***The teaching system with gesture recognition***

***using Microsoft Kinect version 2 and Holographic***

|  |  |
| --- | --- |
| Le Quoc Khoi  Cantho FPT University  FPT University  Cantho, Vietnam  khoilqce130023@fpt.edu.vn  Le Nguyen Huu Tai  Cantho FPT University  FPT University  Cantho, Vietnam  tailnhce130024@fpt.edu.vn | Kieu Hieu Thanh  Cantho FPT University  FPT University  Cantho, Vietnam  thanhkhce140026@fpt.edu.vn  Vo Hong Khanh  Cantho FPT University  FPT University  Cantho, Vietnam  khanhvh@fe.edu.vn |

*Abstract*

In this paper, we propose a gestures recognition system based on Microsoft Kinect version 2 sensors. Our research identifies 6 gestures applied to control the slideshows support for teaching. We tested this system about 1800 samples. The results show that the identification system with over 95% accuracy, even in difficult conditions such as poor lighting, many garbage objects, etc. Combine with holographic, the research brings the most visual feeling about teaching a presentation, especially in 3D-related lessons. Thereby enhancing the ability to acquire and create strong interest for learners.

Keywords

Microsoft Kinect, Holographic, Gesture recognition, Teaching System, Supporting education, Visual presentation.

# Introduction

Every day, students must research a lot of subjects, resulting in students difficult to acquire new knowledge. To attract students' interest in the subjects of the lecture, the teacher needs to innovate the way of teaching, how to make the lecture interesting to help students access the best knowledge. Visualization is one of the most effective methods. The lectures are visually displayed and attract students' interest. However, it is still possible to enhance the efficiency of visualization to a new level by changing the presentation control method by direct interaction. Gesture-recognition is a promising technology that brings new breakthroughs in direct instructional support.

The system we propose consists of two parts, a gesture sensor and a holographic. The cost of investing in this system is not higher than normal projectors, but the benefits bring immensely, so applying this technology in practice is perfectly feasible. The use of dedicated motion detection sensors is more efficient than using motion detection algorithms on conventional cameras. As a result, the system is very fast and accurate. Kinect - a device developed and supported by Microsoft that can do so, provides a wide range of identification. What's more, our solution requires the use of Kinect version 2 - an upgraded version with more new technology with higher accuracy [1], so that the gesture control signal of the instructor will be more accurate. Combined with the presentation on the holographic, the lecture will become more attractive, making learners feel new and exciting to absorb new knowledge. Especially in lectures related to 3D, holographic will make it easier for learners to understand the lesson, rather than take hours to hear the teacher presenting in the traditional way.

With the efficiency that the solution brings as well as the exigence, the solution promises to have more potential for use and development. Can conclude that "The teaching system with gesture recognition Using Microsoft Kinect version 2 and Holographic" is an essential and viable solution.

*Related research in the field of using Kinect to support the presentation*

In the gesture recognition system, the earliest method was to use 2D imaging devices combined with image-processing algorithms. In 1994, Davis J. and Shah M. developed a system for Recognizing hand gestures [2] with an accuracy of 94% based on the FSM (Finite State Machine) model. The characteristic of this method is that the accuracy is only high when the image is obtained in good conditions. The identification accuracy of this method is very low when the image is captured in low light or near-color conditions with coat color and background color.

In 2012, with the introduction of the Microsoft Kinect sensor [3], a new research direction was opened to solve the problem of gesture recognition. With a reasonable price, the ability to apply this type of device is in fact feasible.

A research by Yi Li on Hand gesture recognition using Kinect (2012) [4] shows the high effectiveness of the Microsoft Kinect sensor in hand-gesture recognition. Based on the ability to receive infrared images and the ability to strip the background supported on this sensor, the gesture recognition system achieves high accuracy when shooting in darkness, when the skin color matches the color of the outfit or background color. The accuracy of this system is from 84% to 99% with single-hand gestures, and from 90% to 100% if both hands perform the same gesture at the same time.

In the solution Using gesture recognition to control PowerPoint using Microsoft Kinect sensor was developed by Chang, Stephen M (2013) [5], this system uses a Hidden Markov Model (HMM) to classify the performed gestures in conjunction with a Support Vector Machine (SVM) to perform real-time segmentation of gestures. The fusion of these two models allows the system to classify gestures in real time as they are being done instead of waiting until completion. The incorporation of speech statements gives the user an additional level of precision and control over the system.

