# Chapter 6: Testing and Evaluation

System testing:

This chapter is concerned with the system testing and evaluation . Although each developed feature was tested during evolutions, a complete test of system has to be carried at the end of development process. Therefore following sections describe how system is tested, what is tested and what the results of testing are.

The system provides functionalities under some conditions. These functionalities and conditions (non-functionalities) are system requirements that were specified in the beginning of the project. Therefore testing is carried against them in order to prove that developed system meets requirements and works as expected.

The functionalities for which the system is tested are:

Numbers recognition

Alphabet letters recognition.

Word recognition

Sentence recognition

Text to speech

The conditions under which the system is tested are:

Light conditions

Real time recognition

The system was tested from one native speakers of sing language and one non-native speakers of sign language. The testers were of different ages, body types and sizes. Different testers assure testing for different sizes and shapes of hands thus providing more insights into using the hand shape feature in systems for sign language recognition.

The system is tested for six numbers, four words and one sentence. Testers performed each signs 10 times. The non-native speakers had to be trained to perform the signs whereas non-native speakers were required just to perform the signs. The results from testing are shown in table 6-1.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Sign** | **Non-native Tester** | **native tester** |
| **1** | Number 0 | 10/10 | 9/10 |
| **2** | Number 1 | 10/10 | 7/10 |
| **3** | Number 2 | 10/10 | 10/10 |
| **4** | Number 3 | 10/10 | 10/10 |
| **5** | Number 4 | 9/10 | 8/10 |
| **6** | Number 5 | 10/10 | 10/10 |
| **7** | Word: انا | 10/10 | 10/10 |
| **8** | Word: حساب | 9/10 | 8/10 |
| **9** | Word:اريد | 9/10 | 9/10 |
| **10** | Word: فتح | 10/10 | 9/10 |
| **11** | Word: بنكي | 6/10 | 7/10 |
| **12** | Sentence: انا اريد فتح حساب بنكي | 6/10 | 8/10 |

Table 6-1: Sign for which system is tested

Graphical representation:

The results of testing are shown in following graphs. The figure 6-1 show number recognition accuracy, figure 6-2 word recognition accuracy, finally the figure 6-3 shows recognition accuracy for each component for different testers.

Figure 6-1: Number recognition accuracy

Figure 6-2: Word recognition accuracy

Figure 6-3: Component recognition accuracy for each tester

Result analysis:

The analysis of each sign language components is **as proceeds:**

**Number recognition** – from above graphs it is seen that best recognition rates are higher. From testers the best recognition rate was achieved by non-native tester.

* **Word recognition –** the recognition rate for words is about 81%.
* **Sentence recognition –** the recognition rate for one sentence from two testers is 73%.

The system is tested against non-functional requirements also and obtained results are summarized **in following section:**

**Light conditions -** the system was tested in indoor environment with lights off and on. The system worked in same way in both cases. It is well known fact that Kinect does not work in presence of sun rays and therefore it was not tested under these conditions.

* **Real time recognition -** the real time recognition was measured by measuring the system response time from the time when sign was performed until it was recognized. Measured within application (while application running) the system detects one hand shape 25 times per second. This means that the sign that represent number one is recognized every 25 time within second or for 0.41 seconds. The response time was unnoticeable for the users when they performed numbers, alphabet letters, words and sentences, since as soon it was performed it was recognized and provided as text/speech.

Evaluation:

The evaluation of first objective that is related with recognition of selected subset of signs from a dictionary of sign language is given **in following section:**

* **Number recognition** – the system is able to recognize numbers from 0 to 5 in sign language with high accuracy. However not all numbers of sign language were recognized such as numbers starting from six and up.
* **Word and sentence recognition** – the system implements a limited vocabulary of words and sentences. More specifically five words and one sentence were implemented and tested. Although recognition accuracy is higher and at acceptable levels.

The sign language incorporates non-manual features (facial expression, body posture, hand movement) that were not taken into consideration in this project. The capability for continuous recognition of sign language communication is not provided. Continues communication in sign language does not require recognition of alphabet letters, since sequential combination of words is what forms sentences that are used for communication. Sentence recognition was tested for one sentence only and followed a simple algorithm that is not scalable for inclusion of new sentences.

**Sing to speech conversation (second objective)** – is done by converting recognized signs to text first and then using Text-to-Speech engines to convert text to speech. Conversation is provided separately for numbers, alphabet letters, words and sentences since system is not capable of determining the context of communication. Although the signs were converted to Kosova’s spoken language, the employed engine for TTS support multiple languages thus allowing conversation into multiple spoken languages.

