Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 59 (2015) 550 – 558

Procedia Computer Science

International Conference on Computer Science and Computational Intelligence (ICCSCI 2015)

The Implementation of Hand Detection and Recognition to Help Presentation Processes

Rhio Sutoyo[a], Bellinda Prayoga[a], Fifilia[a], Dewi Suryani[a], Muhsin Shodiq[a]*

[a]*Bina Nusantara University, Jl. K. H. Syahdan No. 9, Kemanggisan / Palmerah, Indonesia*

Abstract

This research was conducted with the aim of utilizing the human hands as an object to operate computers. It is intended to support and use technologies especially in the field of education for helping teachers to do presentation with ease. The program is developed by using a science field of computer vision, as well as additional libraries which are: FLTK, OpenGL, and OpenCV. In order to use this program, the presenters need to have a webcam and a projector. The webcam will be used to recognize the shape and the pattern of the presenters' hands. Furthermore, the program will send signals to the computers based on the recognized pattern. The result of this research is a program that can improve the teaching process of teaching/presentation.

Peer-review under responsibility of organizing committee of the International Conference on Computer Science and Computational Intelligence (ICCSCI 2015)

Keywords: Presentation, Hand Detection, Hand Recognition, Computer Vision

* Corresponding author. Tel.: +62-838-936-33550; fax: +62-530-0244
E-mail address: rsutoyo@binus.edu

Peer-review under responsibility of organizing committee of the International Conference on Computer Science and Computational Intelligence (ICCSCI 2015)
doi:10.1016/j.procs.2015.07.539

1. Introduction

Presentation process is one of the activities that are often done by humans (e.g. lecturers, employees, students, etc.). Based on our observation in Bina Nusantara (BINUS) University, presenters or lecturers generally teach using wired mouse that have been provided by BINUS as tools to control the presentation slides on the computer. Moreover, there are some of them who use other tools such as remote pointer and wireless mouse. For the presenters, the usage of wired mouse to control the presentation slides on the computer is not quite effective in term of time. The reason for this is because the presenters need to move from the computer to the projection screen or center stage during their presentations to control the slides. In another case, presenters who use a remote pointer or wireless mouse can control their slides easily. However, the need to always bring their personal devices for the presentations.

Based on the given problem of controlling computer for the presentation purposes, we have an idea to create a tools that is able to control computer by using presenters' hands and finger shapes. With this solution, hands will be used to substitute remote pointer and wireless mouse to control the slides. The application, named "Touch Anywhere", use computer vision to detect hands and to be able to recognize hand shapes of the presenters. Computer vision is the knowledge of teaching computers to recognize real-world objects[1].

In order to handle the problem of controlling computer, the idea to create a computer controller application only with presenters' hands and fingers shape as an alternative solution occurs. Based on observation that has been done, the hypothesis is obtained as follows:

a. The number of data samples that will be trained according to the needs, will greatly affect the results of the hand detection.
b. The light intensity when doing hand detection greatly affects the result of detection.
c. The size of data samples that are trained affects the result of object detection.
d. The incorporation of Convex Hull and Haar-Training algorithms can detect hand patterns correctly.
e. The distance between the object with web camera and background should correspond at the time of detection so that the object can be recognized well.

The main purposes of this application are:

a. Build the application to assist in the presentation.
b. Implement the hand patterns in controlling the computer applications, such as mouse movement, click, and double click.
c. Implement the hand patterns for controlling the presentation slides in power point, such as start slide show, next slide, previous slide, and end slide show.

The benefits of this application is to help the user operating computer applications during the presentation easier so that the presenter or lecturer does not need to always move from the projection screen to the computer and does not need to use other presentation tools. For example: to open, next, or previous slide simply by using the pattern of hand, without having to use the other tools or by moving from the computer to the projection screen.