In Vietnam, the researchers Nguyen Cong Tin, Ha Kim Tung, Vo Thanh Nghia, and Le Thanh Cong (2016) development research Applying Kinect to presentation documents [6] show that the minimum distance between the operator and the Microsoft Kinect for sensor recognition is 1.5 m, command response time is 0.03s. However, this research only works on Windows 7 and Microsoft Office PowerPoint 2007.

In general, the application of Microsoft Kinect to gesture recognition has been studied extensively. However, the previous researches used only the Microsoft Kinect version 1 sensors to control PowerPoint slides on the screen. No research yet combined the use of Microsoft Kinect sensor version 2 and Holographic to support teaching.

# PROBLEM and Solution plan

1. *Problem*

This solution is based on the Microsoft Kinect version 2 sensor to recognize the user's hand gestures. The PowerPoint slides are controlled based on the input data that is hand gestures of the user. PowerPoint slideshow controlled for instance are turn off slides, turn on slides, slide transitions... Moreover, the presentation content is displayed on the hic, aiming to increase the attractiveness of lectures, create new experiences for users.

1. *Solution Plan*

*2.1. The research about Kinect and operational principle of Kinect*

1. *Compare Kinect version 1 and version 2*[7, p. 2]

As you can see from the table, the Kinect v2 is equipped with a full HD resolution color camera and a higher resolution depth camera. As a result, the quality of images captured by the version 2 is much better than the version 1.

|  |  |  |
| --- | --- | --- |
| **Feature** | **Kinect version 1** | **Kinect version 2** |
| Color Camera | 640x480 @30fps | 1920x1080 @30fps |
| Deep Camera | 320x240 | 512x424 |
| Max Depth Distance | ~4.5m | 8m |
| Min Depth Distance | 40cm | 50cm |
| Depth Horizontal Field of View | 57 degrees | 70 degrees |
| Depth Vertical Field of View | 43 degrees | 60 degrees |
| Tilt Motor | yes | no |
| Skeleton Joints Defined | 20 joints | 25 joints |
| Full Skeleton Tracked | 2 | 6 |
| USB Standard | 2.0 | 3.0 |
| Supported OS | Windows 7, 8 | Win 8 or higher |

Table 1: Technical information of Kinect v1 and Kinect v2

Kinect version 2 can handle 2 gigabytes of data per second, USB 3 provides almost 10x faster bandwidth for the data transfer, 60% wider field of vision, and can detect and track 25 joints from 6 people’s bodies including thumbs. In comparison, the Kinect v1 could only track 20 joints from 2 people. The Kinect v1.0 device doesn’t have the fidelity to individually track fingers and stretching and shrinking with hands and arms but the Kinect v2 has these capabilities. For these reasons, we chose the Kinect v2 to be our main gesture-detecting sensor to reduce wrong-gesture detections.

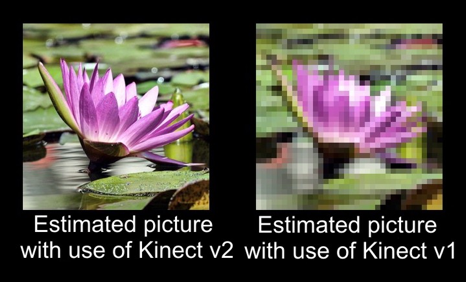


Figure 1: The difference between the qualities of image captured by the Kinect v1 and Kinect v2 [8]

1. *Structure and operational principles of the Kinect v2*

For the hardware, the Kinect v2 has 2 division[1]: the depth and the color. The depth division has two sensors: an infrared (IR) light and a depth image CMOS. Using the Light Coding Algorithm, the Kinect v2 can capture and create a depth map. The Kinect sensor can also draw a skeleton graph from the depth map. The color division has a color image CMOS to capture color images and data. Using data captured from the sensor, the processor then used special algorithms to process data. With the given Kinect libraries in the Kinect SDK, we can program it, extract processed data for our own purposes and in this case, control the Microsoft PowerPoint presentations.

