FINGER SPELLING FOR SPECIFIC WORD IN SIGN LANGUAGE USING VIOLA AND JONES METHOD

Wildan Fauzan[1], Nelly Indriani W., S.Si., M.T.[2]
[1,2] Program Studi Teknik Informatika
Fakultas Teknik dan Ilmu Komputer Universitas Komputer Indonesia
Jl. Dipatiukur No. 112-114 Bandung
E-mail : fauzan.wildan@email.unikom.ac.id[1], nelly.indriani@email.unikom.ac.id[2]

ABSTRACT

Finger Spelling is commonly used to provide information about the person's name or the name of the city to communicate with sign language. “People who cannot understand will need a system that can detect sign the finger spelling.” “There are so many method to perform signs recognitions one of them is Viola and Jones. The advantages of this method are that this method can detect object relatively faster than any other methods.”

The data used in this research in the form of sample images that are captured through the camera as much as 80 webcam image which consists of 78 image form of hand signals, and 2 the image is not a form of hand signals that are carried out in real time. The stages performed by the method of Viola and Jones. Filter with the greatest weight is placed on the process the first time with the aim to clear the area of the image instead of hands as soon as possible. Testing conducted with the environmental conditions in the room with light and outdoors with light enough, and distance from the camera.

Based on the results of the research, analysis and design, the implementation and testing phase of finger spelling with sign language using Viola and Jones methods, produces a reading accuracy of 55%. It is obtained that is still less accurate to identify a series of letters.

Keywords: Image, Finger Spelling, Sign Language, Indonesian Sign Language, Viola and Jones.

1. INTRODUCTION

Sign language is a media of communication for the Deaf people. Sign language can be a movement of the hand signals, a face or body Gestures. People generally have the ability to communicate with sign language. The problem arises when the deaf people want to communicate with normal people that are less or do not understand sign language. So it is reqiured to normal people who can understand the sign language to have a media of communication to interpreter sign language that is easily understood by most people.

Sign language essentially has a dynamic nature, because it uses the movement of hand or body gestures in exchange for speak to communicate. Gestures are a form of body language or non-verbal communication. One part of sign language is the sign language that is practiced only by using hand gestures especially change the pose of the fingers that is called sign language hand or finger spelling [1]. Finger spelling is commonly used to inform the name of the people, places, and other objects that are not yet known in sign language such as company names, brand names and other [1]. In the realm of sign language recognition research using computer, sign language differed into static and dynamic sign language based on the nature of the input system. Following similar research that aims to *make software applications using the static method of sign language* PCA (*Principal Component Analysis*) and *Haar Like Feature* [2]. From the results of the design and testing of programs, which obtained percentage accuracy read letters of 80,42%. With deficiencies in reading the letters M, N, S, T are the accuracy of reading. It cannot recognize the reading of the letters I and J due to cue the letters I and J use the dynamic movement of the hand [2].

Viola and Jones is one method for pattern recognition. Viola and Jones is a method of object detection framework with in real-time that can deliver results with a fairly high level of accuracy and with a very high speed. This method, proposed by Paul Viola and Michael Jones in 2001 [3].

By utilizing methods of viola and jones on the research of this time, allows the spelling (finger spelling) could recognize a series of letters for the name of the person or place, for example.

Based on problems that have been observed, the intent of the research was to analyze the spelling letters (finger spelling) with sign language hand using methods of Viola and Jones.

Meanwhile, the goals that must be achieved in the study is to know the accuracy of the method of Viola and Jones against spelling letters (fingerspelling), sign language hand that can translate sign language which practiced through hand gestures and implemented in the form of prototype applications

2 THE CONTENS OF RESEARCH

2.1 Problem Analysis

The analysis of a problem is the first step of an analysis system with identify problems or obstacles in the research. There needs to be an analysis of the problem formulation issue through pre-set.

To understand the sign language as a medium of communication for the deaf people is main thing to do. As a person who has no special needs, we usually rarely learn the sign language. Therefore, it is required a medium of communication interpreter of sign language into language that is easily understood by normal people.

On the research of a type previously aimed at making software applications using the static method of sign language PCA (Principal Componen Analysis) and Haar Feature, the system can read the letter but it low accuracy of reading [1]. The problems of the research of this time is how to spell (finger spelling) for special needs such as spelling the letters for the name of the person or place.

