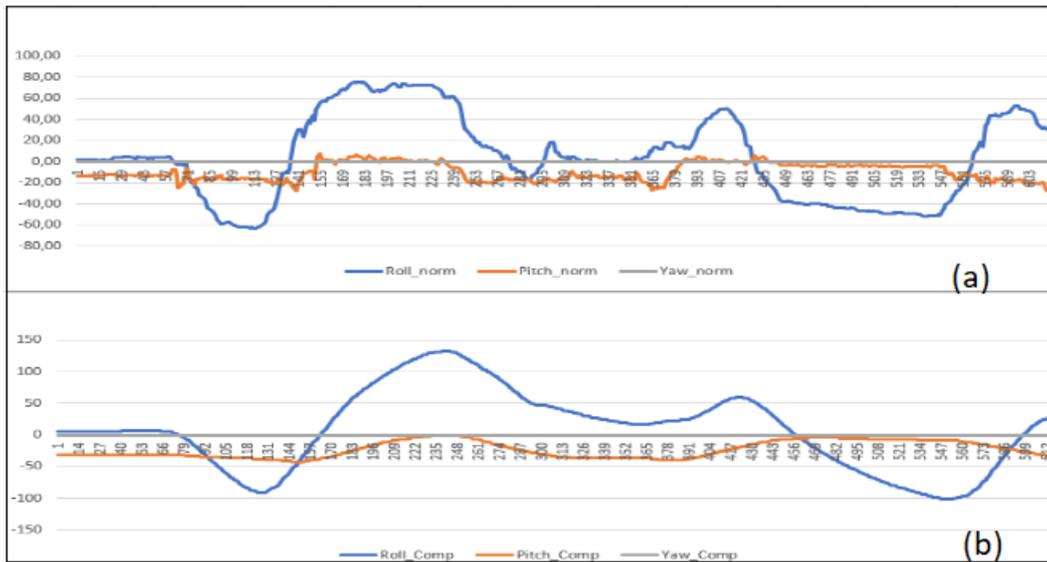


Figure 3: (a) Raw data, and (b) Kalman filter results

4 Database and methods

4.1 6DOF IMU Head gesture sensor

In the present work, the sensor used to determine the position and movements given by the orientation in 3D space is represented by the MPU6050. The MPU6050 sensor is placed on an audio headset in order to be able to provide much easier information about orientation in space, information that is transmitted to the MEGA2560 board. To respect the previously mentioned conditions, we did decide to place the 6DOF IMU sensor above the

head in the center position and equal distance between ears, such as presented in Fig.4. This approach leads to getting the head orientation command with accuracy and independent by the head circumference of each user.

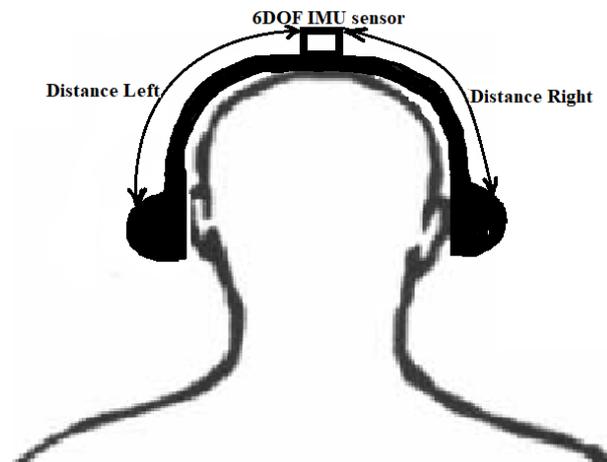


Figure 4: The 6DOF IMU sensor placement

The experimental setup created taking into account the right positioning of the IMU sensor can be seen in Figure 5.