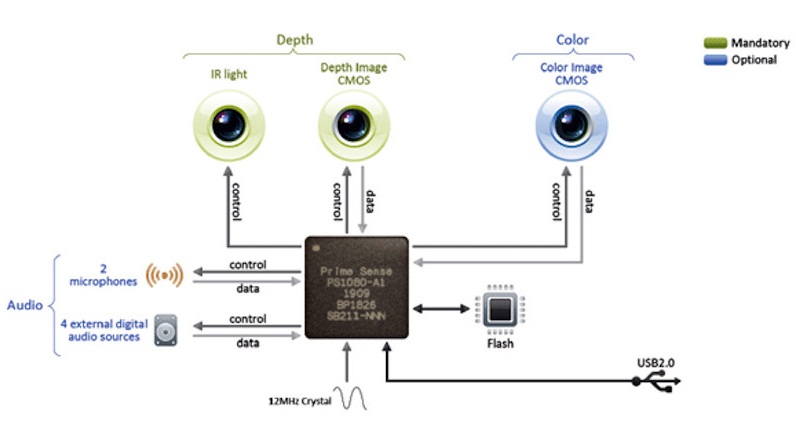


Figure 2: The structure of the Kinect

The software is what makes the Kinect a breakthrough device. Developers for the Kinect gathered an incredible amount of data regarding motion-capture of actual moving things in real-life scenarios. Processing all this data using a special artificial intelligence machine-learning algorithm allows the Kinect to map the visual data it collects to models representing people of different backgrounds (age, height, gender, body type, clothing and more).

1. *Introduce the Light Coding algorithm*

Kinect uses a Light Coding algorithm[10] to recognize gestures. Kinect is, deep down, a structured light scanner, meaning that it projects an infrared pattern (so invisible for us). First, it projects a known pattern (speckles) in near-infrared light. Then an IR (infrared) camera observers the scene. Bases on the projection generated by a diffuser and diffractive element of IR light, the Kinect uses a special algorithm to create a depth map.

* 1. *Presentation contents on holographic*

Holography is a technique that allows the light scattered from an object to be recorded and later reconstructed so that it appears as if the object is in the same position relative to the recording medium as it was when recorded. There are many new technologies applied for holographic, so there are many types of holograms: transmission holograms, white light reflection holograms, multiple channel holograms, real image holograms[11].

In this research, we used holograms with four sides and one bottom. Depending on the presentation conditions, the size of the holographic will change flexibly to meet the needs of many subjects.

# plan implementation

1. *The structure of the system*

As we can from figure 3, Kinect is an input device. After the data had been processed by the program, the computer executes the instructions and then controls the objects (PowerPoint slides). Projector, laptop’s screen, holographic, etc. are output devices that display PowerPoint presentations. In this system, the controlling presentation function is configured and include the following actions:

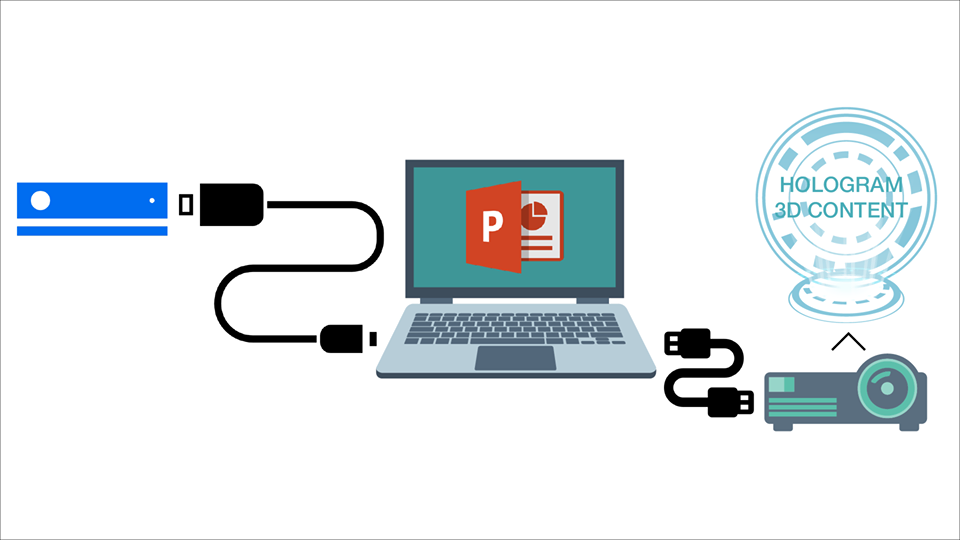
* Change animations/slides (previous animation, next animation, first slide, last slide, etc.)
* Taking note when presenting
* Laser pointer