Based on our literature reviews, there are also other researchers that have explored this topic. For instance, the research by Ramadijanti in his "Tracking Jari dengan Haar Cascade dan Filter Kalman pada Virtual Keyboard" discusses how to use web camera to identify finger shape. The output of that process will be used to become an event for type in the virtual keyboard. The result of this study concluded that the detection of an object depends on the data training process. A similar research also has been done by Hojoon[3] in his paper "A Method for Controlling Mouse Movement using a Real-Time Camera". In this research, he develops the method to control the mouse cursor with a camera by using hand gesture. Hojoon said in his paper that it is difficult to achieve stable result with the difference of light situation and skin color of humans. Furthermore, Young Jae Lee and Dae Ho Lee[6] in their paper "Research on Detecting Face and Hands for Motion-Based Game Using Web Camera", discussed the use of face and hands detection in the game production area. Although their implementation are quite different from our work, it gives the insight about hand location and movement detection.

In this paper, the first chapter discusses matters concerning the background, formulation of the problem, hypothesis, scope, objectives, and benefits. In the second chapter, we will discusses the framework, the methodology used in the research, applications design, and system design. The third chapter will discusses the test algorithm used with other algorithms, analysis of the experimental results, the user evaluation, and systems evaluation. The theoretical basis used to support this research and comparison with other studies will be discussed in chapter four. The final chapter will discusses the conclusions of this study and suggestions for further development.

2. Related Works

2.1 Tracking Jari dengan Haar Cascade dan Filter Kalman pada Virtual Keyboard[2]

Hand detection and hand gesture of using web camera has been widely applied in the world of technology as a medium to operate computers, such as: move mouse cursors, operate applications, and other simple actions. In this work, Ramadijanti and team using the help of a web camera to see the objects, the haar-training to detect objects, and Kalman Filter to predict position of the fingers. In this study there are some major steps in object detection, which are: 2500 negative and 2400 positive data as the training data, cascade conversion into xml, and object detection. This research concludes that the success of object detection process is depends on the training of positive and negative data. Moreover, the number of data also affect the accuracy when detecting objects. Other factors (i.e. light intensity, distance, and object movement) also affect the accuracy.

2.2 Dynamic Training of Hand Gesture Recognition System[4]

In addition to the research of Ramadijant, similar studies also has been done by Licsar Attila and Tamas Sziranyi from the University of Veszprem and Analogical and Neural Computing Laboratory, entitled "Dynamic Training of Hand Gesture Recognition System". This study uses hands to control computer with the help of projector camera. This research concludes that dynamic data training will enhance the accuracy of the detection. The gesture detection can also be trained on a limited number of training and also training "the supervised training system" also can correct the error in the detection of gesture.

2.3. Computer Music Controller Based on Hand Gestures Recognition through Webcam[5]

The study, titled "Computer Music Controller Based on Hand Gestures Recognition through Webcams", is a study that uses a web camera as a tool in controlling the computer. The research conducted by Junhao and colleagues is focused on controlling the music player. The results from this study is the application of several algorithms such as Convex Hull in recognizing patterns of hand, as the introduction of contour.

In this work, we will focus only on controlling the presentation slide as and basic control of computer mouse. Moreover, we will use web camera as the “eyes” for the computer and Haar-Training and Convex Hull algorithms as the methods to capture hands and fingers.

3. System Overview

In order to build this application, we have formulated a research methodology. This methodology is used to explain the steps of this research from the start to the beginning. The output generated from this work is an application that is able to detect the presenters’ hands and recognize their hands’ patterns. Moreover, those patterns will be used by the application to determine what action will be made in the presentation process. For example, hand shape that shows number two is used for “next slide action” and hand shape that shows number three is used for “previous slide action”. This research methodology could be seen in the Fig. 1 below:

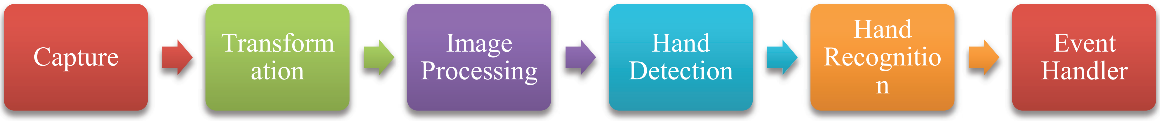

Fig. 1. Research Methodology

Every step has a different function. They are performed sequentially where the outputs from a particular step will be used in the next step. Below are the explanation of the steps:

1. Capture
   At this stage, the application will capture the projection screen as the background of the object. The process at this stage just show display captured by the web camera.
2. Transformation
   In this stage, projection screen captured by web camera will be transformed by the file that stores the coordinates taken at the calibration stage. The calibration stage is the stage of making the coordinates of the projection scre en captured by the web camera to be used as a scope limitation that will be projected on the screen.
3. Image Processing
   At this stage, the display / image captured by the web camera will be processed through some kind of process in image processing, and it will undergo stages of the crop, gray scale, saturate, and the threshold image.
4. Hand Detection
   Hand detection is the most important stage. At this stage, the hand object captured by the web camera will be recognized as the hand that will be further processed to be used as an input to operate the computer. In this phase the Haar-Training algorithm that is used. Haar-Training algorithm requires the data to be used as samples in the training process in order to detect objects well. The required data are the positive and negative data. The positive data are shown in the Fig. 2 where hands are clearly seen and perfectly cropped. Furthermore, the negative data are shown in the Fig. 3 where hands are not clear and there is no objects in the captured images.

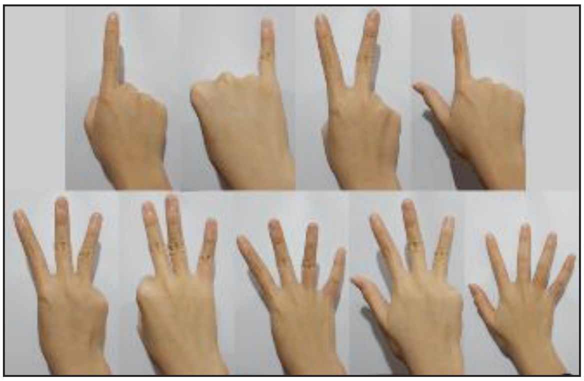

Fig. 2. Positive Image Example

Fig. 3. Negative Image Example

Once the positive and negative data is obtained, this image data can be processed first through the resize or change of image size according to the needs of the crop or cutting along with the amount of area to be trained for the adjustment of the data on the application. Coordinates taking is done after the positive data have been collected. The taken coordinate of positive image will be chosen and selected in accordance with the display position of the hand which is the object. Coordinate data is stored in a file which will be trained to be cascade before it is converted to XML. Then there will be stages of sample preparation. This stage is the process of making a sample of some of the images that have been collected. After all phases are completed it will be followed by a phase of training the data, at this stage, the sample will be processed in a pixel size of A * B, where A is the width of the image and B is the height of the image. The final stage is the result of the training that has been done will be converted into an XML file.

5. Hand Recognition

After going through the hand detection stage then in the next stage, there will be the recognition of a pattern or shape of a hand. At this stage, the hand pattern recognition is determined based on the number of fingers of the hand which is detected which would then be used as an input in this application. In this phase, it was using the Convex Hull algorithm.

6. Event Handler

This step is the final stage in building this application. At this step all the processes prior will be implemented to be used as an input in operating a computer.

4. Result and Discussions

Based on researches and experiments that have been conducted using Haar-Training algorithm in object detection that uses XML method as much as 360 positive and 174 negative images, which have been resized and crop to a size of 600 x 600 pixels, the experimental results are obtained as following:

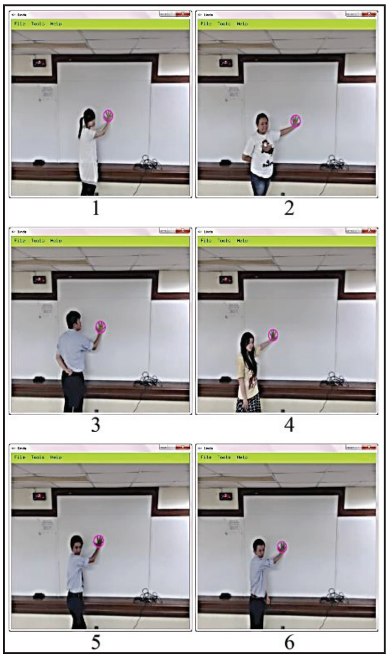

Fig. 4. Hand Detection

Table 1. Hand Detection

Sample	True	False
1	1	0
2	1	0
3	1	0
4	1	0
5	1	0
6	1	0

The result of the hand detection is shown in Fig. 4 and summarized in the Table 1. It can be seen that the results of the experiment resulted in as many as 10 correct results or percentage value is 100% and the false hand detection is 0 or the value of the percentage is 0%. The detected hand in the image is marked with a purple circle by the system. Thus, it can be concluded that the amount of positive and negative data used during the training process influence the detection result. Based on the experimental results of hand detection and tracking as well as references from various sources on the previous discussion, it was found that the Haar-Training algorithm detects the hand properly.

