Implementation of Classification Algorithms in a Smart Glove for Hand Gesture Detection

Felipe Quirino, Marcelo Romanssini, Mathias Baldissera and Alessandro Girardi
Computer Architecture and Microelectronics Group - GAMA
Federal University of Pampa - UNIPAMPA
Alegrete, RS, Brazil
Email: felipeantunesquirino@alunos.unipampa.edu.br

***Abstract*—The use of portable device is diffused around the world. It is becoming popular the use of these devices to mitigate the problem that visually and hearing impaired people face in communicating, mainly through the capture and identification of natural gestures. We developed a gesture detection glove capable to translate the hand movement into useful information, such as a letter of the alphabet, allowing typing words and phrases. The present work demonstrates the performance of different machine learning algorithms to recognize gestures corresponding to the letters of alphabet in LIBRAS signal language. We measured the prediction time and accuracy of six different classification algorithms. All algorithms achieved acceptable results, even though the slowest prediction time was about 130 times higher than the one provided by the faster algorithm. The precision varies from 88% to 94% for the evaluated models. Support Vector Machine and Linear Discriminant Analysis achieved the best results in terms of precision. Logistic Regression and Linear Discriminant Analysis are best in terms of time prediction. These three algorithms are candidates for future implementation in a dedicated embedded system to be integrated on the glove.**

I. Introduction

Social integration of people presenting some kind of physical impairment - such as the inability to speak or to hear - is mandatory in the modern world. Usually these people communicate by means of a sign language using gestures. However, this language, in general, is still restricted to a small part of the population, making the communication difficult. The use of modern electronic devices can mitigate this hurdle. A portable device for translating sign language to spoken language in real time is of great importance for the complete social integration of hearing and visually impaired people [1].

Some approaches for implementing this kind of device are proposed in the literature. The methods can be classified in glove-based [2], vision-based [3], depth-based or a combination of them [4] [5]. The data glove method has the advantages of providing a small amount of input data and high speed. Moreover, it can provide 3D information about the hand and detect the independent position of each finger. The method of vision-based gesture recognition uses camera to collect gesture image sequence and identifies the gestures by processing the image. This method provides a good precision, but it is constrained by lighting sensitivity and the distance of the user from the camera. The depth-based method relies on a distance measuring hardware, which provides 3D geometric information. The camera must have a resolution that allows to recognize fingertips and small changes of finger position. The quality of the analysis furthermore depends on the position of the user in front of the camera. All methods use machine learning techniques for identifying gestures, such as supervised classification algorithms. However, there is not a consolidated methodology capable to result in an end-user popular device. So, further research is necessary in this area.

This work has the goal to analyze the performance of classification algorithms implemented in a gesture detection glove. The used glove contains gyroscope sensors located at each fingertip and at the dorsal part of the hand, allowing the development of several applications. The glove may replace conventional joystick, mouse and keyboard devices with the advantage of inserting more degrees of freedom to the control. The application here addressed uses the glove as an input device to interpret the Brazilian Sign Language (LIBRAS) [6], thus translating the posture of the hand into the corresponding letter of the alphabet. There are different algorithms capable to evaluate the sensor data and identify the corresponding letter. Different approaches can result in different performances related to accuracy and prediction time. This paper intends to compare machine learning algorithms for the interpretation of hand gestures in LIBRAS sign language aiming a posterior implementation in a dedicated embedded hardware.

II. Gesture Detection Glove

The gesture detection glove used in this work was fully implemented by our group. It is composed of six sensors, a microcontroller, wires and a Bluetooth module. Fig. 1 depicts the prototype of the gesture capture glove [1]. Hand movement data is captured by six MPU 6050 sensors arranged at each fingertip and at the dorsal area of the hand. Each MPU 6050 has an accelerometer and gyroscope, providing linear and angular acceleration data on three-axes (X, Y, Z), which allows good precision for motion capture. The accelerometer is capable to measure accelerations in the three axes separately up to 16g (156.896 m/s^2) and has four programmable ranges: 2g, 4g, 6g and 16g. The unit g refers to the value of the Earth gravitational acceleration. The gyroscope is capable to measure instantaneous angular variations on the three axes separately with four programmable ranges: 250°/s, 500°/s, 1000°/s and 2000°/s.

