

Teachme, A Gesture Recognition System with Customization Feature

Hazem Qattous

Department of Computer
Information Systems
Applied Science Private University
Amman, Jordan

Bilal Sowan

Department of Computer Network
Systems
Applied Science Private University
Amman, Jordan

Omar AlSheikSalem

Department of Software Engineering
Applied Science Private University
Amman, Jordan

Abstract—Many presentation these days are done with the help of a presentation tool. Lecturers at Universities and researchers in conferences use such tools to order the flow of the presentation and to help audiences follow the presentation points. Presenters control the presentation tools using mouse and keyboard which keep the presenters always beside the computer machine to be close enough to the keyboard and mouse. This reduces the ability of the lecturer to move close to the audiences and reduces the eye contact with them. Moreover, using such traditional techniques in controlling presentation tools lack the communication naturalness. Several gesture recognition tools are introduced as solutions for these problems. However, these tools require the user to learn specific gestures to control the presentation and/or the mouse. These specific gestures can be considered as a gestures vocabulary for the gesture recognition system. This paper introduces a gesture recognition system, TeachMe, which controls Microsoft PowerPoint presentation tool and the mouse pointer. TeachMe also has a gesture customization feature that allows the user to customize some gestures according to his/her preference. TeachMe uses Kinect device as an interface for capturing gestures. This paper, specifically, discusses in details the techniques and factors taken into consideration for implementing the system and its customization feature.

Keywords—Microsoft Kinect®; Gesture recognition system; Gesture customization

I. INTRODUCTION

These days, presentation tools, such as MS PowerPoint and Prezi, are used by lecturers to present lectures in a controlled flow. Presentation tools are also used in the domain of conferences [1]. Such presentations are controlled using keyboard and mouse which requires the lecturer to be standing beside the computer table all the time to start a presentation, move the presentation slides and ending the presentation. However, the lecturer still need to move back and forth between the computer and projection plan from one side and the students or the audiences for better communication from another side [2] and [3]. Additionally, using keyboard and mouse reduces the interactive and naturalness in communication with audiences [4] and [5]. References [2] and [1] introduce that even the Bluetooth connected devices, which are designed to control presentations have some drawbacks regarding the problem of keeping the lecturer close the computer because of its limited functionalities. As an example, such devices cannot control the mouse. A subjectively noticed

problem of wireless connected devices is that the lecturer always forgets bringing the device to lecture. This returns the lecturer back to use keyboard and mouse.

All the above researchers introduced gesture capture devices as a solution for the above addressed problems. Although such gesture recognition devices were originally designed for interactive video games, several research found them suitable to be used as presentation controlling tools. Their suitability comes from their ability to provide an intuitive, natural and effective way for communication between human and machine [6] and [4]. Between all the research conducted in this field, very little focused on the problem of customizing the gesture. Basically, in any gesture recognition system, the implementation depends on a gesture vocabulary that is hard coded as that introduced by [7]. The user then should learn how to use the system based on the implemented vocabulary [8]. However, very little research focused on the customizing the gesture vocabulary to meet the preference of the user. This research addresses the customization problem and provide a simple solution that can be considered as an entrance for more research and implementations in this field. Providing the ability to customize gesture to a specific command is considered as the main contribution of this paper. This feature was implemented mainly for controlling MS PowerPoint presentation.

This paper presents a gesture recognition system, TeachMe. TeachMe is prototype that depends on Microsoft Kinect to recognize gestures. It is implemented to control Microsoft PowerPoint presentations and mouse pointer. The paper discusses implementation issues and obstacles faced during the system implementation. Additionally, it shows the parameters and factors taken into consideration during implementation. More importantly, the paper presents the new gesture customization feature implemented in TeachMe.