Here is the result of merging the Haar- Training algorithm for hand detection with the Convex Hull algorithm for hand pattern recognition:

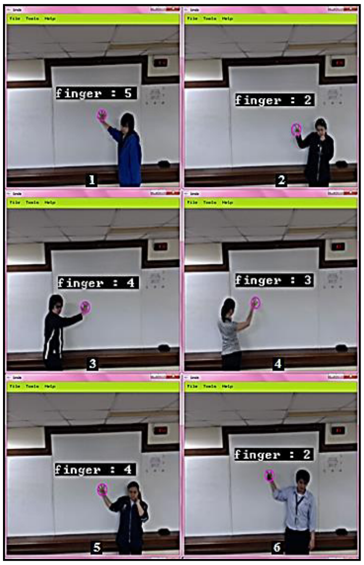

Fig. 5. Finger Detection

Based on the above experiments it can be concluded that the merging Haar-Training with Convex Hull algorithm can detect and recognize patterns well. In addition to hand pattern detection and recognition experiments, the experiments on three different types of distance between the projection screen and webcam is also done as follows:

a. Distance of 1.70 meters
   In Fig. 6, it appears that the results of detection at a distance of 1.70 meters between the webcam with projection screen are accurate. With this short distance, the camera is able to capture adequate image for the system to be analyzed. Hence, the system could easily marked the user's hand and show it by purple circle.

b. Distance of 2.85 meters
   In Fig. 7, it appears that the results of detection at a distance of 2.85 meters between the webcam with projection screen are accurate. Although the image captured by the camera is not clear because of the medium distance, the system is still able to detect user's hand.

c. Distance of 5.92 meters
   In Fig. 8, it appears that the results of detection at a distance of 5.92 meters between the webcam with projection screen are inaccurate. Lastly, we also try to put the camera in far distance. Unfortunately, the system could not detect the user's hand because the image captured by the camera is not good enough.

Fig. 6. The results of experiments at a distance of 1.70 meters between the projection screen and webcam

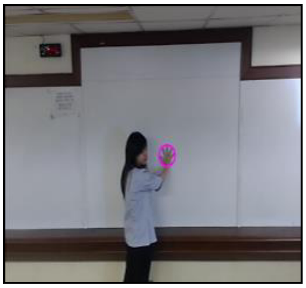

Fig. 7. The results of experiments at a distance of 2.85 meters between the projection screen and webcam

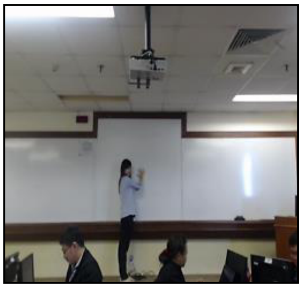

Fig. 8. The results of experiments at a distance of 5.92 meters between the projection screen and webcam

d. Conclusion on Distances

Based on experiments that have been carried out, it can be concluded that at a distance of 5.92 meters between the projection screen and webcam produces inaccurate detection whereas, at a distance of 2.85 meters and 1.70 meters yield accurate detection.

The experiments of this application also conducted with 3 types of different light intensity, which are:

a. The light intensity value from default settings

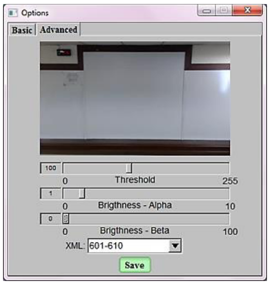

Fig. 9. Display menu options advanced tab settings to the default value

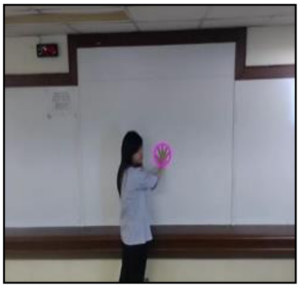

Fig. 10. Results of experiments with indoor light intensity using the default value settings

Fig. 9 and Fig. 10 shows that the detection results with light intensity using the default settings values are accurate. The hand of the presenter is detected by the camera.