Figure 3: System diagram

1. *Conventions of the gestures*

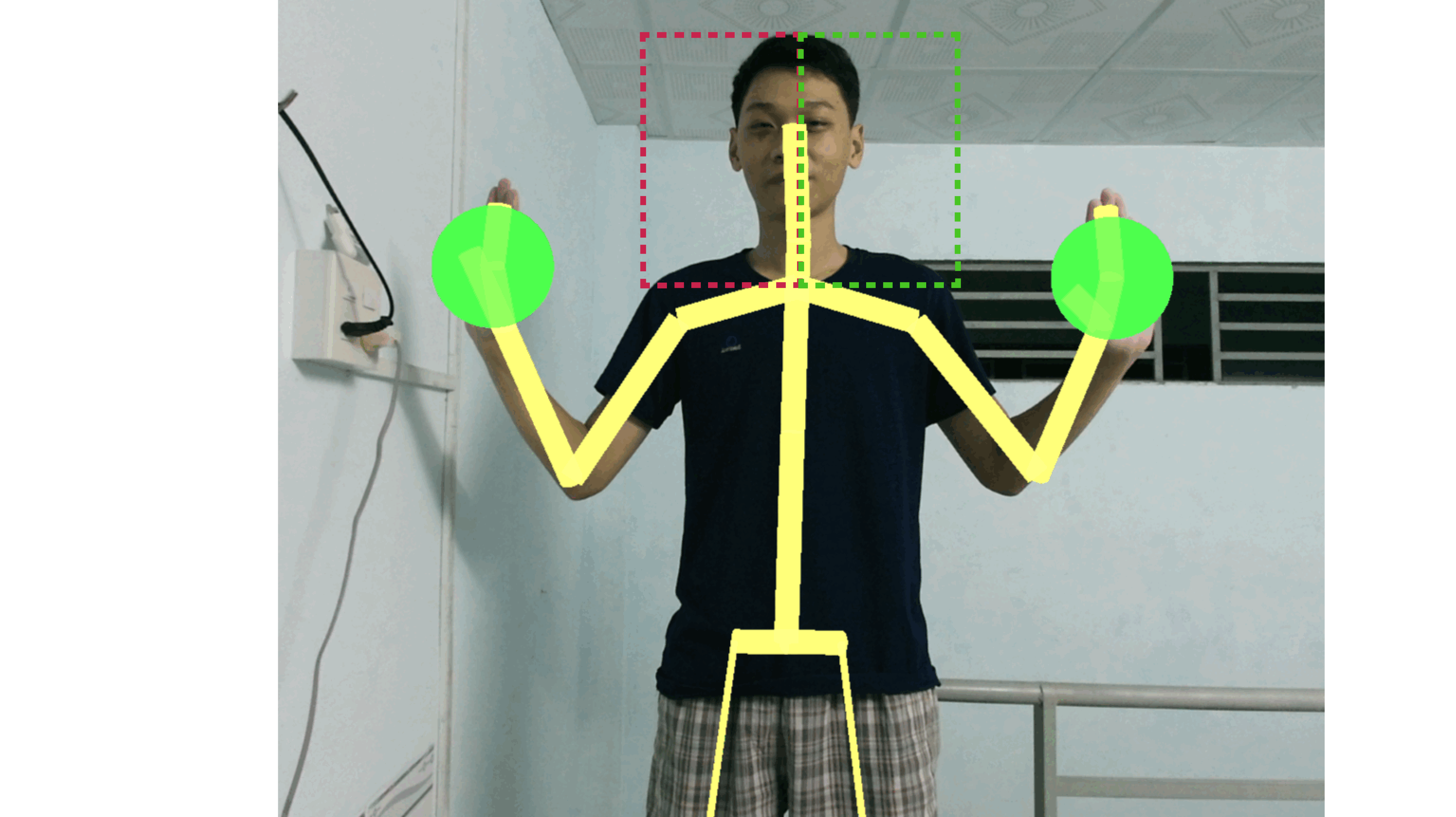
**

Figure 4: Describe gestures which perform next animation or previous animation in the slide