**The third objective** – translation of speech to signs, was left optional from the very beginning because the technologies and applications that enable it are available, whereas the translation of sign to speech is challenging and not yet provided. Therefore, due to the focus on investigating the Kinect device for this purpose and limited time of the project it was not implemented. Nevertheless this feature is important in systems that try to facilitate the communication between hearing impaired and normal hearing people and its integration in current system is suggested as future work. Finally the last objective - the real time performance of the application was achieved as it was presented in testing section, but application testing in more different computers would reveal its real performance.

Evaluation of Software Development Process:

The Evolutionary Software Development process that is an agile methodology proved to be better choice than traditional waterfall model. The project was considered as high risk project because within limited time it tries to implement a feature (hand shape recognition) using a framework that is under development, that is not fully tested and that does not provide documentation. Risk avoidance was done through not trying to define all requirements at the beginning and developing an evolutionary prototype for better understanding of requirements and technical difficulties.

Furthermore providing incomplete but working versions and then getting feedback, proved to be more important than documentation that is emphasized in traditionally water fall model.

However the risk of liking and following initial version of system prototype became true. During late evolutions the code complexity increased and code factoring was inevitable. It was necessary step in order to make the system more maintainable. It should be noted that code refactoring or changing internal structure without changing it external behavior should be done as soon as risks have been avoided and not at the end of the development cycle. Nevertheless the software development process paid itself since the risks were resolved in time and the software delivered required functionality.

Classification Measures

**Accuracy:** is our starting point. It is the number of correct predictions made divided by the total number of predictions made, multiplied by 100 to turn it into a percentage.

**Precision:** is the number of True Positives divided by the number of True Positives and False Positives. Put another way, it is the number of positive predictions divided by the total number of positive class values predicted. It is also called the [Positive Predictive Value](http://en.wikipedia.org/wiki/Positive_predictive_value) (PPV).

Precision can be thought of as a measure of classifiers exactness. A low precision can also indicate a large number of False Positives.

**Recall:**  is the number of True Positives divided by the number of True Positives and the number of False Negatives. Put another way it is the number of positive predictions divided by the number of positive class values in the test data. It is also called Sensitivity or the True Positive Rate.

Recall can be thought of as a measure of classifiers completeness. A low recall indicates many False Negatives.