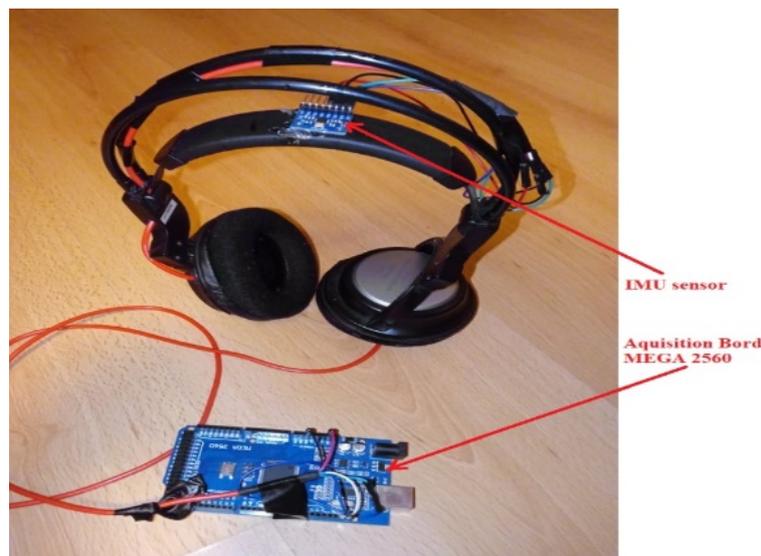


Figure 5: The wearable system

The proposed solution composed of an IMU sensor and development board MEGA2650 further captures and sends the information to the VBA application, using a serial link for signal acquisition through the COM protocol. The analysis and management of the information obtained will be done offline using the PyCharm development environment together with the python language.

The MEGA2650 development board is the main component of the acquisition system. The MEGA2650 offers a good ratio between performance and cost.

4.2 Database

The final database was implemented by acquiring data from two types of human subjects, six trained persons, and one untrained person. That six human subjects, from which the measurements were taken, were instructed before performing the head movements. All the subjects were sitting on a chair, simulating a spastic tetraparesis or tetraplegia patient. Another significant thing is that each person from whom measurements were made was trained to execute the head movement commands without moving the rest of the body. For the second case, the human subject did have the freedom to execute all commands from sitting without any restriction.

For this paper, a unique database was created after the preparation of each subject database. Each repetition frame was offline processed, and the garbage signal part was eliminated to have a pure command signal. Also, for the validation of the proposed classification system, a database that contains a set of random measurements performed by an operator that has not been trained before to perform the correct movement command was used. This thing leads to obtain a real situation classification result when a human subject uses our proposed solution to control a wheelchair or a windows application without any knowledge about command control. The acquisition of head commands was made with the help of the prototype presented in Fig. 5. The recordings were made on people who have been trained to execute the movements established for this work and implicitly to simulate the behavior of people suffering from similar tetraparesis as in a previously published work [4]. We considered a number of 8 control commands. These commands were classified using the following algorithms: Support-vector machine, Random Forest, k-Nearest Neighbors, Decision Tree Classifier, Extra Tree Classifier, Gaussian Naive Bayes, Logistic Regression, Quadratic Discriminant Analysis, and Adaptive Boosting Classifier. Because in the present work, the database was created following the data acquisition from each human subject, the method of Principal Component Analysis (PCA) was used to obtain fast and accurate classifications. The commands had been chosen to provide the possibility to control through the head movements an audio/video player or a wheelchair. The initial command sets are Move Right 2s (1-Jump to another future time period), Move Left 2s (2-Jump to another previous period), Move Next right (3-Skip to next track), Move Next left (4-Skip to one last track), Volume up (5-Volume increase), Volume down (6-Volume reduction), Start video (7-Start play), and Stop video (8-Stop play). The commands mentioned above have been established to offer the possibility of control a multimedia application that can play audio/video content, thus providing the possibility for persons with tetraparesis to be able to control certain audio/video playback states alone. For the present work, 40 repetition measurements for each human subject were purchased. Each movement of the head will be performed starting from a forward position until a return to the initial position. All lateral tilt movements of the head to the left and right will be performed with respect to the sagittal plane, and the rotations of the head to the left and right will be related to the horizontal plane. The forward-backward movements are associated with the frontal plane. The following head movements describe all 8 commands:

1. The movement of the head will be done by rotating the head twice on the right side around the yaw axis with the return of the head orientation in the right position.
2. The movement of the head will be done by rotating the head twice on the left side around the yaw axis with the return of the head orientation in the right position.
3. The movement of the head will be done by tilting the head twice in the right side around the roll axis with the return of the head orientation in the correct position.
4. The movement of the head will be done by tilting the head twice on the left side around the roll axis with the return of the head orientation in the right position.
5. The movement of the head will be done twice by turning the head towards the ceiling around the pitch axis with the return of the head orientation in the right position.
6. The movement of the head will be made by lowering the head twice to the floor around the pitch axis with the return of the head orientation in the right position.
7. The movement of the head will be done by lowering and rooting twice the head towards the ceiling and the floor around the pitch axis with the return of the orientation of the head in the right position.
8. The movement of the head will be done by turning the head twice from the right maximum to the maximum left around the yaw axis with the return of the head orientation in the right position.

