ISSN 1870-4069

Data Acquisition Software for Sign Language Recognition

Manuel Eleazar Martínez-Gutiérrez, José Rafael Rojano-Cáceres, Edgard Benítez-Guerrero, Héctor Eduardo Sánchez-Barrera

Universidad Veracruzana, Facultad de Estadística e informática, Mexico

eleazar.mg.2000@gmail.com, {rrojano, edbenitez}@uv.mx, eduarsb4@gmail.com

Abstract. The research of sign language recognition is a topic of interest in the recent years in the fields of computing vision, artificial intelligence and human-computer interaction. The sign language recognition has at least three steps such as data acquisition, data classification and results. This paper introduces a software capable of facilitating the data acquisition step with the Intel Realsense camera. This software digitize 22 points from the hand of the user in coordinates (x,y,z) and save this data in a comma separated values (CSV) file. Finally, in this paper we tested the software with 19 users recording the alphabet of the Mexican sign language, with the dataset from the software, we trained an artificial neural network to obtain 80.1192% of precision. While this percentage can be increased with a variety of techniques, the test shows that users can create the dataset using this software and that those values can be used in a classification algorithm with acceptable results.

Keywords. Sign language, data acquisition, pattern recognition.

1 Introduction

Sign language has the characteristic of being a visual language, this means that the medium of communication is the sight and the way to express a message is the use of gestures from the hands (including the form of the hands, the movement and the space where it is made), face and the combination of both [1].

The characteristics of the sign language allow the use of different types of research to try to digitize it, there are research with the context of sign language recognition, more specifically in the areas of: computer vision, artificial intelligence and human-computer interaction. In computer vision the researchers develop algorithms to digitize an image or video, more precisely to obtain data of the sign in the image, this can be done with filters [2,3]. The filters can find the edge of the hand, eliminate the background noise from the picture, find the person doing the hand sign, obtain the pixels where the hand is, etc. (this depends on which filter is used). The artificial intelligence takes the data digitized from the computer vision as a dataset, this dataset is taken to a pre-process step to clean the data for the classification algorithm. The purpose of the algorithm is to train from the dataset to be able create a classification model capable of classifying another dataset with a high success rate [4-6]. Finally, the human-computer interaction takes the challenge of incorporating the model from the artificial intelligence into a computer system, the computer system needs to design interfaces to allow sign language users to interact with the software, to do so it is needed an evaluation of the software with sign language users to measure the usability of the system. The system should be able to communicate with the user using the sign language, to perform this, the system should take the filters from computer vision to digitize the input form the user into a dataset to be classified by the classification model, the output of the model will be the input form the user.

In this work it is proposed the use of a software capable of digitize the data of the hand with the use of the Intel Realsense f200, the software can detect 22 points from the hand and saving them in a CSV file, this file can be used as a dataset to work with an algorithm from artificial intelligence.

The paper is organized as follow: Section 2 cover the process of data acquisition. Section 3 is about the software for data acquisition and is explained how the software takes the work capturing the data. Section 4 use the dataset from the software to create a model capable of classifying the signs. Section 5 brings a conclusion to this work.

2 Data Acquisition

As mentioned in the introduction the data acquisition is one step of the sign language recognition, this step consists of digitize the information from the user to make it usable to classification algorithms, the information from the user can be acquired in different ways depending on the tools used, such as visual based like cameras or wearable like the use of electromyogram, sensors in a glove, etc. Those tools have different advantages such as low-cost, hardware and software compatibility, user detection, etc., in the same way those tools have disadvantages such as occlusion, cost, trouble with the users, etc. Depending on the scope of the research the tools to obtain information from the user can change.