Next section of this paper presents the general structure and some general implementation information about TeachMe gesture recognition system. Implementation of TeachMe for controlling PowerPoint presentation and mouse pointer are then discussed in details. Later on, the paper presents the gesture customization feature implementation with some examples. After that, the paper introduces some related work and trials for implementing customization features at other gesture recognition systems. The limitations of TeachMe and its gesture customization features are discussed in detail

followed by a list the future work and plans for further development of the customization feature. Finally, the paper concludes and summarizes the work presented in the paper.

II. TEACHME GESTURE RECOGNITION SYSTEM

TeachMe is a research gesture recognition prototype that has been initially developed as a graduation project at Applied Science University. The system captures the lecturer body gestures through Microsoft Kinect® device. TeachMe development uses the standard Kinect SDK 1.8 as library. The Microsoft SDK uses three streams of data, RGB, depth data stream and skeleton data stream. As the coloring does not give useful information for this research purposes, it was neglected during implementation. The interest was directed to the depth and skeleton data streams. Once a gesture is captured, the system analyzes it. Analysis process maps the captured gestures to actions that are executed by the controlled part (presentation or mouse pointer). The mapping depends on a predefined gesture vocabulary that determines what gesture does which action. TeachMe recognizes two types of hand gestures. The first controls Microsoft PowerPoint presentation while the second controls the mouse pointer. Gestures controlling Microsoft PowerPoint includes starting a slideshow, moving slides forward and backward, and ending the slideshow. Gestures controlling mouse include moving mouse over the desktop, selecting an object (file or folder) by clicking on it, dragging an object, and double click on an object to open a file or explore a folder. Additionally, TeachMe provides the ability to customize some gestures to suit a lecturer preference. Customization feature will update the mapping between the recognized gesture and the action to be executed in the gesture analysis process.

III. CONTROLLING MS POWERPOINT

The system recognizes gestures that are required to control MS PowerPoint presentation. For this purpose, initially, the system focuses on three points, the user's head, left hand and right hand. It tracks the movements of both hands in reference to each other and in reference to the head point.

When the user moves their right hand from right to left, MS PowerPoint slideshow moves one slide forward (the current slide moves left) and the result is showing the next slide in the slideshow. If the user moves their left hand from left to right, the slideshow moves one slide backward (the current slide moves right), that is, showing the previous slide in the slideshow. This implementation was adopted as the most suitable one after several trials and experiments. An older version of the system was developed to depend only on the right hand. In that case, when the user was trying to move one slide forward by moving the hand from right to left, the system responds correctly. However, when the user returns their hand back to right, the system was wrongly recognizes the movement as a command to move the slide backwards again. The problem was handled, firstly, by taking time required for each movement into account. Reference [7] introduced that time is one of the factors that is used for controlling gesture recognition. In TeachMe, the case was that if the movement is quick, this means to perform the command while slow movement of the hand means to ignore the gesture. It was noticed that this puts some limitations on the user movement,

which contradicts with the purpose of the system. Accordingly, the solution of recognizing the movements of two hands was adopted.

One of the earliest problems faced the research and the implementation of TeachMe was that the system limits the lecturer movements. This means that the lecturer once recognized by the system, they cannot move their body freely during the lecture as their gestures will be analyzed and recognized by the system even if the gestures are directed to the students not to the system. To increase the system smoothness and gives the flexibility to the lecturer to move their hands during the lectures for purposes other than controlling the slideshow or the mouse, some other factors were added to control the gesture recognition. One factor, again, is the time. From experience gained in the older version of the system, time can play a major role in gesture recognition. The time required to perform the gesture was taken into consideration. During the system testing, it has been noticed that the user moves a hand from one side to another with a reasonable speed. The user focuses on the gesture and tries to make it recognizable for the system. This makes the user uses a reasonable speed that is neither very quick nor very slow. The reasonable speed was determined subjectively by five persons who were working on the project and implemented as a factor in the gesture recognition system. Note that the system does not recognize the start of the gesture. Instead, the system knows the gesture when it ends. Based on this, the system measures the time required for any movement of the user's body (including hands) and sends the data to be recognized as gesture or not.