The orientations of the head were made in the 3-dimensional domain related to the 3 axes given by the values of the yaw, pitch, and roll. The acquisition of the received and filtered signals was made using a simple application in the VBA that takes the information from the serial port and insert the values in a CSV file from which they were copied, processed, and added in the database. For the final database, which was used for analysis, we did perform the preparation of each measurement data set for clear the garbage part of the acquired signal. This procedure has helped us to maintain just a utile part of signals for training the used algorithms. This step enabled us to train the intelligent algorithm with a pure pattern signal without the noise or garbage part. Also, for validation of the proposed classification system, the final database includes a measurement which coming from a subject that was not trained to execute the desired command correctly. This fact will provide us information about the system performances based on an unpredicted data set. Because the similarity between each command frame provided by all subjects is very important, in this paper, we chose to have a length in a range of 167 samples to 310 samples.

5 Results

For this research paper, the training and test data set were variable represented for each established 8 classifications classes. The length size for each control class was chosen independently for each human subject. The main idea was to simulate an unpredictable behavior with different speed of execution control commands. The acquisition of data from MEGA2650 was made at the maximum system speed of 16MHz. As a direct result, the sampling frequency for the head movements (acceleration and gyroscopic information) was 50 Hz. For this research paper, 6 unique databases were created for each of the 6 subjects, which performed all head moving command from sitting position on a chair with a backrest, thus simulating a patient suffering from spastic tetraparesis. Besides the previous measurements which were done from 6 unique subjects, we did perform another database from the subject that was not trained before. With this database, we did try to simulate and validate the real condition when it is possible to execute the command in the wrong way. For the seventh database, the human subject naturally performed the movements, regardless of whether the body and head position were kept in a vertical position.

In this part of the paper, we will present the performances obtained from the offline analysis using the predictive algorithms mentioned in the previous chapters. The offline analysis was done in the 2017 PyCharm development environment. All created Data Bases (DB) were split for each time in two parts. The first one is the training set, and the test set represents the second part. For this paper, the test set was chosen to 20%. The training set was 80%. In Table I, the results are presented. These outcomes helped us to select the best predictive algorithms for further analysis in future research. Extensive analysis was done comparing results using PCA based feature vectors, Linear Discriminant Analysis (LDA) based feature vectors, and raw based feature vectors. The distribution of DB with all measurement values from 7 people subject can be seen in Fig. 6. From Fig. 6 we can get some conclusion regarding the range of movement reported at the reference Yaw-Pitch-Roll (YPR) system. The maximum range for Yaw signal is between 60° and -40° , Pitch was stimulated between 30° to -80° , in finally the Roll value were in range -45° to 40° this information can be seen in left representation from Fig. 6. Also, if we take into consideration Fig. 6 we can make some conclusions about database and value manipulated on each axis yaw, pitch, and roll. From this figure, we can see the fact that all acquired signals are repetitive information provided by diagonal representation in time what relation between axis yaw, pitch, and roll provided us information if data can be grouped at the specific classes. For performing a proper evaluation of the proposed system in this paper we use some metrics as Classification Accuracy, Precision factor, F1 Score, and Recall factor. Before starting to present all results provided by the proposed system, we will present all specified metrics. In this paper, classification accuracy is used to get the ratio of correct prediction from a maximum number of inputs, the equations which describe this metric can be seen in the (11):

$$Accuracy = \frac{\text{Number of correct prediction}}{\text{Total number of prediction}} \quad (11)$$

Another interesting metric is represented by a precision factor who tell us how many positive results get according to the number of positive predicted value by prediction algorithm. The equation which describes this factor is specified in (12):

$$Precision = \frac{\text{True positive}}{\text{Total Predicted Positive}} \quad (12)$$

The recall factor provides us information about the predicted class if truly belongs to the predefined class. The equation (13) describes this factor.

$$Recall = \frac{\text{True positive}}{\text{Total Actual Positive}} \quad (13)$$

Based on the F1 score, we will estimate how many instances can be classified correctly or how many are classified incorrectly. The equation which described the previous affirmation can be seen (14):

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

Because the size of the created database is large, we chose to use an additional technique to reduce the dimensionality of the data set and also to reduce the running time for each algorithm. For this situation, we did use PCA and LDA techniques. Those techniques presented in the previous section are linear transformation techniques, and the difference between them is provided by the fact that LDA is supervised, whereas PCA is unsupervised. According to the information presented in table 1, we got a good and similar classification ratio for all of each 3-analysis technique applied to the final database, which contains data from 7 persons. After the offline analyses, we got a set of algorithms that provide an accuracy higher than 80%, few from these good

classifiers are: Random Forest, Extra Trees Classifier, Support Vector Machine, Decision Tree and k-Nearest Neighbors.