2.1 Visual Base Acquisition

The process of data acquisition can be seen as a technical work with images, as in the work [2] where they used a Sobel edge filter to eliminate the background noise from the image and obtain the outline of the hand as a binary image, then a neuronal network is used to generate tokens from the outline of the hand, this tokens represent the points of the shape of the hand, finally the set of tokens from the combination of the images is the dataset used in a classification algorithm. This process can start since the image acquisition as it can be seen in [3] he begins explaining about the resolution from the camera, this is important because the more pixels the image has it is needed more processing power and time to digitize the information, the environmental noises like the light, the position of the cameras and the background of the images. Once the set of images are ready, he smoothed the images with a spatial averaging filter of mask 3x3, then he uses morphological operations of dilation and erosion, next he uses a spatial difference to obtain segments of the hand, then to improve the contrast of the segment image he used the adaptive histogram based contrast enhancement, finally he uses an elliptical Fourier descriptors to create the dataset used in a neuronal network.

The process is similar from the previous works using a special camera like Kinect, the advantages of depth cameras are the ease of digitize the visual data and include it into a software, as it can be seen in [7] the data acquisition with Kinect may require the use of affine transformation to cancel the noise of the signer position and rotation. So, with that in mind, even if the depth camera can obtain the digitize data it may require a few steps to prepare the data.

2.2 Wearable Tool Data Acquisition

The wearables tools have the characteristic of being intrusive, so it can be bothersome to the user, however these tools can approach problems like occlusion or can be used to test different kind of sensors in data acquisition for sign language recognition. For example, in [8] they used the TMS porti electromyogram as a data acquisition tool, unlike with the use of a camera, this one obtains signals that needs to be extracted to be used as a dataset, in this case it is used the mean absolute value and moving variance.

As mentioned before, the use of wearable tools can be because of different reasons like the use of colored gloves to make the hand easier to detect or to evade problems like occlusion that the use of cameras present, in the work [9] they utilize a data glove with 15 sensors divided 3 per finger, with the use of this glove they can digitize the information from the user and evade the problem of occlusion because of the use of sensors the data does not change even if the hands overlap with each other.

3 Sign Language Data Acquisition Software

As seen in the previous section of this paper, the data acquisition is a step in sign language recognition that depending on the tool used it can involve the use of filters or preparations of the data, similarly the choice of the tool depends of its advantages like to evade occlusion, easier hand detection, usability, etc.

In this paper, we propose the use of a software with Intel Realsense f200 depth camera as a data acquisition tool, this camera has the advantages of digitize 22 points from each hand and is capable of facial recognition, the range of capture is between 20-120cm so it is recommended to use this tool for one user at a time, while this is a small range of capture to use in an open space for full body capture like with Kinect, it is optimal for a context when the user is seated and close with the computer.

The specifications to work with the Intel Realsense f200 camera are: 4th generation intel core processor or higher, Microsoft windows 8.1 or higher, 4gb of RAM, 4gb of hard disk space and a USB 3.0 port. The software proposed in this paper works with JAVA and the Intel Realsense SDK, the scope of the software is to capture 22 points of the hand in coordinates (x,y,z), this coordinates are saved in a CSV file with values measured in meters, while this software can capture the face and hands, in this paper we will focus in the use of only one hand.

One of the advantage of the software is the ease of acquiring the hand data from the user, this easiness is compared with the process of data acquisition from the perspective of using a web camera which involves the use of filters and transformation on the image to identify the hand and digitize it to obtain data. Another advantage of the software is the simple design so that the user is capable of digitize his information without the need of an expert to help him.

The software shows the user using the depth camera as shown in figure 1, when the software detects the hand from the user, an icon will appear in the bottom right corner of the interface and the 22 points from the hand will be drawn in colors, the red dot is the center of the palm, the blue dots are the tip of the fingers and the green dots are the rest of points that are the base of the fingers and the wrist. Simultaneously the software draws a line through the points of each finger to help the user represent the finger, this is useful when the user is making a sign and it is confusing to know if the points are in the correct position.