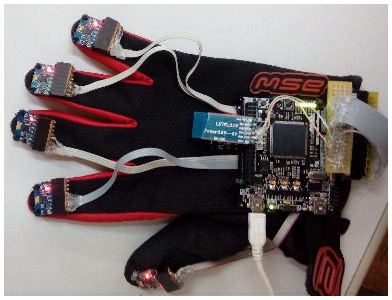

Fig. 1. Prototype of the gesture capture glove.

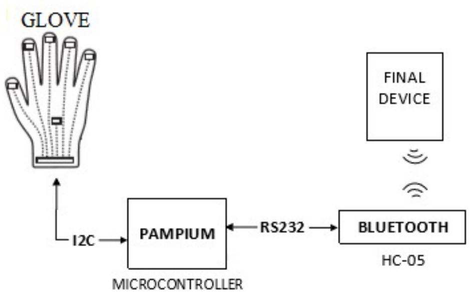

Fig. 2. General scheme of the entire system of the gesture capture glove.

The control of the glove is performed by the PAMPIUM microcontroller, which is responsible for configuring and reading the sensor data. Both the microcontroller and the communication interfaces are implemented in FPGA. A Bluetooth module is used to transmit the data read from the sensors to the final device (a computer or a smartphone). This wireless transmission allows great mobility for the user. The overall power consumption of the entire system is estimated in 0.45 W. Fig. 2 shows the general scheme of the system.

The recognition of letters from gestures is a practical application for the glove, in which the user makes a certain hand signal and the application understands it as a key pressed in the keyboard.

We adopted the LIBRAS alphabet, as shown in Fig.3, as standard for the gesture recognition. LIBRAS is the official Brazilian signal language and its use is widespread along the country. Each letter corresponds to a different static finger configuration. With the combination of various gestures it is possible to spell complete words and phrases.

The procedure for identifying the letters of the alphabet is made with the aid of machine learning. For supervised learning, we collected 10 gesture repetitions corresponding to each letter of the alphabet from 8 volunteers. To collect the data, we implemented a Matlab interface in a computer, in which it is possible to visualize and organize the raw data read from the sensors. We tested six classification algorithms in order to identify which algorithm offered better performance considering accuracy and prediction time.

Fig. 4 shows the patterns of the values of gyroscopes corresponding to the letters A, B and W captured with the glove. It is possible to notice a clear difference between the

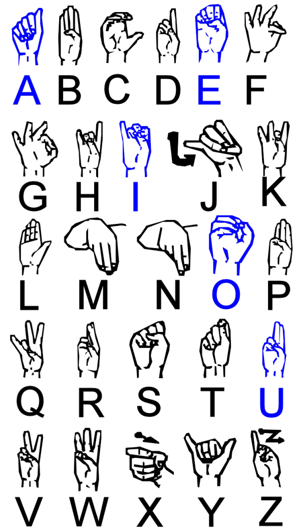

Fig. 3. Gesture corresponding to the letters of the alphabet in the Brazilian signal language.

data behavior for the 3 letters - specially between B and W, although the gestures corresponding to these letters are quite similar.

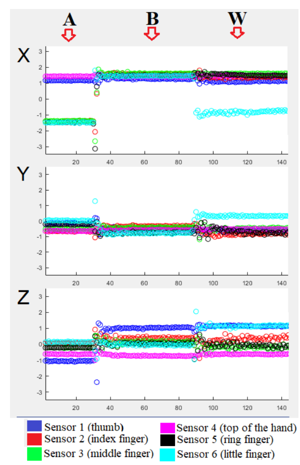

Fig. 4. Gyroscope patterns captured with the glove for the gestures corresponding to the letters A, B and W in LIBRAS.

With a trained model it is possible to perform classification of the gestures in classes and predict the corresponding letter of the alphabet. The prediction algorithm flowchart is depicted in Fig. 5. It receives the gyroscope data read by the 6 sensors referring to a certain static gesture. First, it verifies if the hand is stable (no movement) prior to classifying. If the hand is not stable, the algorithm returns to reading the sensors again. When the hand is stable, the algorithm performs the classification. For the classification, the gyroscope information is the input vector of the trained model, which returns the class with the higher probability of containing the gesture. Finally, the predicted letter of the LIBRAS alphabet, corresponding to the returned class, is printed on the computer screen.

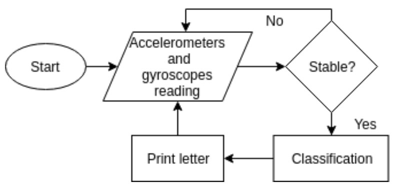

Fig. 5. Flowchart of the prediction algorithm.