At this time the research is going to implement the method of viola and jones where this method is an object detection framework with real-time which can deliver the results with a fairly high level of accuracy and a very high speed. It is expected by the method of viola and jones the system can detect or recognize letters better.

2.2 The System analysis to be Built

An overview of the stages system will be done in Figure 1.

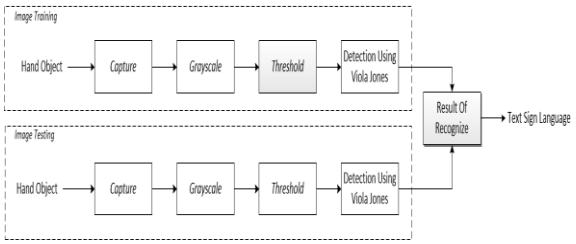


Figure 1. An Overview Of The System

In building a system of Finger Spelling took several stages. The Stages of training data will be done in Figure 2.

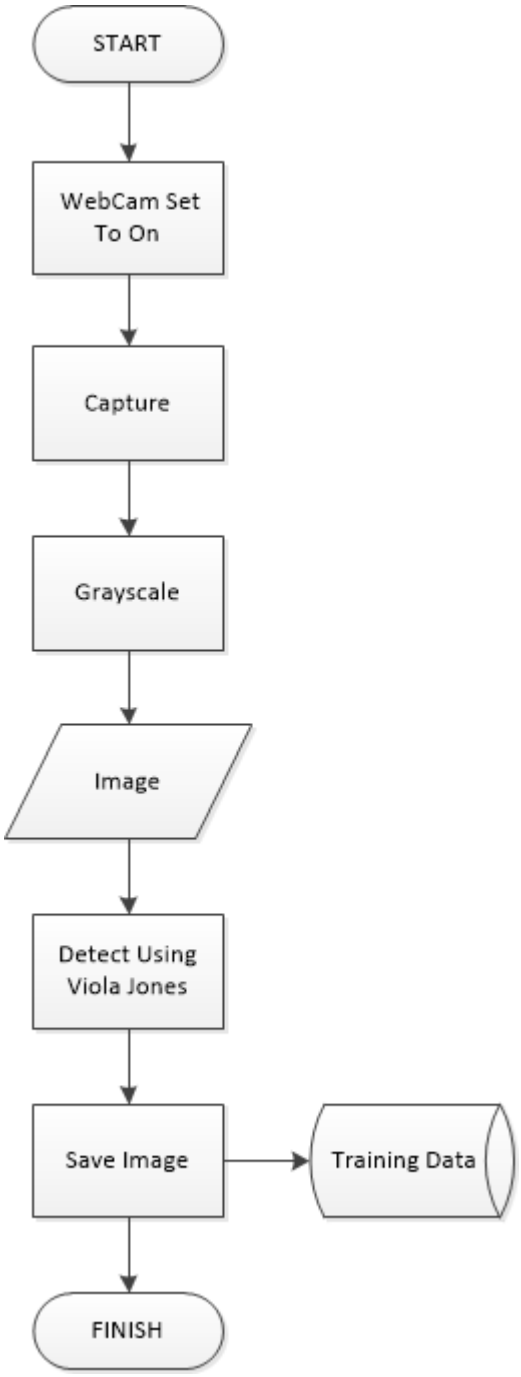


Figure 2. *Flowchart* Add Training Data

Based on Figure 2, the addition of training data on the system to built, namely the user activate the camera webcam as input data and to capture the shape of the hand signals. Then data capture results image into the praproses stage where there are stages in it cropping, grayscale and thresholding.

Next step data image results praproses log onto the stage of the process of classification. The image data is already classified in store in the storage medium.

Stages for the introduction or detection of a hand sign language finger spelling contained in Figure 3.