All the analyses were performed on the previously presented database, containing data from the recording obtained from 7 different subjects. From these subjects, 6 were trained people, and 1 on was an untrained subject. This fact generates more accurate and realistic results if we compare this with the previous paper [4]. This conclusion comes from the fact that the solution presented in this paper contains a large and unique data from two kinds of situations which can simulate a real condition. Also, considering that the approach of the performance analysis is similar, the results are superior for the present case. Another thing that confirms the good performances obtained in this work is offered by the values of the performances obtained for each database corresponding to each human subject. In the previous publication [4], we got a maximum precision equal to 91.67% on an independent database; in time, what in this paper, we got a maximum value equal to 93% . This thing leads to the fact that our classification system presented here is more accurate than the previous research publication.

The obtained accuracy is a competitive result if we report our findings to the solution provided by the predictive system using a web camera [14] where accuracy was 87% . The video system complexity [14] is higher, and the possibility to include that solution to control another system, such as a wheelchair, can be a complicated approach. Other reasons that confirm the excellent performance provided by the proposed system can be obtained through comparison with the solution provided by an IMU with nine degrees of freedom (9DOF) [15]. In [15] the accuracy achieved is equal with 87.53% with data acquired from 10 subjects. The main advantages provided by our proposed solution is that regarding: cost, number of commands, good classification, and the possibility to integrate our system easily in a more complex application.

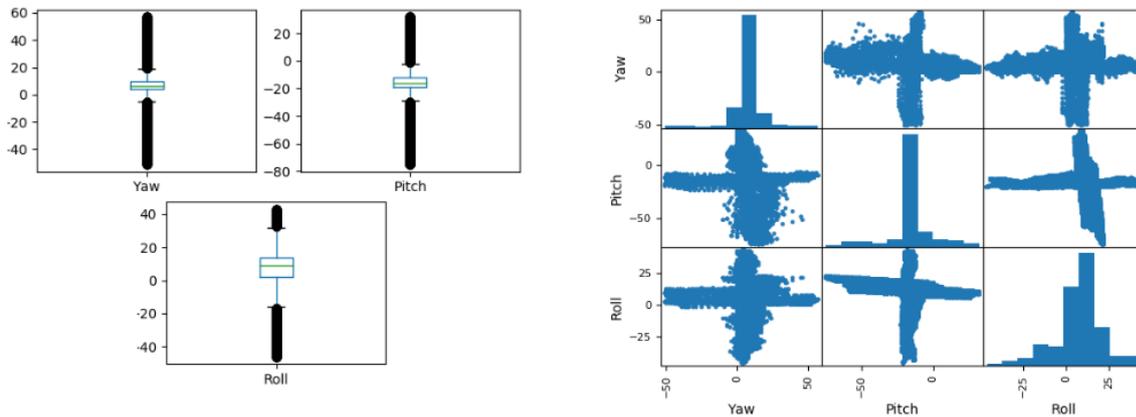


Figure 6: Data distribution and valid range of the DB which include 7 people subjects

Table 1: Result obtained on the test set 20% of the DB which include 7 people subjects

Classifier	PCA-based feature vectors Best accuracy	LDA-based feature vectors Best accuracy	Raw data-based feature vectors Best accuracy
Random Forest k-Nearest	78.03%	78.46%	80.4%
Neighbors(kNN)	77.58%	77.66%	78.1 %
Decision Tree	77.96%	78.13%	78.9%
Gaussian Naive Bayes(NB)	45.08%	44.64%	41.5%
Logistic Regression	39.97%	39.97%	41.8%
Extra Trees Classifier	78.87%	78.68%	80.49%
Quadratic Discrimination Analysis	45.60%	45.60%	45.60%
ADA Boost Classifier	33%	35.08%	77.6%
Support Vector Machine (SVM)	61.62%	62.57%	81.26 %

In order to be able to determine the performances offered on each specified class, the Random Forest algorithm was chosen from the list of the algorithms determined to be with the best classification rate, see Table 2. Table 2 presents the obtained performances on each independent database.