III. Analysis of Classification Algorithms

The classification algorithms must be trained with samples of known patterns in a supervised learning process. The model resulting from the training process is used to predict the corresponding letter of unknown samples.

The design space is divided in 28 classes corresponding to all letters of the alphabet plus the symbols for ”space” and ”idle” (hand in resting mode). The sensor data has 18 features, composed of the measurements of six 3-axes gyroscopes previously described, captured from the glove at a given instant.

We selected six machine learning algorithms in order to evaluate which one adapts better to our application and produces the best performance in terms of accuracy and prediction time. They are the following: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Logistic Regression (LR) and Random Forest (RF) [7].

The LR algorithm can be understood simply as finding the β parameters for the $F(x)$ function, which is the probability of the dependent variable x belonging the a given class:

$$F(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \tag{1}$$

The class with highest probability is considered the predicted class.

The strategy of SVM is to trace a hyperplane that maximizes the distance between classes. It can use different kernels for the similarity function, such as linear, polynomial, Radial basis function kernel (RBF) or sigmoid. In this work we use the RBF kernel.

The KNN algorithm uses the distance between two points to classify the samples. It identifies a point in the middle of each class and evaluates the Euclidean distance between this and the sample to be classified with the following function:

$$F(x) = \sqrt{(p - q_x)^2} \tag{2}$$

where p is the point to be predicted and q_x is the central point of each class. The point p is then assigned to the nearest class.

Decision trees are based on a set of true/false decisions that causes the algorithm to arrive at a determined class. Random forest is a generalized form of decision trees, assigning weights to the features in order to avoid overfitting [7].

The QDA algorithm approaches the problem by assuming that the conditional probability density functions are normally distributed. Under this assumption, the Bayes optimal solution is used to predict points as being from a class if the logarithm of the likelihood ratio is bigger than some threshold T.

The LDA algorithm finds a linear combination of features that characterizes two or more classes of objects. The resulting combination may be used as a linear classifier. It is quite similar to QDA, with the difference that LDA makes the additional simplifying homoscedasticity assumption that the class covariances are identical and that the covariances have full rank.

The metrics for evaluating classification algorithm performance are precision, recall and F_1-score. A predicted class can result in four cases: true positive (TP), false positive (FP), true negative (TN) and false negative (FN). Precision is moreover referred to as positive predictive value (PPV) and can be defined as:

$$PPV = \frac{TP}{TP + FP} \tag{3}$$

Recall is true positive rate (TPR), given by:

$$TPR = \frac{TP}{TP + FN} \tag{4}$$

The harmonic mean of precision and recall is the balanced F_1-score (F_1):

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{5}$$

The precision decays if the algorithm predicts many false negatives, and the recall decays when there are many errors in a single class [8].

The prediction time demanded by the generated model is moreover important for evaluating the classification method. The algorithm may be unfeasible for the application if the test time is bigger than the data update. It can be estimated by measuring the average time needed for testing a vector of input instances.

A. Model precision

We used labeled data collected by volunteers corresponding to the letters of the alphabet in LIBRAS. We trained the classification algorithms with a dataset composed of 10 gestures for each letter from 8 different people. This dataset was divided in a training set composed of gestures from 6 random people and a test set with the gestures from the 2 remaining people.

The algorithms which presented best precision were SVM and LDA, with 94% of correct predictions. Worst result was obtained by KNN, with 88%. Fig. 6 shows the results. There is a small difference between the best and the worst algorithms, but the effect of a smaller precision is relevant if considered a high number of predictions.

Table I shows the comparative for precision, recall and F_1-score for the algorithms. Considering F_1-score and the recall, SVM and LDA presented again the best values. SVM achieved 0.93 for TPR and F_1 and LDA obtained 0.91 for both metrics. The high F_1-score demonstrates that few classes have the high error rate and that the remaining present low error rate.

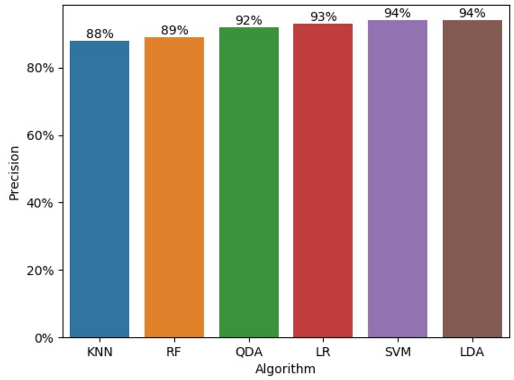

Fig. 6. Precision for each evaluated algorithm after training and testing.