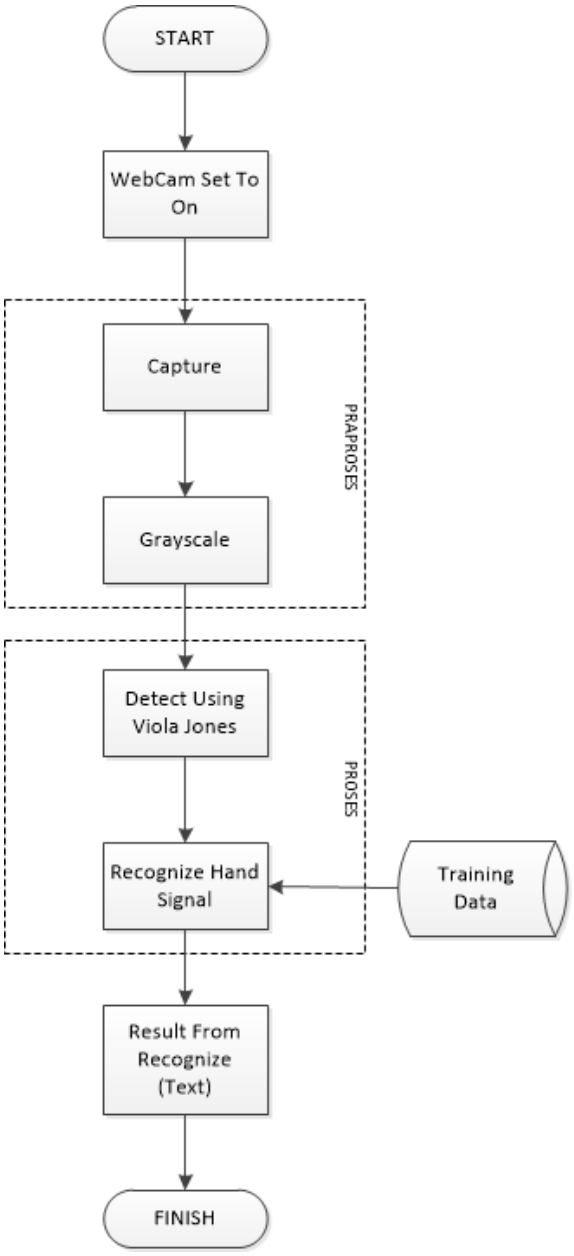


Figure 3. *Flowchart* The Process Of Recognize Of Hand Signals

Based on Figure 3, the introduction of a system of hand signals are divided into pre process and process. The main process, the user activate the camera webcam then capture (record) the hands and moves it correspond to the hand signals.

The first step is praproses, this stage prepares the data until the data is ready to be processed in the main proceedings. The processes at this stage are:

a. The user enable the webcam camera. (laptop webcam).
b. Capture the hand signals.
   Capture or record a step movement hand signals on media input into the system in real-time using webcam.
c. Preprocess
d. Before the main process, it must be cropping, grayscaling and thresholding on the input image.

The next process is the main process of identification. The following process on this stage are :

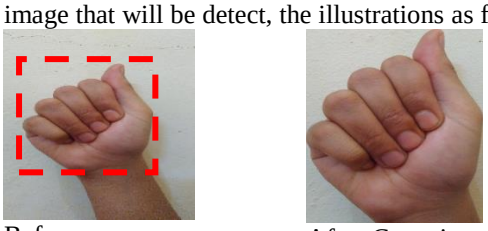
a. Detecting the hand signals using Viola and Jones method.
   The image detection process in this hand using viola-jones method.
b. The process of identification of hand signals.
   The process of identification or matching of hand signals which have been recorded via webcam with the trainer saved in the data storage media.

The last stage is the process of Identification results. At this stage the image which already known to be displayed into the text form is appropriate with the given signal. To spell the letters, make a word or sentence.

2.3 Preprocess

a. *Cropping*

This stage is to determine the part of an image that will be detect, the illustrations as follows:



Before After *Cropping*

Figure 4. *Cropping ilustration*

The result of cropping image is converted into grayscale image.

b. *Grayscale*

The stage is the conversion of grayscale images to RGB values of each pixel in the image into the degree of grayscale . In general color used black as a minimum value of 0 and white as the color of a maximum value of 255, so the value is between the color gray [4].

The basics of process grayscale pixel values from leveling perform 3 value of R, G, B by using the following equation:

Grayscale = 0.2989R + 0.5870G + 0.1140G [4]

The examples to determine the value based on the grayscale values of R, G, B on the image using the equation. E.g. for the pixel at coordinates (0.0) with a value of R = 205, 206, and G = B = 199.