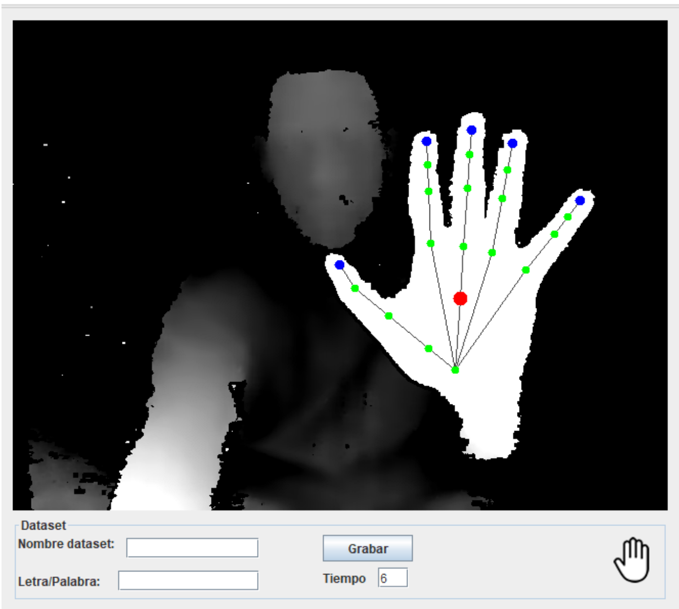

Fig. 1. Data acquisition software with Intel Realsense.

The way to use the software as shown in figure 2 is to fill the data before pressing the button with the name of the CSV file, the letter or word that the user is going to sign and the time that the software is going to record. This inputs are important to create a dataset, in this software if the name of the CSV file is the same of a previous one, the new records are going to be in the end of the previous ones, this is recommended to take control of the number of records per sign and to evade the noise when the user is changing from sign to sign, the letter or word represent the expected value in a classify algorithm so depending on the algorithm it is possible to use a number, the value of the time is because of the number of records that the software is going to write in the CSV file, the camera take 30 frame per second so if the time is in 6 seconds the software is going to write 180 records.

Finally, when the button is used, it will start recording and change its label to a timer to let the user know how much time left to finish the record, simultaneously an icon REC will appear at the bottom right corner of the interface and it will disappear when the time is up.

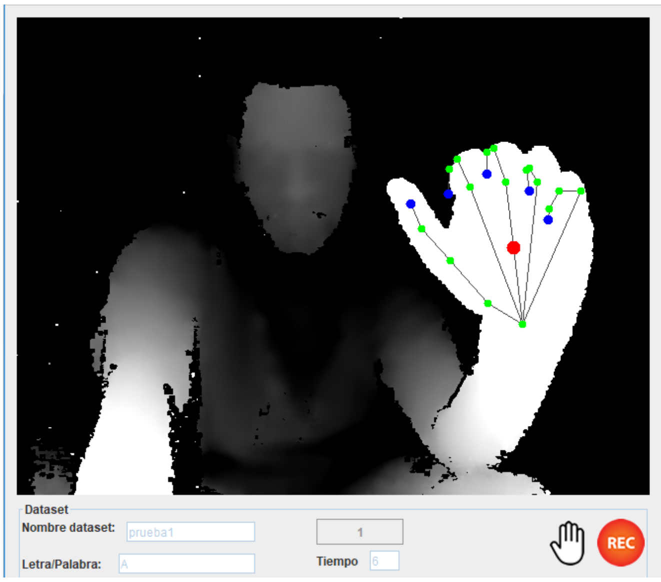

Fig. 2. Recording the letter “a” from the Mexican sign language.

4 Evaluation of the Software and the Dataset

So, to evaluate the software in this paper we recorded the alphabet of the Mexican sign language with a group of 19 people with basic knowledge of the Mexican sign language, the process was as follow:

1. Equip a place with two computers with the Realsense camera and the software.
2. Schedule the users so only two will use the software at the same time, one in each computer.
3. Prepare a consent document to use the digitize data from the user.
4. Explain the user how to use the software.
5. For the static signs, the user will sign then it will start the record until the time is up.
6. For the signs with movement, the user will repeat the sign until the time is up.

7. The record will be the alphabet twice per user.
8. Combine all the CSV files into one file.
9. Eliminate null records.