Another factor taken into consideration was the second hand position while one hand is making the gesture. Moving right hand from right to left is not enough to generate the command of moving the slides forward, or "show next slide", in TeachMe. Instead, left hand should be in the left part and below the head point while the right hand is moving from right to left. Note that right and left sides of the body are recognized based on the head point that is considered by TeachMe implementation, as the "middle of the body" point reference.

One more factor that was added as a feature to recognize is the distance a hand moves from one side to another. The initial and the final positions of a hand are captured. A minimum threshold of 25 cm is set. Using threshold helps the lecturer to move their hands freely while lecturing without generating undesired commands because threshold adds more features to satisfy a gesture recognition. Threshold is introduced by [2] to recognize gestures. As a summary, the hand position according to the head point, the second hand position, the distance of the movement and the time required to go from first point (position) to the final point are all combined to recognize a gesture and generate a command. In general, it has been noticed that adding more features helps in better recognition for the required command and gives more freedom and flexibility to the user to complete their lecture smoothly.

In addition to moving slides forward and backward, TeachMe enables the user to start and end a PowerPoint presentation. Starting a presentation is done through moving the right hand side from down to up. The hand should start

moving from a point below the head to end up above the head point. Moving the left hand similarly ends the presentation. In both cases, the other hand should be in its normal relaxing position referring to the head point. Similar to previous movements, speed and distance of the movement are all required to recognize the start and end presentation commands. Apparently, TeachMe depends on a set of features that should be tested and satisfied before a gesture is recognized.

Reference [7] introduced an approach to recognize gesture using gesture description language. This language depends on text script written as rules with an expert system approach to manage the rules and recognize the gesture. The above-described way of implementing rules in TeachMe is a similar approach followed by [7] as each rule defines specific positions for joints. Rules implemented in this research are simpler, however. Reference [7] took into consideration the distance between joints as a factor for gesture recognition as this research does. Although they mention that time is an important factor in gesture recognition, they used the time in segmentation of the captured images not in the rules. On the contrast, TeachMe takes time directly as a factor for gesture recognition. Instead, they depended on time sequence for the poses to form a gesture. They say “our research presented in this paper proves that it is possible to unambiguously recognize, in real time (online recognition), a set of static poses and body gestures (even those that have many common parts in trajectories) using forward chaining reasoning schema when sets of gestures are described with an “if-like” set of rules with the ability to detect time sequences”. The if-like in their implementation is similar to the thresholds considered in this research.

IV. CONTROLLING MOUSE POINTER

TeachMe requires two hands to control the mouse pointer. The user should raise their left hand to be approximately beside the head point and keep it fixed in that position. A 5 cm threshold gap is set between the left hand position in reference to head point. The user uses right hand to control the mouse pointer movement by placing the right hand in front of his/her body and moves the hand. The mouse pointer moves with the hand movements. Mouse pointer control is possible using the same gesture on Windows desktop or any other running application.

To generate a left mouse click command, selecting an icon on the desktop as an example, the user should move the mouse over the required item to be selected, lower down the left hand side to be under the head point, which prevents the mouse pointer movement, and then, move their left hand forward away from the body. This hand push movement uses the Z-axis. There should be a specific distance between the body point and the hand point in Z-dimension to recognize the gesture as a click. Of course, there is a space threshold set here for the distance in Z-dimension. Double click command is achieved when the user repeats the forward movement twice with the left hand. The time between the two movements is one of the major features to recognize the gesture as a one click or double click. If the user leaves a significant period of time, between the two movements, the gesture will be recognized as two gestures of a click which may lead to a different behavior.

Traditional mouse uses the same concept to differentiate between the one click and double click. The period between the two movements was set to be less than two second to be recognized as a double click.