Grayscale = 0.2989(205) + 0.5870(206) + 0.1140(201)

= 61.27 + 120.92 + 22.91

= 205.10

=

Those results do rounding, then result = 205. the calculations for all image pixels of RGB values, so it can obtained the image grayscale. This image of the results of the grayscale.





RGB *Grayscale*

Figure 5. *Grayscaling ilustration*

c. *Thresholding*

Thresholding phases are used to change the value of the image binary values into grayscale. Threshold value 128 taken because that value is a between 0 to 255 and the condition of the black and white image on the threshold clearly visible.

$$g(x,y) = \begin{cases} 1 & if\ f(x,y) > T \\ 0 & if\ f(x,y) \le T \end{cases} \quad [4]$$

At coordinates (0.0) with value of 205 205 > 128, so the threshold value at (0.0) coordinates become 1. Then do the same process to all coordinates in the greyscale image, so it becomes the value of the binary image. The following image threshold results in Figure 6.





Grayscale *Threshold*

Figure 6. Ilustration *Thresholding*

2.4 *Viola and Jones*

The method of analysis is conducted to find out the process of the methods used and can be enumerated into the software system. The hand detection software use viola and jones methods that Can be seen in Figure 7 below:

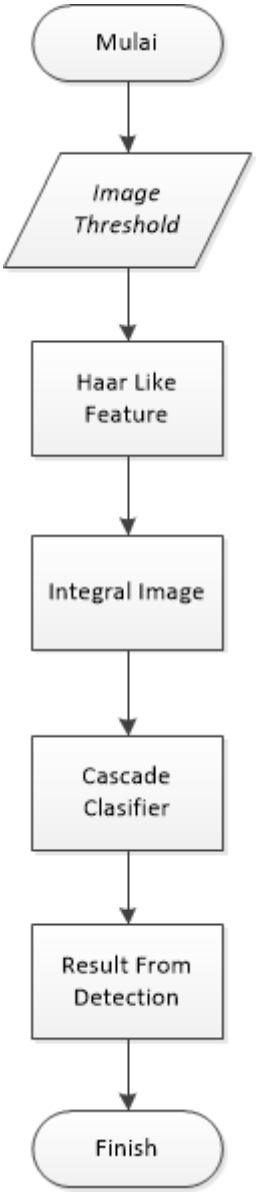


Figure 7. *Flowchart* Method

a. Integral Image

After the value of threshold image obtainable, calculate the integral image of image threshold. An Integral image is an image that is the value of the pixel it self, it is the accumulation of the values of pixels of top and left [3].

The example of calculation of the integral image of Figure 6 with coordinates (0.0) to (2.2).

1	1	1
1	1	1
1	1	1

1	1+1	1+1 +1
1+1	1+1+1 +1	1+1 +1 +1+ 1+ 1

1	2	3
2	4	6
3	6	9

1+1 +1	1+1+1 +1+1+ 1	1+1 +1 +1+ 1+ 1+1 +1 +1

Figure 8. Ilustration *Integral Image*

b. *Haar Like Feaature*

To etermine the type of haar by evaluating every feature toward the training data by using the value of the feature. The features that have the biggest limit between the form of hand signals and not a form of hand signals, is considered the best features. The features used in this study using a kind of edge feature.

The use of haar features can be seen in Figure 9.



Figure 9. Haar Like Feature

c. Calculating the value of a feature

After determining the next haar features calculating an integral feature of the image. With an integral matrix examples in table 1.

Table 1. Matrix Integral Image

1	2	3
2	4	6
3	6	9

For calculating the equations feature [4]:

$$f(x) = Sumblack\ rectangle - Sumwhite\ rectangle = (D_{black} + A_{black} - (B_{black} + C_{black})) - (D_{white} + A_{white} - (B_{white} + C_{white}))$$

Where the coordinates for the values of A, B, C and D for black or white rectangle, the illustration in the following table 2.

Table 2. Ilustrasi Coordinates *Rectangle*

A	…	B
. . .		
C		D

Then the values of the features of the image of the matrix in table 1 is:

= (0 + 3 – (0 +9)) – (6 +1 – (2 + 3)
= (3 – 9) – (7 – 5)
= -6 – 2
= -8

The value of the is -8 is the value of the difference between the value of features of black and white feature value. This value is stored and used as classification parameters object as a form of hand signals whether it is detected or not.

d. *Cascade Classifier*

The values obtained-8 is the value of the difference between the black and white features. This value is stored and used as classification parameters object as a form of hand signals is detected or not. If the result of the value from the features of the test under the threshold value (the value of the feature set of process training) it mean that it is not detected.