From Table 1 we got the best prediction algorithm, which is represented by Random forest. This classification algorithm is composed by a number of 100 trees which combine all result from them using the random state for controlling randomness of the bootstrapping, samples used for building trees and sampling of the features used to choose the best split on each node from the predictor. K-Nearest neighbors' algorithm represents

Table 2: Result obtained with the best classifier for each subject people

Random Forest	Overall accuracy [%]													
	S1 (90.12%)		S2 (83.75%)		S3 (91.25%)		S4 (93%)		S5 (90.87%)		S6 (80%)		S7 (85.25%)	
	Prec	Recall	Prec	Recall	Prec	Recall	Prec	Recall	Prec	Recall	Prec	Recall	Prec	Recall
JumpNextLeft	0.89	0.93	0.87	0.89	0.95	0.97	0.96	0.96	0.95	0.97	0.91	0.89	0.89	0.96
JumpNextRight	0.93	0.92	0.78	0.80	0.94	0.94	0.93	0.91	0.91	0.94	0.96	0.88	0.82	0.76
MoveRight_2s	0.89	0.95	0.81	0.85	0.85	0.88	0.92	0.87	0.85	0.87	0.96	1.00	0.83	0.84
MoveLeft_2s	0.95	0.97	0.92	0.97	0.95	0.98	0.90	0.90	0.95	0.96	0.95	0.97	0.82	0.91
RisVolume down	0.92	0.94	0.95	0.94	0.89	0.87	0.91	0.95	0.89	0.88	0.37	0.44	0.85	0.90
RisVolume up	0.87	0.92	0.71	0.72	0.89	0.90	0.92	0.97	0.89	0.89	0.90	0.80	0.87	0.86
StartVideo	0.89	0.80	0.76	0.70	0.94	0.88	0.95	0.94	0.94	0.91	0.36	0.35	0.88	0.75
StopVideo	0.87	0.84	0.90	0.86	0.89	0.89	0.95	0.93	0.89	0.89	0.99	0.97	0.86	0.86

Table 3: Result obtained with the best classifier for all subject people DB

Random Forest	Precision Factor	Recall Factor	F1-Score	Accuracy Factor
JumpNextLeft	0.84	0.88	0.86	0.80
JumpNextRight	0.81	0.81	0.81	0.80
MoveRight_2s	0.75	0.76	0.75	0.80
MoveLeft_2s	0.83	0.87	0.85	0.80
RisVolume-down	0.70	0.75	0.72	0.80
RisVolume-up	0.79	0.81	0.80	0.80
StartVideo	0.75	0.67	0.71	0.80
StopVideo	0.78	0.74	0.76	0.80

another classification algorithm used with a good result for this paper. This algorithm is based on the similarity between different points and takes into consideration the fact that similar data points are usually near to each other. For this paper k-nearest data point factor was chosen to be equal with 5. As a metric, we used the Makowski metric with distances calculated based on the Euclidian distance. The best performance for this paper using the kNN predictor algorithm was obtained with k factor equal with 3, and for this situation, the accuracy was equal with 78.1%. From Table 2 we can see the precision on each command for all 7 subjects. This parameter empowers this solution as a good choice for the determination of the head movements if it takes into consideration the cost and complexity factor. The average value of the precision factor for all 7 human subjects on each command can be seen in Table 3 where the precision factor is up to 70%. From Table 3 we can also confirm the excellent performance of the proposed solution with the F1 score, which tells us about how many instances of each command can be correctly classified or not. In our case, we got a good score of up to 71%, until 86%. Confirmation if the predicted value truly belongs to a specified class is provided by the Recall factor, which is up to ~70% for each command.