TeachMe also allows the user to move objects in the desktop using the mouse pointer. The user may use his/her both, left and right, hands in a push gesture in front of the body that leads to a holding of a left mouse click behavior. In this case, the gesture uses the z-dimension but for both hands this time. When the user moves his/her both hands together, the mouse pointer moves accordingly. Using this gesture, the user can drag and drop folders into others or into the recycle bin as an example. The gesture depends on the keeping the two hands beside each other and in front of the body with a distance at the same time.

V. GESTURE-COMMAND CUSTOMIZATION

TeachMe can be customized to suit the user preference. The customization tool is implemented to help the user specify a specific gesture to a specific command. In TeachMe current implementation, the commands that can be customized are those controlling a PowerPoint presentation but not the mouse pointer controlling. Once the system starts, the above described gestures and related commands are set as default for the user. If the user requires to modify a specific gesture to perform another command, customization tool can help. The tool shows four commands that can be customized, moving a slideshow forward, moving a slideshow backward, starting a slideshow and ending it. As an example, if the user wants to customize moving a slideshow one slide forward using the gesture “moving right hand from right to left” instead of “moving left hand from left to right”, the default, they can specify this by selecting the “Moving Forward” command using a radio button. This opens a dropdown list with possible gestures from which the user can select, “Right” in this case. To continue the example, the user also customize “Moving Backward” command to be “Left”. This customization appears in (Fig. 1). Similarly, the user can customize gestures for starting and ending slideshow commands (Fig. 2 and Fig. 3). According to the literature reviewed, such customization feature has not been implemented at any system before. The following figures show how the user can customize the system.