**Training and testing examples:**

**Word:** أنا

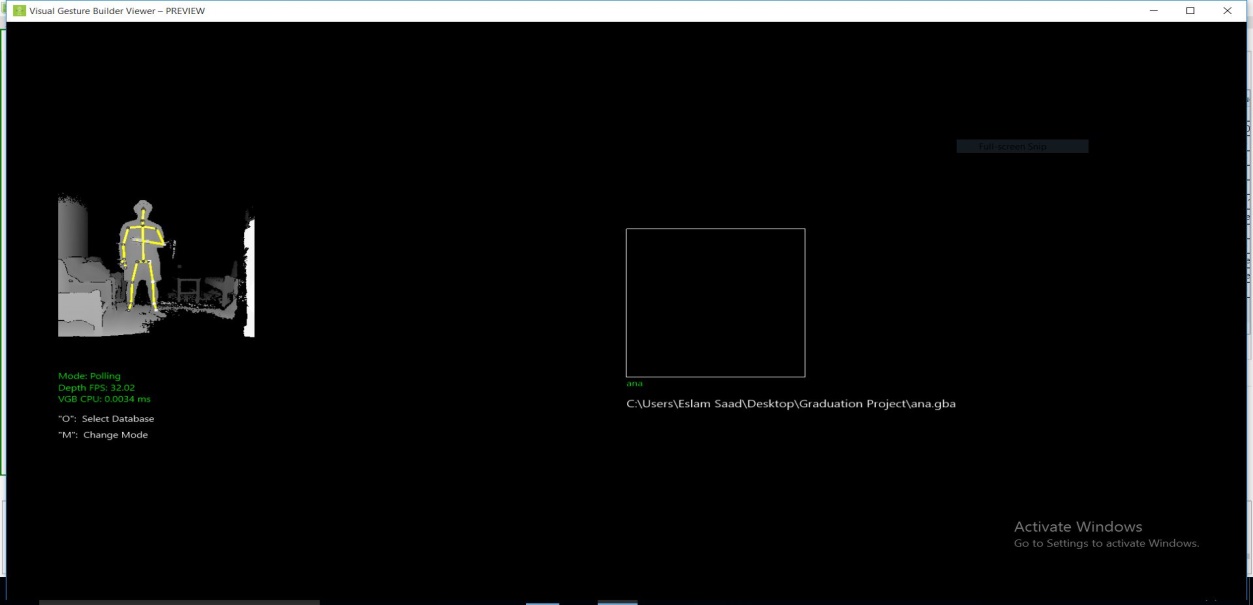
****

Figure 6-4: wrong movement

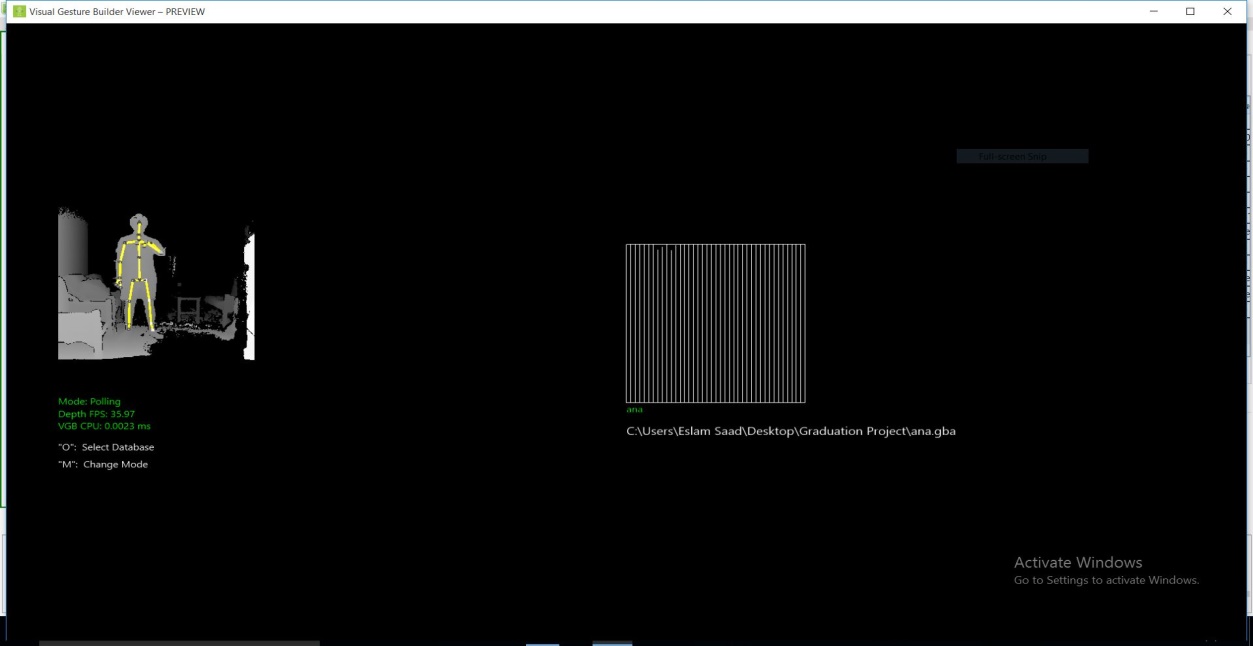


Figure 6-5: correct movement