6 Conclusion

The system presented in this paper was proposed to provide the possibility to classify a number of 8 commands. Such a system offers the possibility to control a media player by tetraplegic people. After the different analyses presented above, we got good results according to the main proposed scope. As a first conclusion, our idea can be very competitive if we take into consideration the fact that for the moment, only a few similar solutions exist which make head movement classification using a web camera [14] or a more complex IMU 9DOF sensors [15]. In this paper, we got a good classification accuracy of 81%, which can be a competitive result if we report to a web camera solution [14], which got an 87% correct classification rate. Still, the video system is a more complex one, challenging to include it in another complex system which means both more processing times and higher implementation costs. Another solution, presented in the literature, is based on a head gesture acquisition system using a 9 DOF IMU [15]. This solution provides 87.56% accuracy with the classification of 4 head commands. The data were obtained from 10 subjects. Our solution provides an 81% accuracy obtained on a number of subjects as half as in [15] and on a number of classification commands double than in [15], while the implementation cost is smaller than in [15]. In addition to the solutions mentioned above [14], [15], the idea proposed in this publication has been evaluated using our previous system [4], where for getting the head position, we used 4 capacitive sensors. In the previous approach [4], we got an accuracy of 70% for a database containing raw, unprocessed data from 5 trained human subjects. When the PCA techniques were used, we succeeded to raise the accuracy to 72-75%. In the previous approach [4], we got an accuracy of 70% on a database containing raw, unprocessed data from 5 trained human subjects. When the PCA techniques were used, we succeeded to raise the accuracy to 72-75%. Unlike the previous solution [4], the current one has achieved far superior results: a precision of up to 80% obtained even if a higher number of both commands (i.e., eight commands) and subjects (i.e., 6 trained and one untrained human subjects) was used. These results may claim for our idea to be considered as one of the good alternatives for getting the head orientation instead of the existing classical solutions [4], [14] or [15]. There are two main approaches to both test and use an intelligent

system: offline and online. In the offline approach, the system is tested based on an independent dataset, other than the data set on which the system parameters were extracted. Due to the nowadays existing large number of open-source data sets, the offline testing method has been far more used. However, in a real-time scenario, the newly developed intelligent system must be embedded into a specific application and tested in interaction with a real environment. Our new proposed system (see Fig. 4 and Fig. 5) was built, having in mind the online scenario. The intelligent system was developed as a system able to interact with an unconstrained environment, receiving inputs from different specific head movements, each of them being executed in variable time intervals specific to each user. If we will have to use within an online environment our offline trained system, we can expect a degradation of the classification performances mainly due to some unlearned particular head movement characteristics that were not previously encountered by the system. However, this decrease in classification rate can be taken away by further employing the main idea of an online system, namely, by continuously updating the parameters of the classification system during and after each data input. So, in order to implement such an online system, we will have to adjust our software components according to this rule, following that a special focus to be put on critical processing times. In conclusion, with some future improvements done, we consider that our system will meet all the required prerequisites in order to work in a real-time scenario. In this research paper, we had proposed a new kind of prototype for the determination of head movements for tetraplegic people who can not control their limbs. This solution provides as main advantages fact that is easy to be implemented, good acquisition cost, possibility to be included easily in the more complex system, and excellent performances reported at a similar solution. The solution described in this paper can be successfully framed in the field of computational intelligence because the movement evaluation, which was executed for each command, was done with help from intelligent supervised algorithms. Through the solution described in this paper, we tried to highlight a new method of monitoring the movements determined by the human body, using a system with a low cost of implementation and with high performances regarding the performed activities classification rate. Also, with the proposed solution in this paper, we wanted to highlight the possibility of creating a simple and intelligent HCI system that can determine with high precision the activities performed by the human operator. This solution can come to the aid of people who suffer from spastic tetraplegia, to control specific devices or systems of movement (wheel chair) without high implementation costs. The method proposed by us is an intelligent method that can offer the possibility of its integration in complex and smart systems in order to make it possible to control certain processes by creating a favorable human-computer interface environment. In this paper, the main applicability of our idea was in the medical area. Still, our solution can have a lot of applicability in other fields such as Automotive, IoT, HCL, etc. Also, we propose to implement the established best classifier in this paper, in the real-time application for controlling a media player or for an intelligent wheelchair system with the dynamic classification of head gestures. Another aim was to increase the database with more acquired data from a different human subject and to increase the number of classified head gestures. In this way, we try to obtain an excellent system able to classify a high number of head movements with a good rate of classification.

Funding

The publication of this research was sponsored by the Bosch Romania, Department: Electrically Assisted Steering Systems (EES).

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

- [1] Ariffin, N.; Arsad, N.; Bais, B. (2016). Low cost MEMS gyroscope and accelerometer implementation without Kalman Filter for angle estimation, *2016 International Conference on Advances in Electrical, Electronic and Systems Engineering (ICAEES)*, IEEE, 77-82, 2016.
- [2] Bankar, R.; Salankar, S. (2015). Head Gesture Recognition System Using Gesture Cam, *2015 Fifth International Conference on Communication Systems and Network Technologies*, IEEE, 8, 341-346, 2015.