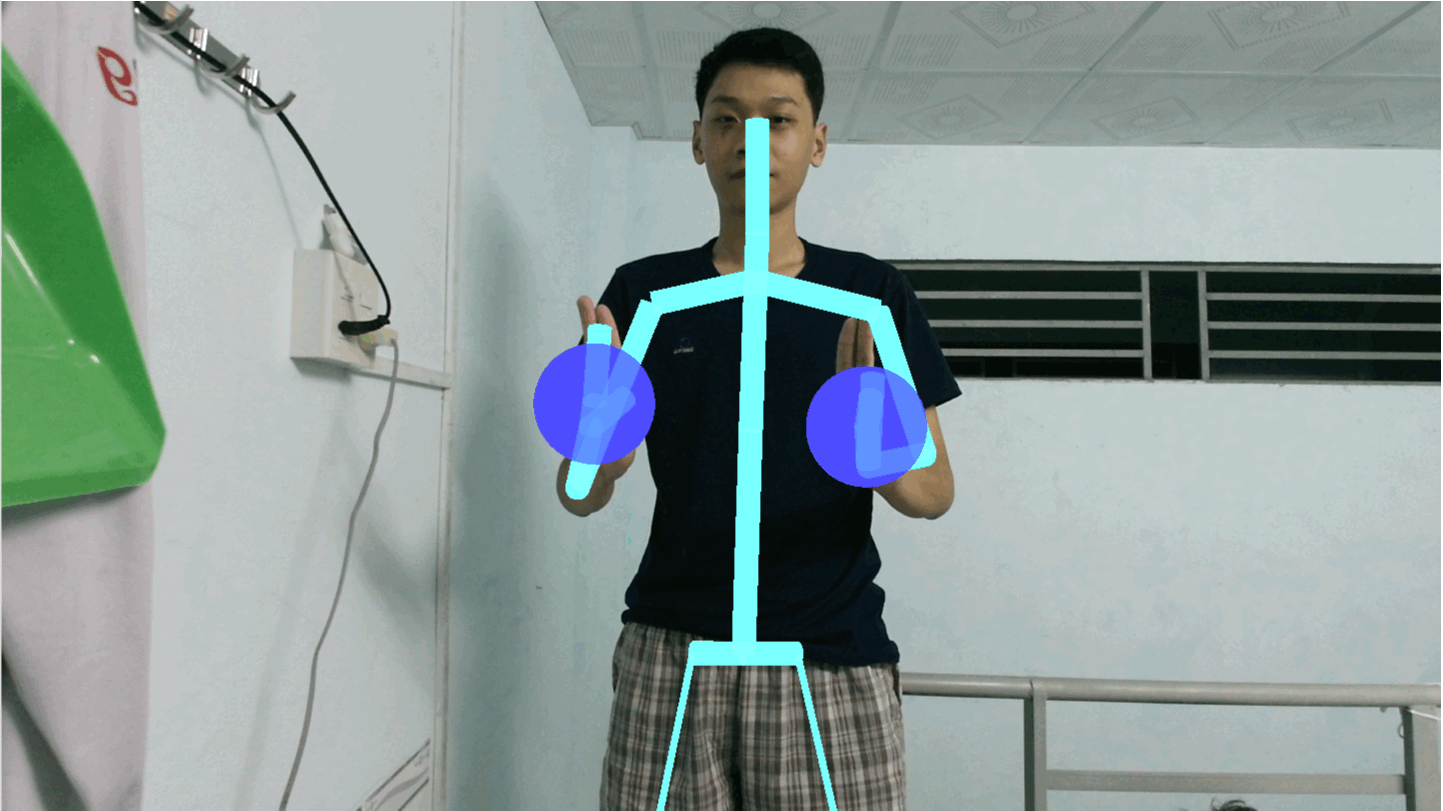


Figure 5: Describe the gesture which performs return to the first slide



Figure 6: Describe the gesture which performs go to last page

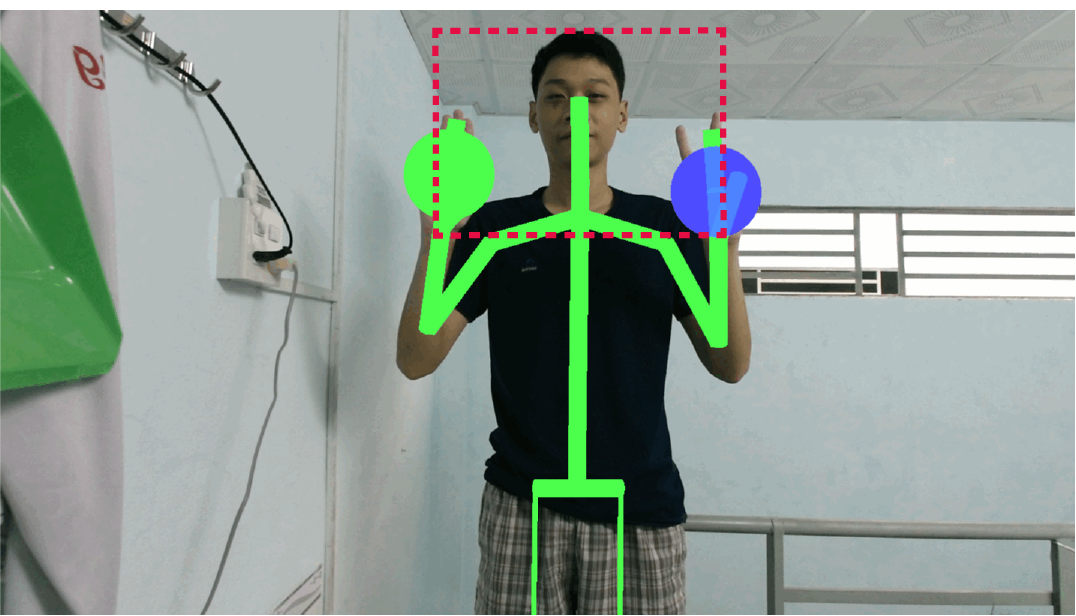


Figure 7: Describe the gesture which performs take note on the slide

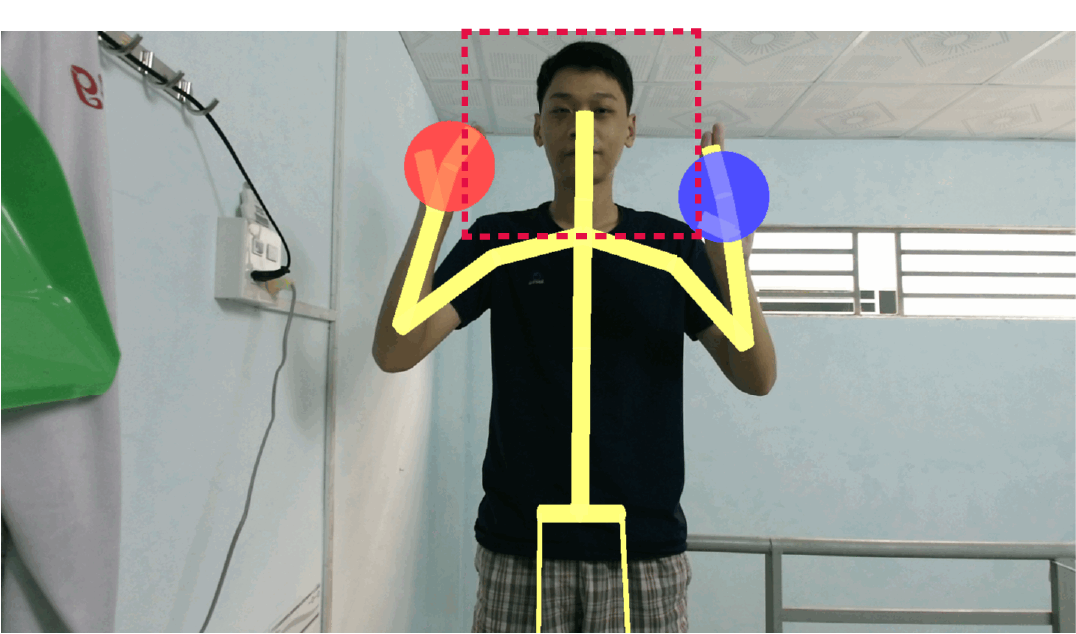


Figure 8: Describe the gesture which performs Laser Pointer

Below is a table of conventions and significances of gestures to control the slide:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| # | **Function** | **Description** | **Picture** |
| 1 | Back to the previous animation | Open right hand and swipe left in the green zone with a speed ~0.3-0.5m/s | Figure 4 |
| 2 | Switch to the next animation | Open left hand and swipe right in the red zone with a speed ~0.3-0.5m/s |
| 3 | Return to the first slide | Both hands in lasso state | Figure 5 |
| 4 | Go to last page | Close left hand and lasso right hand | Figure 6 |
| 5 | Take notes on the slide | Left hand open and right hand in lasso state. (Note: both hands must be in the red zone) | Figure 7 |
| 6 | Laser pointer | Left hand close and right hand in lasso state. (Note: both hands must be in the red zone) | Figure 8 |

*Table 2: The conventions of the gestures*

We have researched carefully to make the conventions so that the presenters can comfortably carry out the presentations. Among the most commonly used gestures, we must list the next and previous animation. We chose an open left hand and swipe right will perform the next animation function, open the right hand and swipe left will perform the previous animation function because they are the most natural and comfortable gestures for presenters. In the return to the first slide function, we require users to lasso both hands. This function is rarely used so we have assigned it a hand gesture and the presenter will never be confused. Accordingly, the move to last page function is performed by grasp the left hand and lasso right hand. To take note, users simply need to raise both hands over the shoulder with the left hand in open state and right hand in lasso state. To use a laser pointer, do as the above but left hand in the closed state.

1. *The operating principles of systems*

First, the Kinect uses the Light Coding algorithm to collect depth data. Then based on the gathered data, it analyzes and aligns the data with a stored collection of skeleton structure to interpret our movements. This result in a simplified 3D skeleton picture with joints and hand states (if available). The developer can set conventions like the x, y, z coordinates of hands, the distance between joints or hands, etc.… to limit the available zone of a gesture so that the presenter can comfortably perform gestures without the worrying about the misunderstanding and wrong-gestures performing of the Kinect.