**Word**: أريد

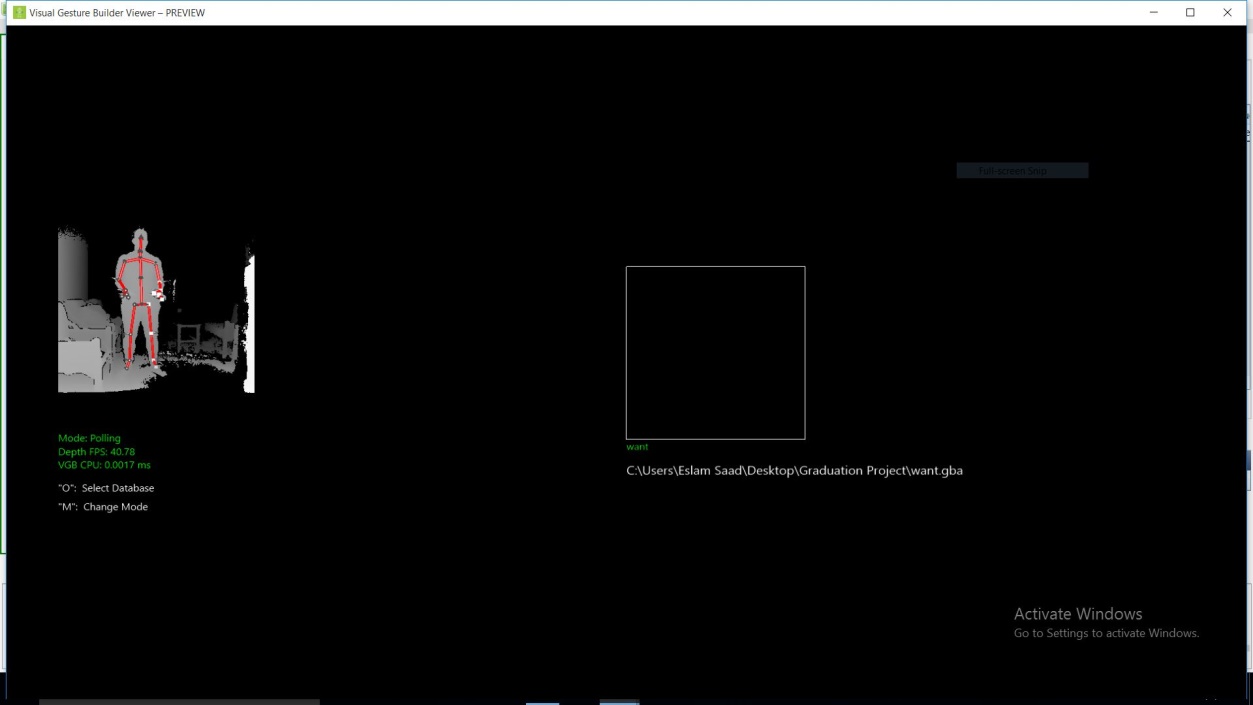


Figure 6-6: Wrong movement

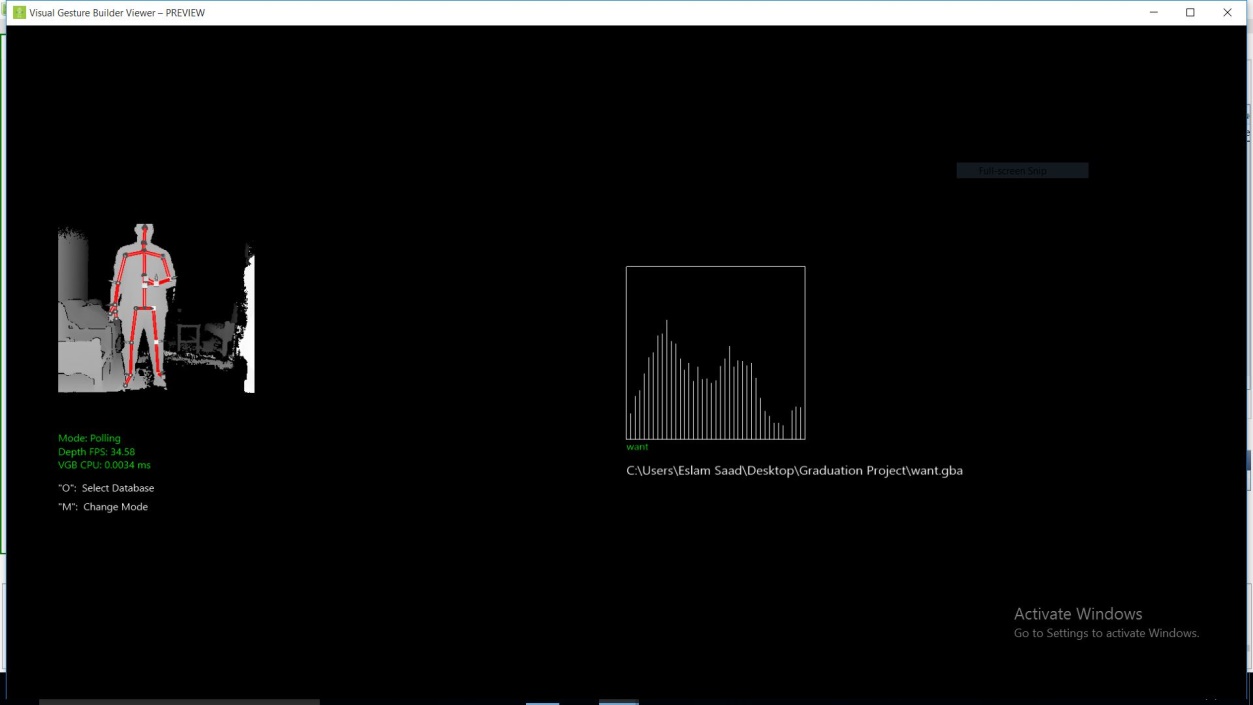


Figure 6-7: Correct movement