- [3] Chen, Y.; Yang, J.; Liou, S.; Lee, G.; Wang, J. (2008). Online classifier construction algorithm for human activity detection using a tri-axial accelerometer, *Applied Mathematics and Computation*, 205(2), 849-860, 2008.
- [4] Dobrea, M.; Dobrea, D.; Severin, I. (2019). A new wearable system for head gesture recognition designed to control an intelligent wheelchair, *The 7th IEEE International Conference on E-Health and Bioengineering - EHB 2019*, IEEE, 1-5, 2019.
- [5] Feng, K.; Li, J.; Zhang, X.; Shen, C.; Bi, Y.; Zheng, T.; Liu, J. (2017). Correction: A New Quaternion-Based Kalman Filter for Real-Time Attitude Estimation Using the Two-Step Geometrically-Intuitive Correction Algorithm, *Sensors*, 17(9), 2146, 2017.
- [6] Ferdinando, H.; Khoswanto, H.; Purwanto, D. (2012). Embedded Kalman Filter for Inertial Measurement Unit (IMU) on the ATMega8535, *2012 International Symposium on Innovations in Intelligent Systems and Applications*, IEEE, 1-5, 2012.
- [7] Guan, Y.; Song, X. (2018). Sensor Fusion of Gyroscope and Accelerometer for Low-Cost Attitude Determination System, *2018 Chinese Automation Congress (CAC)*, IEEE, 1068-1072, 2018.
- [8] Islam, T.; Islam, M.; Shajid-Ul-Mahmud, M.; Hossam-E-Haider, M. (2017). Comparison of complementary and Kalman filter based data fusion for attitude heading reference system, *AIP Conference Proceedings*, 1919, 020002, 2017.
- [9] Jeong, G.; Truong, P.; Choi, S. (2017). Classification of Three Types of Walking Activities Regarding Stairs Using Plantar Pressure Sensors, *IEEE Sensors Journal*, 17(9), 2638-2639, 2017.
- [10] Kumar, V. (2015). MEMS based Hand Gesture Wheel Chair Movement Control for Disable Persons, *International Journal of Current Engineering and Technology*, 5(3), 1774-1776, 2015.
- [11] Lara, O.; Labrador, M. (2013). A Survey on Human Activity Recognition using Wearable Sensors, *IEEE Communications Surveys & Tutorials*, 15(3), 1192-1209, 2013.
- [12] Ludwig, S.; Burnham, K.; Jiménez, A.; Touma, P. (2018). Comparison of attitude and heading reference systems using foot mounted MIMU sensor data: basic Madgwick and Mahony, *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018*, 10598, 2018.
- [13] Mukhopadhyay, S. (2015). Wearable Sensors for Human Activity Monitoring: A Review, *IEEE Sensors Journal*, 15(3), 1321-1330, 2015.
- [14] Ng, P.; De Silva, L. (2001). Head gestures recognition, *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, IEEE, 3, 266-269, 2001.
- [15] Rudigkeit, N.; Gebhard, M.; Graser, A. (2015). An analytical approach for head gesture recognition with motion sensors, *2015 9th International Conference on Sensing Technology (ICST)*, IEEE, 1-6. 2015.
- [16] Thacker, N.; Lacey, A. (1998). *Tutorial: The Kalman Filter*, Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester, 1998.
- [17] Truong, P.; You, S.; Ji, S.; Jeong, G. (2019). Wearable System for Daily Activity Recognition Using Inertial and Pressure Sensors of a Smart Band and Smart Shoes, *International Journal of Computers Communications & Control*, 14(6), 726-742, 2019.
- [18] [Online]. Available: <https://www.pieter-jan.com/node/11>, Accessed on 21 Dec 2019.



Copyright ©2020 by the authors. Licensee Agora University, Oradea, Romania.

This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.

Journal's webpage: <http://univagora.ro/jour/index.php/ijccc/>



This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).

<https://publicationethics.org/members/international-journal-computers-communications-and-control>

Cite this paper as:

Severin, I. C.; Dobrea, D. M.; Dobrea, M. C. (2020). Head Gesture Recognition using a 6DOF Inertial IMU. *International Journal of Computers Communications & Control*, 15(3), 3856, 2020.

<https://doi.org/10.15837/ijccc.2020.3.3856>