The plot of Cascade Classifier in Figure 10 below.

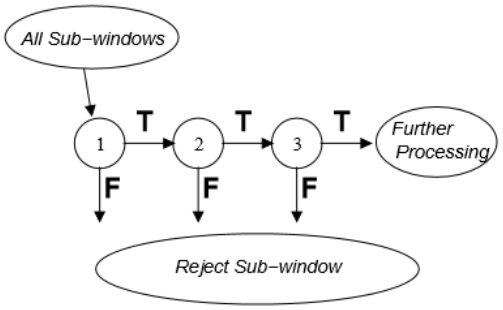


Figure 10. Cascade Classifier

On the classification of the first filter, each images will be classified using a single feature. If the value of the filter features do not compatible with the criteria, the outcome is rejected.

Move to the next window, and calculate the value of a feature again. If the obtained results correspond treshold is compatible with the criteria, then continue to the next filter stage. Next to the amount of sub window which passes the classification will also be reduced to approaching the existing image on the training data.

2.5 Testing

Testing system is the stage to perform a series of tests to try out the system that has been built with the purpose of knowing the quality of the system being built. The test aims to find out the software created meets the criteria in accordance with the objective of designing.

The testing method is the performance testing phase of the implemented methods to build the sign Introduction system . The method will be tested is a method of Viola and Jones. The method to be tested using the trainer's image data 78 and 80 of the test image data. These details can be seen in table 3:

Table 3. Details of the Training Data and test Data

No	Type Of Image Signals	The Number Of Alternative	The Number Of Test Data
1	A	3	3
2	B	3	3
3	C	3	3
4	D	3	3
5	E	3	3
3	F	3	3
7	G	3	3
8	H	3	3
9	I	3	3
10	J	3	3
11	K	3	3
12	L	3	3
13	M	3	3
14	N	3	3
15	O	3	3
13	P	3	3
17	Q	3	3
18	R	3	3
19	S	3	3
20	T	3	3
21	U	3	3
22	V	3	3
23	W	3	3
24	X	3	3
25	Y	3	3
26	Z	3	3
27	Not A Form Of Hand Signals	0	2

The accuracy of this test is determined by comparing the status of the signals which tested with the pose of the results using the method of system introduction by viola and jones. With this test, it is expected to be showing the accuracy of the method of viola and jones.

The testing was done using the Confusion Matrix which is matrix of predictions will be compared to the original classes of the data input or data test [5]. The results of positive test means that it is appropriate, whereas a negative test result it means that it is incorrect detecting system. the testing was done using the data from the test image 3 of each sign form with a distance of testing toward the webcam range from 20 cm up to 100 cm.

a. Scenario Test 1

In scenario 1 testing conducted in the room with the light conditions of the room. The test results in the Confusion Matrix can be seen in table 4:

Table 4. *Confusion Matrix* Test 1

		Predicted Class	
		Positif	Negatif
Actual Class	Positif	42	36
	Negatif	0	2

After doing the testing, then calculate the accuracy value using the equation on a Confusion Matrix.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$
$$= \frac{42+2}{42+36+2+0}$$
$$= 0{,}55$$

Then calculate the value of the precision and recall the value using the equation of the precision and recall as follow:

$$Precision = \frac{TP}{TP+FP}$$
$$= \frac{42}{42+36}$$
$$= 0{,}53$$

$$Recall = \frac{TP}{TP+FN}$$
$$= \frac{42}{42+0}$$
$$= 1$$

The testing of accuracy using the data as much as 80 image, which consists of 78 images include the sign language and the images that not include to a sign language. The classifications generated by the system as much as 42 classified sign language with precise, 36 classified inappropriately on sign language and 2 classified is not sign language with the right hand, then the proper classification number is 44 hand-sign language.