The dataset is conformed of 90,305 of records after eliminating null records, those happens when the hand was not recognized while the camera was recording. The dataset is composed of 45 variables, 44 of them are the coordinates (x,y) of the 22 points acquired from the camera and one variable representing the letter of the alphabet, with this dataset we trained an artificial neural network (ANN) multilayer perceptron and backpropagation algorithm.

To train the ANN we split the dataset in 70% to train it and 30% to test it, the structure of the ANN was of 44 neurons in the input layer, 34 neurons in the hidden layer and 27 neurons (representing each letter of the alphabet) in the output layer. To test the classification model, we used cross validation with 30% of the dataset, the results of the test were 81.1192% of correctly classified instances and kappa statistic of 0.8039 which is represented as substantial [10].

5 Conclusion

Based in the state of art we can see the process of data acquisition which involves different approaches, tools, scopes, filters and results but the purpose of them is to take the digitize information from the data acquisition to train a classifier algorithm. In this paper we proposed a software that helps in the data acquisition step with the Intel Realsense camera, this software allow the user to digitize the hand in 22 points of the hand in coordinates data with the flexibility of using one button to record and saving the data in a CSV.

Based in the test of the data presented in this paper, the precision of the ANN was above the 80% only using the data form the software, this percentage can be increased with a variety of techniques such as translation, selection of significative attributes, rotation, etc.

Finally the purpose of the software was a success since in the test, the users were capable of using it and it was possible to create a ANN with the dataset from this test, so in the future with this kind of interaction it will be possible to create applications centered in deaf people with the ease of acquiring the digitize data from the user in the acquisition phase or in the application itself.

References

1. Aldrete, M.: Grámatica de la lengua de señas mexicana. El colegio de México (2008)
2. Raheja, J., Singhal, A., Chaudhary, A.: Android based Portable Hand Sign Recognition System. arXiv Preprint arXiv:1503.03614 (2015) http://doi.org/10.15579/gcsr.vol3.ch1
3. Kishore, P.V.V., Prasad, M.V.D., Prasad, C. R., Rahul, R.: 4-Camera model for sign language recognition using elliptical fourier descriptors and ANN. In Signal Processing and Communication Engineering Systems (SPACES), 2015 International Conference on IEEE, 34–38 (2015)

4. Mohandes, M., Deriche, M., Johar, U., Ilyas, S.: A signer-independent Arabic Sign Language recognition system using face detection, geometric features, and a Hidden Markov Model. Computers and Electrical Engineering, 38(2), 422–433 (2012) http://doi.org/10.1016/j.compeleceng.2011.10.013.
5. Bedregal, B.C., Costa, A.C., Dimuro, G.P.: Fuzzy rule-based hand gesture recognition. In IFIP International Conference on Artificial Intelligence in Theory and Practice, Springer, Boston, MA., 285-294 (2006, August)
6. Islam, M.M., Siddiqua, S., Afnan, J.: Real time hand gesture recognition using different algorithms based on American sign language. In Imaging, Vision & Pattern Recognition (icIVPR), 2017 IEEE International Conference on IEEE, 1-6 (2017)
7. Kumar, P., Saini, R., Roy, P. P., Dogra, D. P.: A position and rotation invariant framework for sign language recognition (SLR) using Kinect. Multimedia Tools and Applications, 77(7), 8823–8846 (2018)
8. Amatanon, V., Chanhang, S., Naiyanetr, P., Thongpang, S.: Sign language-thai alphabet conversion based on electromyogram (EMG). In Proceedings Biomedical Engineering International Conference (BMEiCON), 2014 7th, IEEE, 1–4 (2014)
9. Bedregal, B.C., Costa, A.C., Dimuro, G.P.: Fuzzy rule-based hand gesture recognition. In Proceedings IFIP International Conference on Artificial Intelligence in Theory and Practice, Springer, Boston, MA., 285–294 (2006)
10. Abraira, V.: El índice kappa. Semergen: revista española de medicina de familia, (5), 247–249 (2001)