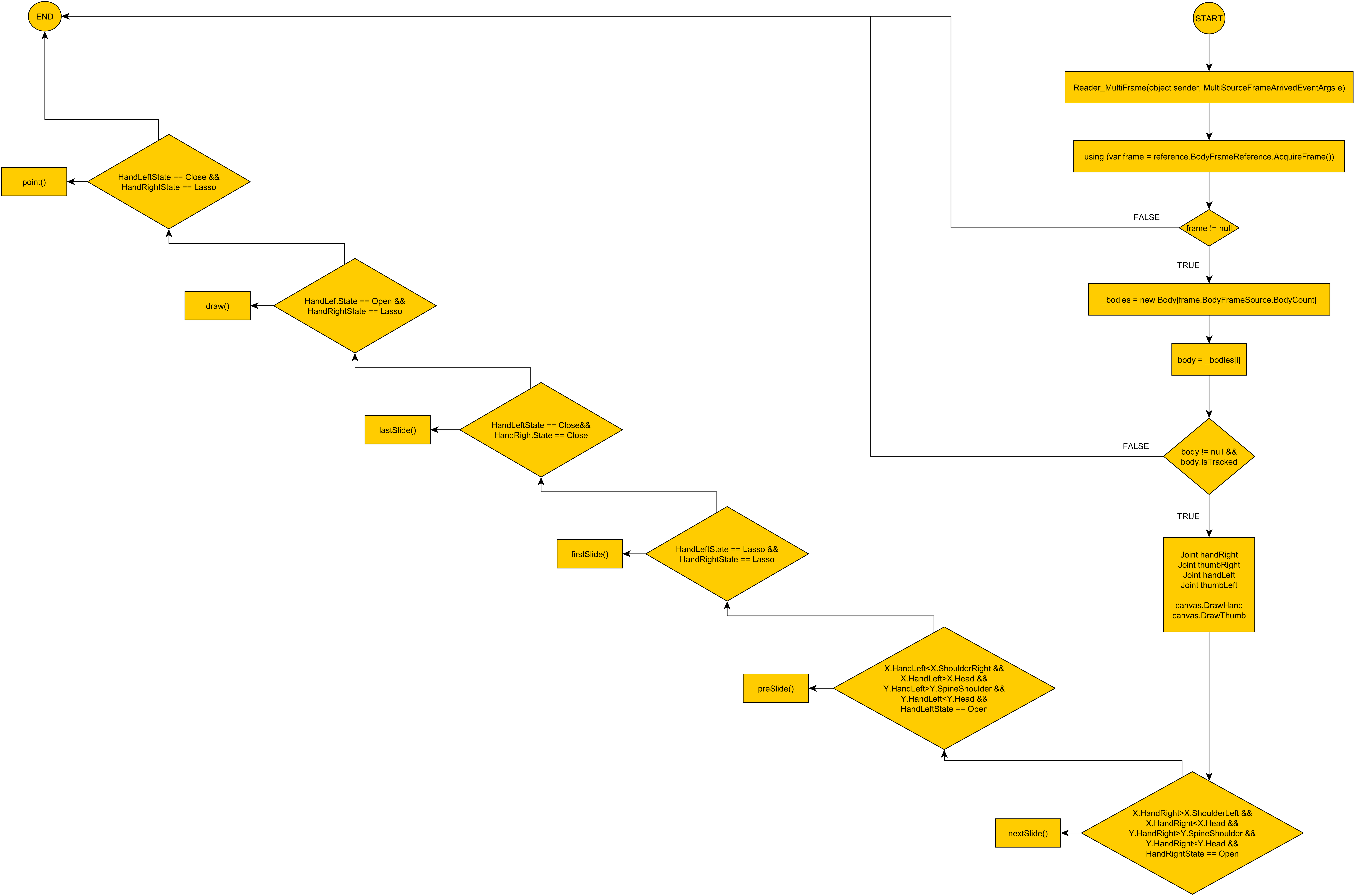
**

Figure 9: The flowchart describes the Algorithm of this solution

# ANalysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Good | Poor | Bad | **Average (%)** |
| Back to the previous animation | 96 | 94 | 96 | 95.33333 |
| Switch to the next animation | 95 | 94 | 95 | 94.66667 |
| Return to the first slide | 100 | 99 | 99 | 99.33333 |
| Go to last page | 98 | 97 | 98 | 97.66667 |
| Take notes on the slide | 91 | 93 | 89 | 91 |
| Laser pointer | 93 | 94 | 91 | 92.66667 |
| **Average (%)** | 95.5 | 95.16667 | 94.66667 | **95.1111** |