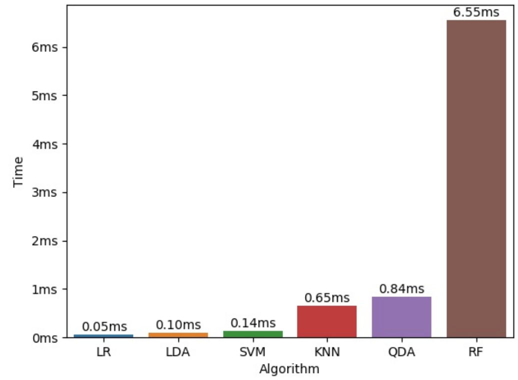

Fig. 7. Time necessary for a single prediction.

TABLE I
PERFORMANCE METRICS ESTIMATED FOR THE CLASSIFICATION ALGORITHMS.

Algorithm	PPV	TPR	F_1
K-nearest neighbors	0.88	0.87	0.87
Random forest	0.91	0.90	0.90
Quadratic discriminant analysis	0.92	0.90	0.90
Logistic regression	0.93	0.91	0.91
Support vector machine	0.94	0.93	0.93
Linear discriminant analysis	0.94	0.91	0.91

B. Prediction time

The glove is capable to send to the computer about 100 new sensor data updates per second. Thus, the prediction model must be able to perform a single test in a time smaller than 10 ms. To evaluate the prediction time, we executed 10,000 runs for each generated model and measured the average time it takes to classify data. This experiment was executed in a computer with a 2.4 GHz i7-5500u processor, using only one thread for each algorithm.

The resulting average prediction times for each model are depicted in Fig. 7. It can be seen that the model produced by the LR algorithm is the faster one, needing only 0.05 ms for a single prediction. The LDA algorithm furthermore produces a fast prediction model, spending 0.10 ms in average. The slower model was produced by the RF algorithm, demanding 6.55 ms for a single prediction.

These results demonstrate that all heuristics can be used for real-time prediction in the target glove. The LDA and SVM, which obtained best accuracy results, produced moreover very fast prediction models, with a mean of 0.10 ms and 0.14 ms for a single prediction, respectively. It demonstrates that both heuristics, together with the LR algorithm, are good candidates to be implemented in a dedicated embedded system with limited hardware resources, although other aspects need to be investigated, such as the requirements of memory and arithmetic functions.

IV. CONCLUSION

This paper described the comparison between different classification algorithms for translating a set of angular acceleration data generated from a gesture capturing glove into a character in a computer. The accuracy results demonstrate that all algorithms are feasible and present high possibility for practical use.

SVM and LDA achieved the best results considering accuracy, with 94% of correct prediction cases. They presented high precision, F_1-score and recall. However, considering the prediction time, the LR algorithm achieved the best performance, being 2 times faster than the second faster model. These 3 classification algorithms are candidate to a future implementation in a dedicated system embedded on the glove.

The hand gesture recognition system herein proposed collaborates to overcome the communication problem of visual and hearing impaired people, providing an efficient tool for their full social integration.

REFERENCES

[1] T. Chouhan, A. Panse, A. K. Voona, and S. M. Sameer, "Smart glove with gesture recognition ability for the hearing and speech impaired," in *2014 IEEE Global Humanitarian Technology Conference - South Asia Satellite (GHTC-SAS)*, Sept 2014, pp. 105–110.

[2] J. H. Shin and K. S. Hong, "Keypad gloves: glove-based text input device and input method for wearable computers," *Electronics Letters*, vol. 41, no. 16, pp. 15–16, August 2005.

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

[4] L. Chen, F. Wang, H. Deng, and K. Ji, "A survey on hand gesture recognition," in *2013 International Conference on Computer Sciences and Applications*, Dec 2013, pp. 313–316.

[5] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. d. R. Millán, and D. Roggen, "The opportunity challenge: A benchmark database for on-body sensor-based activity recognition," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 2033–2042, 2013.

[6] T. A. Felipe, *Libras em contexto: curso básico: livro do estudante*. WalPrint Gráfica e Editora, 2007.

[7] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.

[8] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.