b. The intensity of light with brightness-alpha values by 5

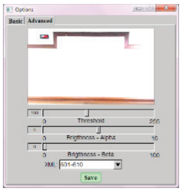

Fig. 11. Display menu options advanced tab brightness-alpha value by 5

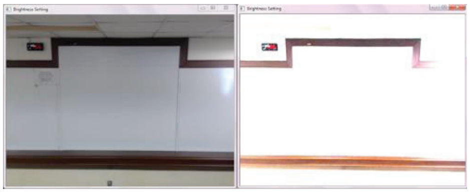

Fig. 12. The difference of light intensity due to changes in the value of 5 brightness-alpha of the default settings

Fig. 13. The results of the experiment with the room light intensity brightness-alpha value is converted to 5

In this step, we tried to change the intensity of light with brightness-alpha value by 5. The setting is shown in the Fig. 11. The result of this changes are shown in the Fig. 12 and Fig. 13. As we can see, the result are inaccurate where the system is failed to detect the user's hand.

c. The intensity of light with brightness-beta values of 50

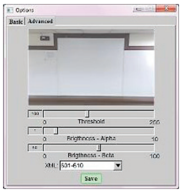

Fig. 14. Display options advanced tab menu with brightness-beta values of 50

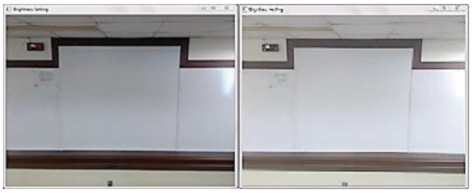

Fig. 15. The difference of light intensity due to changes in brightness-beta value of 50 from the default settings

Fig. 16. The results of the experiment with the room light intensity brightness-beta value is converted to 50

In this step, we tried to change the intensity of light with brightness-beta values of 50. The results of this change can be seen in the Fig. 15. And Fig. 16 and the setting of the light can be seen in the Fig. 14. Although the light's changed, the detection result of the system is still accurate.

d. Conclusion on Lights
Based on experiments that have been carried out, it can be concluded that the intensity of the ambient light using default values and brightness-beta values which were converted into 50 produces accurate detection. Meanwhile, with the change of brightness-alpha values into 5 will produce inaccurate detection.

5. Conclusions and Future Works

In this section it reviews the conclusions of several studies in building the applications Linda obtained from experiments conducted in this research as following:

1. In this research, the number of data samples trained should have a balance between the amount of positive and negative samples.
2. The accuracy rate can reach 100% in detecting and recognizing objects depends on the threshold value setting, brightness (alpha and beta) to the existing light intensity.
3. In this research, the size of trained sample affects the levels of accuracy when conducting the object detection. Hand detection accuracy rate reaches its peak with sample size of 600 x 600 pixels and 30 x 30 pixels during the training process.

4. Based on the study and research conducted in object detection, Haar-Training algorithm is better than the Convex Hull algorithm. Therefore, the Convex Hull algorithm for pattern recognition needs to be merged with the Haar-Training algorithm for hand detection.
5. Web camera distance in this research affects the level of accuracy in object detection.

References

1. Fairhurst, M. C. (1988). Computer Vision for Robotic Systems. An Introduction. Hertfordshire: Prentice Hall, Inc.
2. Ramadijanti, N., Setiawardhana, & Alhaqqi, R. M. (2013). Tracking Jari dengan Haar Cascade dan Filter Kalman pada Virtual Keyboard. Inovtek, 3, 1-9.
3. Park, H. (2010). A Method for Controlling Mouse Movement using a Real-Time Camera. 1-10.
4. Licsár, A., & Szirányi, T. (2004). Dynamic Training of Hand Gesture Recognition System. 17th International Conference on Pattern Recognition, 4, 1-4.
5. Jiang, J., Ma, J., & Jin, Y. (2012). Computer Music Controller Based on Hand Gestures Recognition Through Web-cam. EE368 - Digital Image Processing, Stanford University.
6. Young Jae Lee; Dae Ho Lee, "Research on Detecting Face and Hands for Motion-Based Game Using Web Camera," Security Technology, 2008. SECTECH '08. International Conference on , vol., no., pp.7,12, 13-15 Dec. 2008