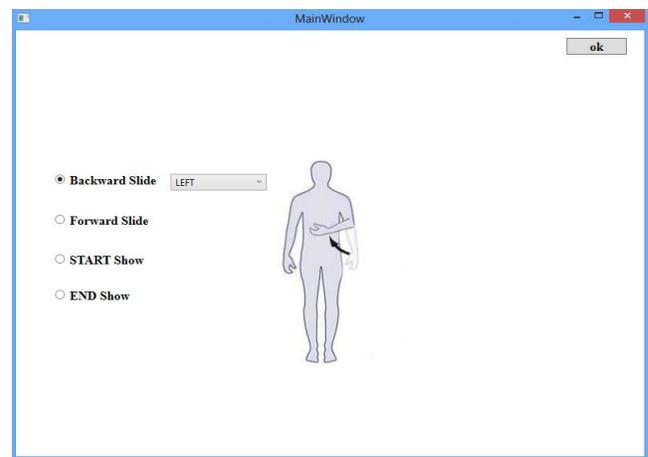


Fig. 1. Customizing moving a slide backward gesture

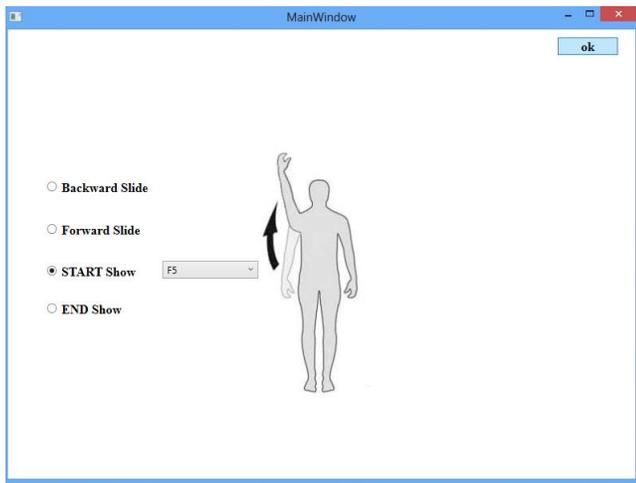


Fig. 2. Customizing starting a presentation gesture

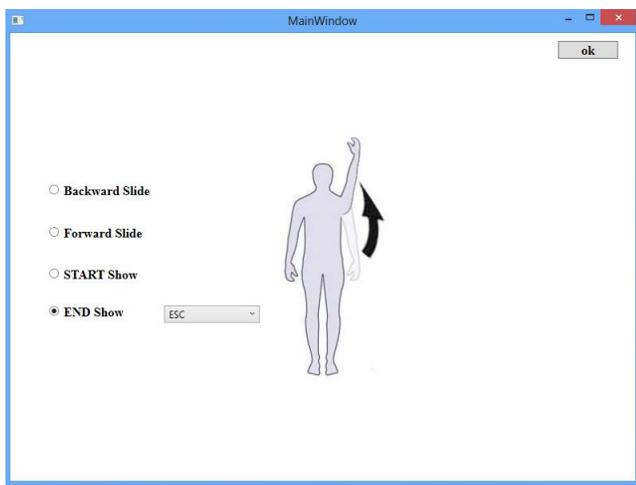


Fig. 3. Customizing ending a presentation gesture

VI. RELATED WORK

The most related work between reviewed papers was [9]. They invented an algorithm called uWave that needs only one time training for each gesture, which provides the user the ability to personalize gestures. Their implementation and recognition depend on quantization and dynamic time warping. Implementation of TeachMe is much simpler and has a simpler customization feature application. Throughout their paper, they try to reduce the complication calculations that result from the algorithm complexity.

Reference [1] presents a gesture recognition system. Their system uses what is called Kinect Presenter approach for controlling MS PowerPoint presentation. They present an empirical evaluation for their system. The system cannot control mouse pointer nor it can control starting and ending a presentation tasks that TeachMe does. Reference [10] introduced a gesture system that controls the mouse pointer. However, it was not controlling slideshow, as it is not implemented to work in the field of education.

A vision-based gesture recognition prototype has been introduced by [11]. The system controls mouse pointer to deal

with tasks related to WIMP (Windows, Icons, Menus, Pointers). The system is different from TeachMe that it does not control presentations and it is not designed for presentation purposes.

The research of [2] was motivated, similar to this research, with the trial to find a replacement technique for the traditional, keyboard and mouse presentation controlling technique. Their way of introducing the solution was through Ki-Prez, a gesture recognition system. Their system uses Kinect for capturing gestures. However, Ki-Prez only recognizes two gestures, swiping right and swiping left that move slides forward and backward. Similarly, [12] and [3] proposed the idea of replacing the traditional techniques with gesture recognition systems using Kinect in the domain of education. Reference [12] described using such systems as an introduction for the e-learning in the future. Reference [12] introduced a prototype that allows personalizing gestures. However, they did not include technical information about the personalization technique or its implementation.

Reference [13] introduced a gesture recognition system that controls presentation. Their system characterized with ability to recognize voice in addition to gesture to refine the response of the system with the required action. Their system is distinguished from other systems with this voice command feature. However, they do not include the option to control mouse pointer into their system as voice and gesture does not leave a space for the need of mouse controlling technique.

VII. TEACHME LIMITATIONS

The main limitation of TeachMe current implementation is the limited number of gestures that can be customized using the customization option. As shown above, the user can choose to customize limited number of gestures and the options of alternative gestures are also limited. Additionally, the user can only customize PowerPoint controlling gestures but not the mouse pointer controlling gestures. It has been decided to work on solving these limitations in the next TeachMe implementation. This is clarified in the future work section of this paper.

TeachMe depends on few number of joint points, three points only, for gesture recognition, reduces its ability to recognize complex gestures. However, from another point of view, using three points only to recognize gestures makes recognition process easy and reduces its complication. Putting in mind that TeachMe is a prototype gesture system designed for research, it was decided to make it simple to concentration on the new feature implementation instead of recognizing complicated gestures implementation. In addition, TeachMe satisfies all the research needs regarding the gestures recognition, as the research requires only simple gestures.

