**VOICE RECOGNITION AND IMAGE PROCESSING ADD-ON FOR EASIER NAVIG­ATION OF SLIDES IN MICROSOFT POWERPOINT**

**Arvin Kent Jacob**

Institute of Computer Science

University of the Philippines, Los Banos, Laguna

Asjacob1@up.edu.ph

**ABSTRACT** In ICS, it is common for professors of large classes to present their lecture with no SAs in front of the computer that’s why most of them uses the remote control to navigate the slides. However, there are a lot of times where the remote control does not work and this is problematic because the professor will move back and forth from the stage to the computer just to change slides and the time that should be used for conducting the lecture is wasted from walking which, in continues succession, is also annoying or awkward to the students. This study aims to create an add-on for PowerPoint presentation to navigate the slides using voice recognition and image processing. The voice recognition will give the user the ability to navigate the slides while image processing will give the user the ability to control the mouse even from afar without using the remote. The study will be evaluated based on the accuracy of how the program performs each commands. Several tests will be made to determine the number of times the program fails to do the desired command.

**INTRODUCTION**

PowerPoint Presentations are mostly used by professors of large classes in ICS to provide visual aids and to simulate the algorithm or code. However PowerPoint Presentations require the professor or a remote control to navigate the slides where in some cases the remote control is not working. Moving from one slide to another is very hard especially if the professor needs to write something in the board.

Most professors does not stay in front of the computer because they need to write codes and algorithms or solve problems in the whiteboard. Some of them want to stand in front to see what the students are doing and look for those who are not paying attention. A few prefer to walk to where the students are seating so that if any of them have any questions the professor can quickly come to them. However, because there are times the remote control is not working. The professor will move back and forth from the stage to the computer just to change slides. Because of this, the time that should be used for conducting the lecture is wasted from walking. This also makes the professor preoccupied which, in continues succession, is annoying or awkward to the students.

This problem can be solved using voice recognition and image processing to navigate the slides easier. The voice recognition will give the user the ability to navigate the slides while image processing will give the user the ability to control the mouse without using the remote. Using C# as the language, the add-on will be coded in visual studio because it has the function to create a PowerPoint add-on. Julius will be used for voice recognition while opencv will be used for image processing. Professors of ICS will greatly benefit from this add-on especially those who are writing things in the whiteboard. Since they only need to speak the microphone to move the slides, it gives them more time to write or to discuss the topic. This will also give additional use of voice recognition and image processing.

**RELATED WORK**

A lot of applications have the ability to control PowerPoint Presentation most of these involves using a wireless remote control. There are also studies that uses voice recognition and image processing.

Fourney, Terry and Mann (2010) research about gesture-controlled presentation [5]. In the study of Cuccurullo, et al (2012), it uses Kinect™ to navigate the slides of Microsoft PowerPoint [4]. Timmins (2004) uses the Tablet as a wireless remote to control the slides [11].

There are a lot of studies involves voice recognition and image processing. Most voice recognition involves controlling a gadget or an appliance using voice while most image processing involves controlling the mouse.

In the study of Shrivastava, Singhal, Das, and Nair (2013) it involves controlling the movement of one to three robots simultaneously using voice recognition [10]. Mustaquim (2011) uses voice recognition in controlling objects within a game [7]. In image processing, the study of Park (N.D.) [8] and Bhadrecha (2015) [1] both focuses on controlling the mouse using different gestures of the hand.

The main difference of this study and other studies is that this involves using both voice recognition and image processing in navigating PowerPoint slides while others uses only voice recognition or only image processing. Voice recognition gives the ability to navigate the slides using voice even if you are far away from the computer while image processing gives the user additional functions like clicking a button, playing a video or going to a link in the slide.

**METHODOLOGY**

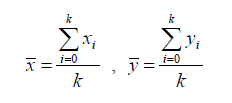
The application will be coded in Visual Studio. The PowerPoint add-in will be added by expanding Visual C# and then Office/SharePoint from the menu. Under these, Office Add-ins node will be selected. In the list of project templates, PowerPoint 2010 Add-In will be selected. The program file will be named PowerPointAddIn.

The application that will be used for voice recognition is Julius because of its flexibility in processing the input from the voice of the speaker [10]. The input will come from the microphone. The function for speech detection will be made by initializing the speech recognizer. The speech recognition grammar will be created. This will detect the voice coming from the microphone. It will then be loaded into the speech recognizer and will be registered for a speech recognition event notification. A handler for the speech recognition event will be created so codes could be added by getting the speech as input. Different functions will be added to the program. The commands will be implement using the modified algorithm of voice-to-text. Instead of writing only the text the add-on will perform each commands using the event handler of PowerPoint add-on. Next mean go to next slide, previous means go to previous slide and page <number>, means go to a certain page.