**Word:** فتح

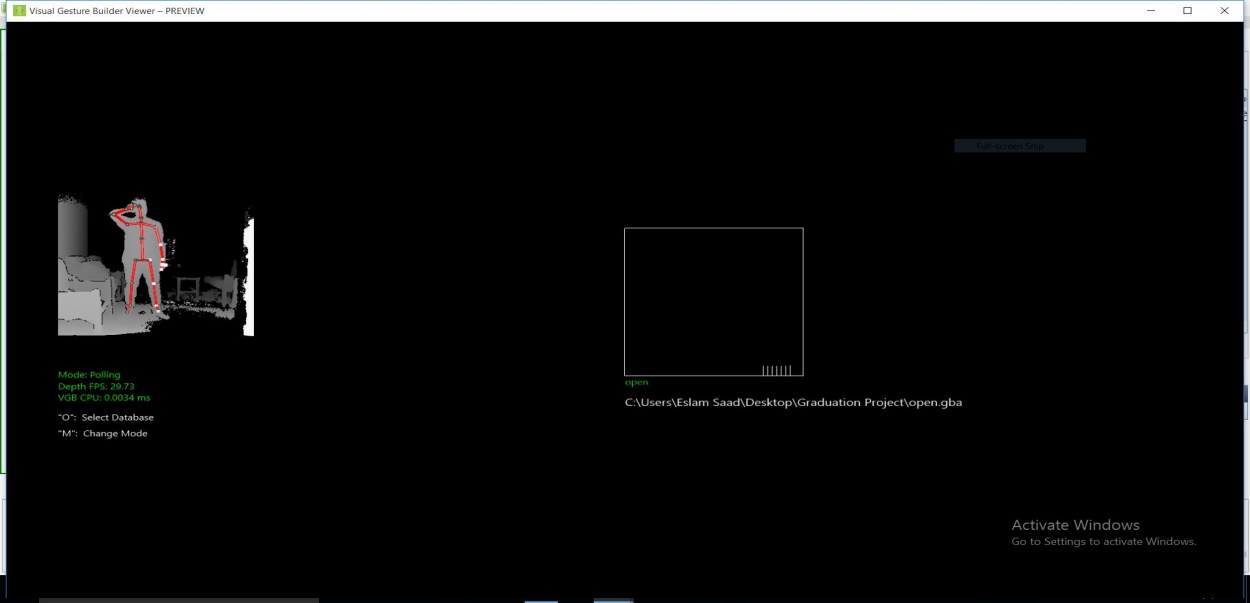
****

Figure 6-8: Wrong movement

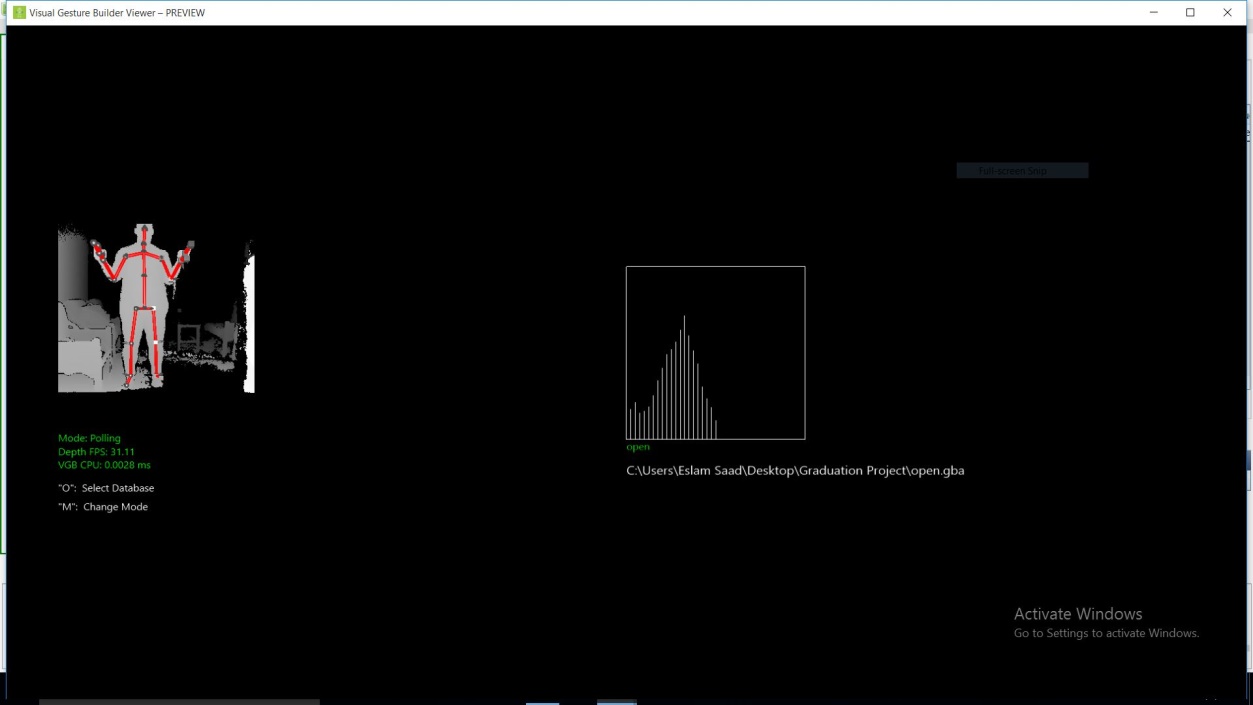
****

Figure 6-9: Correct movement

**Word:** حساب

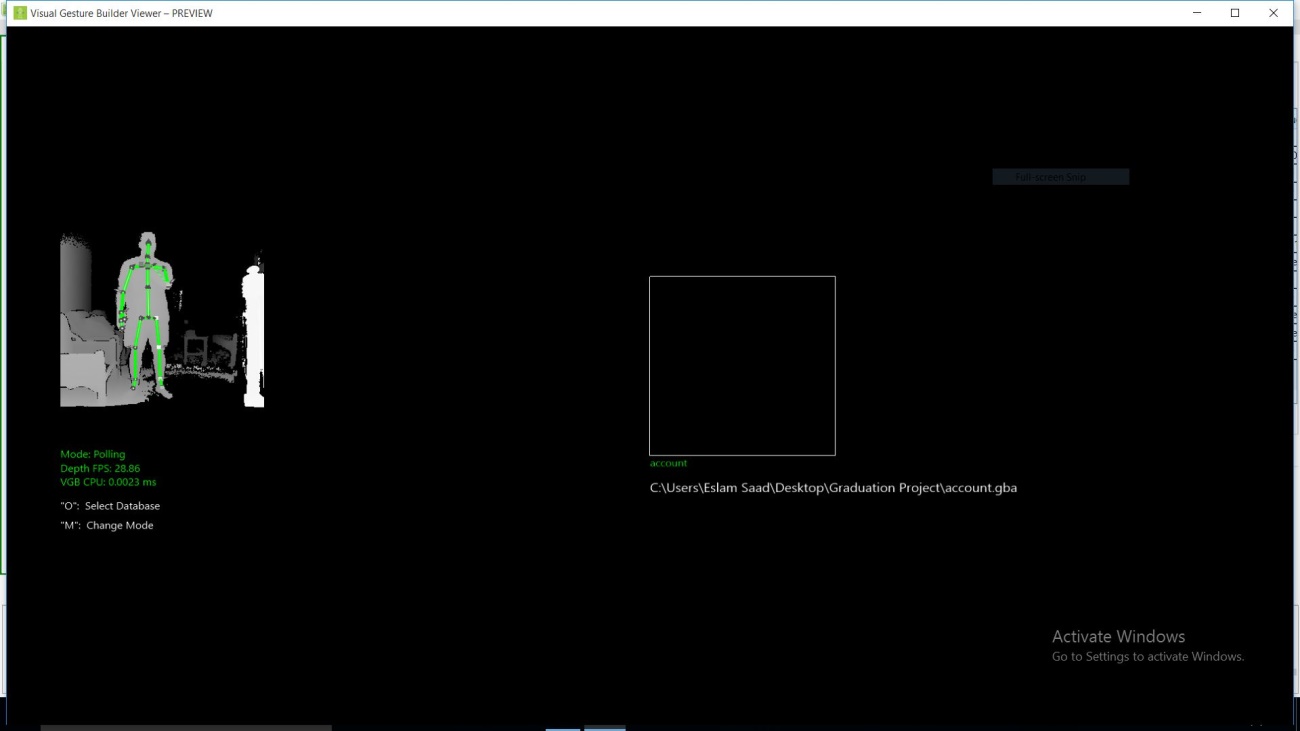