VIII. FUTURE WORK

The following work will be evaluating the new customization feature empirically. This will give an indication of continuing work in this field or find a diversion to another track. One more future work is to enhance the customization feature in TeachMe. Enhancement thought of is to introduce programming by example concept in customizing TeachMe. This will increase the flexibility of customizing gestures and

increase the number of gestures that can be customized according to the user preference.

IX. CONCLUSION

This paper introduced a gesture recognition system, TeachMe. The system recognizes gestures that control MS PowerPoint presentation and mouse pointer. It depends on Kinect as an interface for capturing gestures. The paper discussed the technical implementation details of the system. TeachMe implementation includes a gesture customization feature that allows the user to customize gesture according to their preference. Examples showing the work of the customization feature were included.

ACKNOWLEDGEMENTS

We would like to thank the students who participated in building TeachMe prototype as part of their graduation project, Mohammad Zedan Ismael and Rslan Zedan. The authors are grateful to the Applied Science Private University, Amman, Jordan, for the full financial support granted to this research.

REFERENCES

- [1] S. Cuccurullo, R. Francese, S. Murad, I. Passero and M. Tucci, "A gestural approach to presentation exploiting motion capture metaphors," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, Capri Island, Italy, 2012.
- [2] D. Martinovikj and N. Ackovska, "gesture recognition solution for presentation control," in *The 10th Conference for Informatics and Information Technology (CIIT 2013)*, 2013.
- [3] S. Butnariu and F. Girbacia, "Development of a Natural User Interface for Intuitive Presentations in Educational Process," in *The 8th International Scientific Conference eLearning and software for Education*, Bucharest, Romania, 2012.
- [4] S. Change, "Using Gesture Recognition to Control PowerPoint Using the Microsoft Kinect," MIT, Massachusetts, 2013.
- [5] S. Ha, S. Park, H. Hong and N. Kim, "Study on Gesture and Voice-based Interaction in Perspective of a Presentation Support Tool," *Journal of the Ergonomics Society of Korea*, vol. 31, no. 4, pp. 593-599, 2012.
- [6] T. Osunkoya and J.-C. Chern, "Gesture-Based Human-Computer-Interaction Using Kinect for Windows Mouse Control and PowerPoint Presentation," in *Proc. For the 46th. Midwest instruction and computing symposium (MICS2013)*, Wisconsin, 2013.
- [7] T. Hachaj and M. Ogiela, "Rule-based approach to recognizing human body poses and gestures in real time," *Multimedia Systems*, vol. 20, no. 1, pp. 81-99, 2014.
- [8] A. Butalia, D. Shah and R. Dharaskar, "Gesture recognition System," *International Journal of Computer Applications*, vol. 1, no. 5, p. 48-53, 2010.
- [9] J. Liu, L. Zhonga, J. Wickramasuriya and V. Vasudevan, "uWave: Accelerometer-based personalized gesture recognition and its applications," *Pervasive and Mobile Computing*, vol. 5, no. 6, p. 657-675, 2009.
- [10] A. Argyros and M. Lourakis, "Vision-based interpretation of hand gestures for remote control of a computer mouse," in *Computer Vision in Human-Computer Interaction*, Berlin Heidelberg, Springer, 2006, pp. 40-51.
- [11] F. Farhadi-Niaki, R. GhasemAghaei and A. Arya, "Empirical Study of a Vision-based Depth-Sensitive Human-Computer Interaction System," in *10th Asia Pacific Conference on Computer Human Interaction*, Matsue, Japan, 2012.
- [12] V. Tam and L.-S. Li, "Integrating the Kinect Camera, Gesture Recognition and Mobile Devices for Interactive Discussion," in *IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, Hong Kong, 2012.
- [13] J. Kim, S. Kim, K. Hong, D. Jean and K. Jung, "Presentation Interface Based on Gesture and Voice Recognition," in *Multimedia and Ubiquitous Engineering*, Berlin Heidelberg, Springer, 2014, pp. 75-81.