With the support of OpenCV the functions for Image Processing Detection and Commands will be added into the program. The program must first detect the hand gesture in order to map camera coordinates to screen coordinates. The image will be resized so that the image will be easier to manipulate by computing the ratio of screen resolution to camera resolution using this equation:



where ( *x*' , *y*' ) is the camera position, (*cx*, *cy*) is the current screen resolution, and (*x, y*) is the corresponding screen position of camera position [3]. The background will be deleted to remove unnecessary data into the code using background subtraction. The noise will also be deleted using erosion and dilation functions of opencv. Erosion trims down the image area where the hand is not present and Dilation expands the area of the Image pixels which are not eroded. The Center will be computed using the equation:



where *xi* and *yi* are *x* and *y* coordinates of the *i* pixel in the hand region, and *k* denotes the number of pixels in the region [8]. Then the size of the hand will be computed by computing the area of the circle of the palm. After that, the position of the fingertips will be computed. The convex hull algorithm will determine if the finger is folded or not by finding the biggest polygon including all vertices [8]. This will detect finger tips on the hand. The command functions will be added using the image as input. Opening the palm or 5 fingers controls the cursor of the mouse by moving the hand. Opening only the index finger means left click, the pinky means right click and the index finger and middle finger means double left click.

**EVALUATION**

The study will be evaluated based on the accuracy of how the program performs each commands. If the add-on will perform 90% of the commands correctly and with little delay then it will be marked as *success*. First the voice recognition commands will be tested. The researcher will speak to the microphone and say the commands then time how long the PowerPoint responds with it. If it does not respond within 3 secs it will be marked as *fail*. If it responds, the researcher will check if command performed is correct or not. Doing the wrong command will be marked as *fail*. Each command will be performed and it will be evaluated based on how long is the delay of the command. If the delay if more than one second it will be marked as *bad response*. This will be tested in all of its commands. After that the image processing commands will be tested. First the researcher will test if the computer can read the hand. Then the movement of the hand will tested. The researcher will check the delay of movement from real-time to the response of the cursor. If the delay is more than one second it is marked as *bad response*. If it will not respond it will be marked as *fail*. Next the researcher will check if the add-on could do its desired commands. Each commands will be performed and it will be evaluated based on how long the computer will carry the command. The researcher will also check if it performed the correct task. Doing the wrong command will be marked as *fail*. This will be tested in all of its commands. Several tests will be made to determine the number of times the program fails to do the desired command.

**TIMELINE OF ACTIVITIES**

|  |  |
| --- | --- |
| **Date** | **Activity** |
| December 28, 2016 | Create a add-on for Microsoft PowerPoint |
| January 4, 2016 | Create detection and command functions for Voice Recognition |
| January 25, 2016 | Create detection and command functions for Image Processing |
| February 15, 2016 | Optimize and Test the application |

**REFERENCES**

1. Heera Lal Bhadreca (2015). *Finger Gesture Control Computer Mouse with Image Processing. Thesis.* BK Birla Institute of Engineering and Technology, Pilani Rajasthan, India: Department of ECE, Print.
2. [Vangie Beal](http://www.webopedia.com/author/Vangie-Beal), (N.D). *Voice Recognition*. retrieved on October 14, 2015, from http://www.webopedia.com/TERM/V/voice\_recognition.html.
3. [Computerhope](http://www.webopedia.com/author/Vangie-Beal), (N.D) . *What is voice recognition?*. retrieved on October 27, 2015, from http://www.webopedia.com/TERM/V/voice\_recognition.html.
4. Stefania Cuccurullo, Rita Francese, Sharefa Murad, Ignazio Passero and MaurizioTucci (2012). *A Gestural Approach to Presentation Exploiting Motion Capture Metaphors. Thesis.* Universita deqli Studi di Salerno, Fisciano, Italy, Print.
5. Adam Fourney, Michael Terry and Richard Mann (2010). *Gesturing In the Wild: Understanding the Effects and Implications of Gesture-Based Interaction for Dynamic Presentations*. *Thesis.* University of Waterloo, Print.
6. Tobias Heinroth, Dan Denich, and Wolfgang Minker (2001). *A Multitasking Approach to Adaptive Spoken Dialogue Management. Thesis.* Ulm University, 89081 Ulm, Germany: Institute of Information Technology, 2011. Print.
7. Moyen Mohammad Mustaquim (2011). *Fuzzy-Logic Controller for Speaking-Independent Speech Recognition System in Computer Game*. Thesis. Uppsala University, Uppsala, Sweden: Department of Informatics and Media, 2011. Print.
8. Hojoon Park (N.D.). *A Method for Controlling Mouse Movement using a Real Time Camera. Thesis.* Brown University, Providence, RI, USA: Department of Computer Science, Print.
9. Pcspeak, (N.D). *PowerPoint Commands?*. retrieved on November 5, 2015, from www.pcspeak.com/scripting9/powerpoint\_commands.pdf.
10. Kunal Shrivastava, Nishant Singhal, Pradip K. Das, and Shivashankar B. Nair (2013). *A Speech Recognition Client-Server Model for Control of Multiple Robots a Study*. Thesis. Indian Institute of Technology Guwahati. Assam, India: Department of Computer Science and Engineering, 2013. Print.
11. Steven J. Timmins (2004). *Tablet PC: Blackboard to the Web. Thesis.* University of Delaware, Newark DE, Print.