****

Figure 6-10: Wrong movement

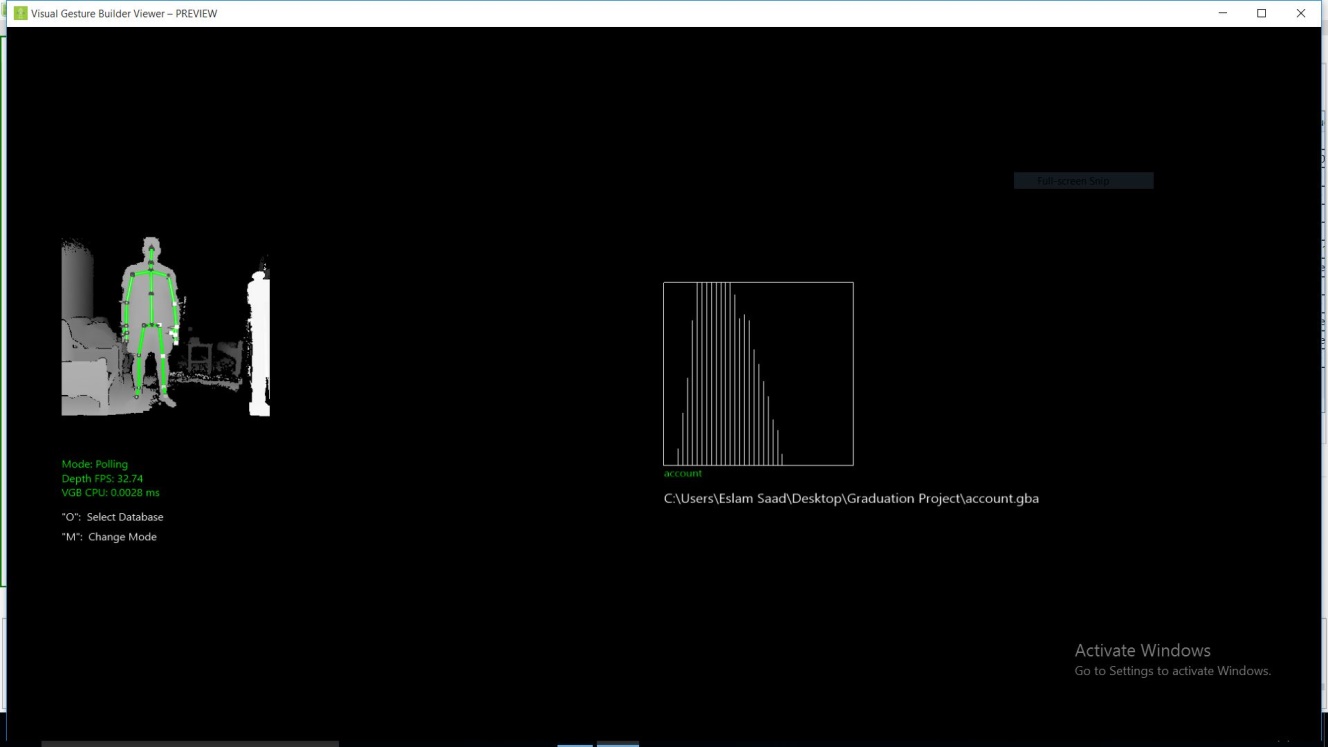
****

Figure 6-11: Correct movement

**Word:** بنكي

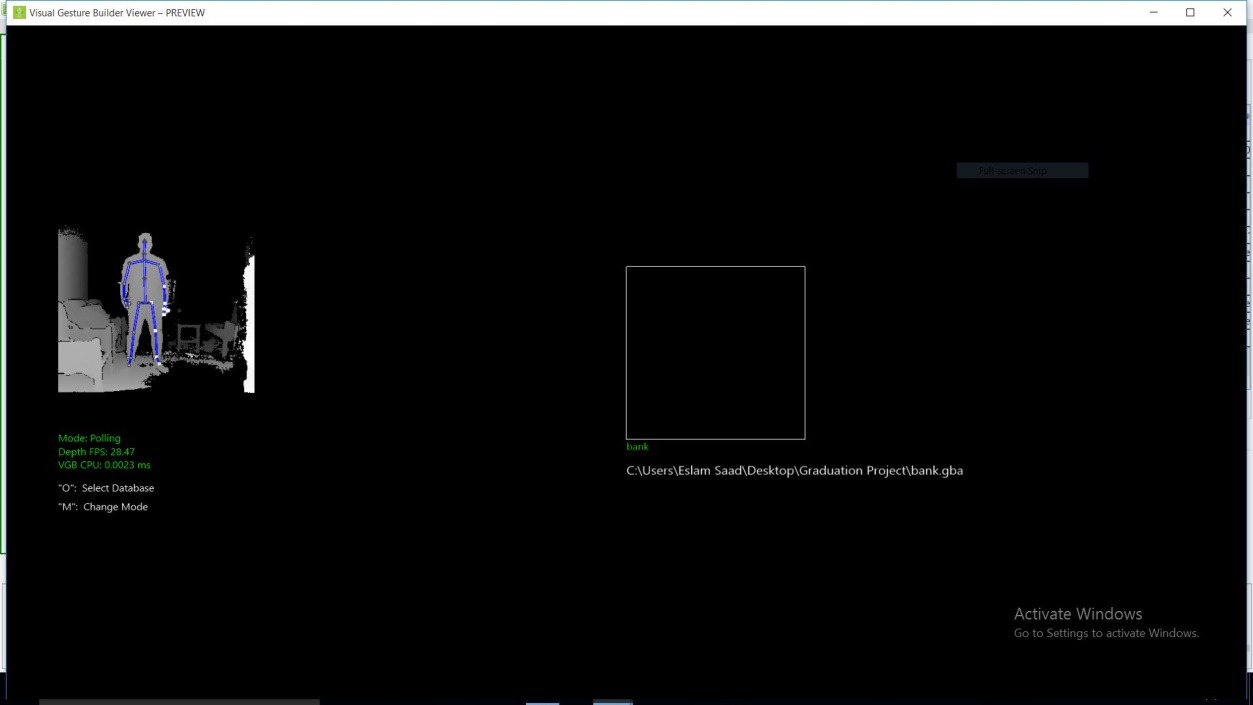