Based on the testing accuracy, the accuracy of results is obtained by 55% with a precision of 53% and recall of 100% to environmental conditions at the outside which the bright room.

b. Scenario Test 2

In scenario 2 the testing done outside the room with the light conditions are bright enough. The test results in the Confusion Matrix can be seen in table 5:

Table 5. *Confusion Matrix* Test 2

		Predicted Class	
		Positif	Negatif
Actual Class	Positif	36	42
	Negatif	0	2

After doing the testing, and then calculate the value of the fit using the equation on a Confusion Matrix.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad [5]$$

$$= \frac{36+2}{36+42+2+0}$$

$$= 0.48$$

Then calculate the value of precision and recall the value using the equation of the precision and recall on the following:

$$Precision = \frac{TP}{TP+FP} \quad [5]$$

$$= \frac{36}{36+42}$$

$$= 0{,}46$$

$$Recall = \frac{TP}{TP+FN} \quad [5]$$

$$= \frac{36}{36+0}$$

$$= 1$$

Testing the test image using data accuracy by as much as 80 image, which consists of 78 images include sign language and image 2 image that does not include a sign language. Classifications generated by the system as much as 36 classified sign language with precise, 42 classified not just in sign language and 2 classified is not sign language hand appropriately. Thus, the proper classification number is 38 hand-sign language.

Based on the testing accuracy, accuracy of results obtained by 48% with a precision of 46% and recall of 100% to environmental conditions, are outside the room with the light is bright enough.

c. Scenario Test 3

The testing conducted with examine some cues for stringing can be a Word, in this case the name of the people. Testing conducted with the environmental conditions in the room with the light being. The test results in the Confusion Matrix can be seen in table 6:

Table 6. *Confusion Matrix* Test 3

		Predicted Class	
		Positif	Negatif
Actual Class	Positif	0	5
	Negatif	0	0

After doing the testing, and then calculate the value of the fit using the equation on a Confusion Matrix.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

$$= \frac{0+0}{0+5+0+0}$$

$$= 0$$

Then calculate the value of precision and recall value using equation the precision and recall the following:

$$Precision = \frac{TP}{TP+FP}$$

$$= \frac{0}{0+5}$$

$$= 0$$

$$Recall = \frac{TP}{TP+FN}$$

$$= \frac{0}{0+0}$$

$$= 0$$

Based of the testing of accuracy results, it is obtained the accuracy of 0% with a precision of 0 and recall of 0. It is note that the system has not been able to construct the letter into a Word with accurate

d. The conclusion Of The Test

Based on the testing that has been done, it can be concluded that by using the method of Viola and Jones for the introduction of the letter on fingerspelling obtained the accuracy of 48% with a precision of 46% and recall of 100%, for conditions outside the room with light enough, and the testing distance towards the webcam range from 20 cm up to 100 cm. As for conditions in the room with enough light , and the distance testing towards webcam range from 20 cm up to 100 cm , get the accuracy of 55% with a precision of 53% and recall of 100%. The system also has yet to be stringing letter into a Word with accuracy. On this testing it also can be concluded that the system can recognize the shape of a similar gesture as the same sign. And cannot recognize some letters due to the factor the training data and environmental such as light and background.

3 CLOSING

Based on the results of the research, it can be drawn that fingerspelling using hand sign language with method of Viola and Jones are produce the accuracy testing of 55% with the environmental conditions in the room with the medium light. However there are a few less hands sign language can be recognized or identified as another hand sign language. And yet be stringing letters accurately.

Based on the results of the research and testing can be given suggestions for future research in the form of the addition of methods to be able to change the contrast of the image captured to be identified clearly and well.

REFERENCES

[1] C. Padden, Learning to fingerspell twice : Young signing childern's acquisition of fingerspelling, New York: Oxford University Press, 2006.

[2] A. Mardiyani, M. Purnomo dan I. Purnama, "Pengenalan Bahasa Isyarat Menggunakan Metode PCA dan Haar Like Future," *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*, pp. 1-7, 2011.

[3] P. Viola dan M. Jones, "RAPID Object Detection Using a Boosted Cascade of Simple Features," *Accepted Conference On Computer Vision And Pattern Recognition*, pp. 1-9, 2001.

[4] U. Ahmad, Pengolahan Citra Digital dan Teknik Pemrogramannya, Yogyakarta: Graha Ilmu, 2005.

[5] M. E. Khan, "Different Approaches To Black Box Testing Technique For Finding Errors," *International Kournal of Software Engineering & Applications (IJSEA)*, vol. 2, 2011.