We have experimented on 20 testers. The test slide has 10 pages. Each person performs a gesture 15 times: 5 times in good light and no obstruction condition, 5 times in poor light condition, 5 times in poor light and cluttered background condition. The total we received is 1800 samples. Experimental results are shown in the following table and chart:

*Table 3: Experimental result in different conditions*

We have conducted the experiment in a 50 square meters classroom in 3 different conditions:

* Good: the illuminance over 400lux and there were no obstacles between the users and the Kinect sensor.
* Poor: the illuminance from 150lux to 200 lux with some obstacles like a student, some laptops, bags, notebooks, etc.
* Bad: the illuminance below 150lux with 2 or more students, many laptops, more bags, notebooks, chairs, desks, etc.

# EXPERIMENTAL RESults and conclusion

We have successfully applied Kinect v2 to recognize the 6 gestures that used to control PowerPoint slideshows combine with holographic to create a presentation system which can support the teaching. We have tested this system about 1800 samples. The results show that the identification system perfectly with over 95% accuracy even in difficult conditions such as poor lighting, many garbage objects, etc. We have conducted a survey with 100 students at Cantho FPT University to evaluate the user's interest in a classroom where presenter use gestures to control the presentation being displayed on holographic. The feedback we have received that up to 96 people have confirmed that they were more interested in learning by this method, they appreciate this solution and hope to use this solution to present their presentation.

##### Development

With many Kinect applications, not only in education but also in other fields. This research has great potential for new developments. For instance, we could integrate voice recognition and facial recognition system to enhance the accuracy and the interaction of the system; Change keyboard shortcuts can control any application on the computer, so the program can become a game system using gestures, the entertainment systems using gestures... If changing holographic to 3D projectors, the presentation content will not be limit at 3D simulations, but 3D augmented reality and the system will perfectly suitable for tasks that need to use holographic such as teaching geometry, biology, designing, ... If the presentation is the image of the medical records, the system will very helpful for surgeons as they can wear gloves while presenting. This system can also be used to control smart devices, smart homes. The development of the topic is not limited to research and lifestyle applications, but also virtual music applications, gesture health systems, or remote mine-clearing robots.

##### Acknowledgment

First, I would like to acknowledge PhD. Vo Hong Khanh, PhD. Nguyen Hong Chi and Mr. Tran Thanh Danh for their guidance and support throughout the process which made this project possible.

Additional thanks to all of our friends in the Cantho FPT University who have offered constant support and encouragement.

##### References

[1] Skarredghost, “The difference between Kinect v2 and v1,” *The Ghost Howls*, 02-Dec-2016.

[2] D. James and S. Mubarak, “Recognizing Hand Gestures.” 06-Feb-1994.

[3] “Starting February 1, 2012: Use the Power of Kinect for Windows to Change the World – Kinect for Windows Product Blog.” [Online]. Available: https://blogs.msdn.microsoft.com/kinectforwindows/2012/01/09/starting-february-1-2012-use-the-power-of-kinect-for-windows-to-change-the-world/. [Accessed: 29-Oct-2018].

[4] Yi Li, “Hand gesture recognition using Kinect,” in *2012 IEEE International Conference on Computer Science and Automation Engineering*, Beijing, China, 2012, pp. 196–199.

[5] S. M. Chang, “Using gesture recognition to control PowerPoint using the Microsoft Kinect,” Thesis, Massachusetts Institute of Technology, 2013.

[6] Tin, N. C., Tung, H. K., Nghia, V. T., and Cong, L. T., “Nghiên cứu và ứng dụng Kinect vào việc trình chiếu tài liệu,” 2016. [Online]. Available: http://kdientu.duytan.edu.vn/vi-vn/nghien-cuu-khoa-hoc/nghien-cuu-va-ung-dung-kinect-vao-viec-trinh-chieu-tai-lieu/. [Accessed: 29-Oct-2018].

[7] O. Wasenmüller and D. Stricker, “Comparison of Kinect V1 and V2 Depth Images in Terms of Accuracy and Precision,” in *Computer Vision – ACCV 2016 Workshops*, 2017, pp. 34–45.

[8] A. Znagui Hassani, B. van Dijk, G. Ludden, and H. Eertink, “Touch versus In-Air Hand Gestures: Evaluating the Acceptance by Seniors of Human-Robot Interaction,” in *Ambient Intelligence*, 2011, pp. 309–313.

[9] “What are the different types of holography? How did each evolve? - Quora.” [Online]. Available: https://www.quora.com/What-are-the-different-types-of-holography-How-did-each-evolve. [Accessed: 09-Nov-2018].