Figure 6-12: Wrong movement

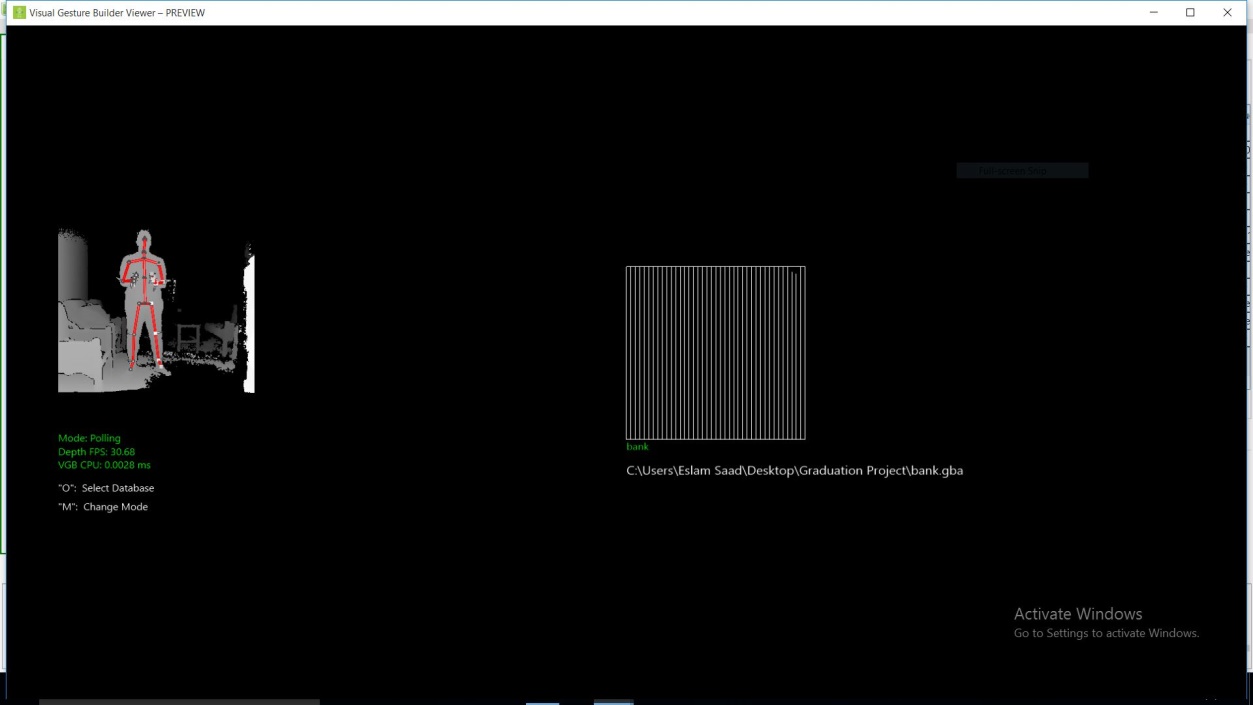


Figure 6-13: Correct movement

^ Accuracy is increase when detecting the correct sign

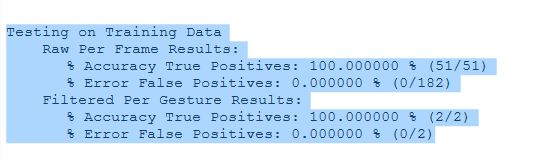
****

Figure 6 14: Accuracy

|  |  |  |  |
| --- | --- | --- | --- |
| **Accuracy** | **Precision** | **Recall** | **Confidence** |
| 96% | 94% | 96% | 95% |

Table 6-2: Classification measure