



CONSTRUCTION OF A SOMATOSENSORY INTERACTIVE SYSTEM BASED ON COMPUTER VISION AND AUGMENTED REALITY TECHNIQUES USING THE KINECT DEVICE FOR FOOD AND AGRICULTURAL EDUCATION



Chao-Ming Wang ^{*1}  , Yu-Hui Lin ²

^{*1,2} Department of Digital Media Design, National Yunlin University of Science & Technology, Douliu City, Yunlin County, Taiwan 64002, ROC



DOI: <https://doi.org/10.29121/IJOEST.v5.i2.2021.162>

Article Type: Research Article

Article Citation: Chao-Ming Wang, and Yu-Hui Lin. (2021).

CONSTRUCTION OF A SOMATOSENSORY INTERACTIVE SYSTEM BASED ON COMPUTER VISION AND AUGMENTED REALITY TECHNIQUES USING THE KINECT DEVICE FOR FOOD AND AGRICULTURAL EDUCATION.

International Journal of Engineering Science Technologies, 5(2), 1-37.
<https://doi.org/10.29121/IJOEST.v5.i2.2021.162>

Received Date: 13 February 2021

Accepted Date: 08 March 2021

Keywords:

Food and Agricultural Education
Game Playing
Human-Machine Interaction
Computer Vision
Augmented Reality

ABSTRACT

A somatosensory interactive system based on computer vision and augmented reality (AR) techniques using the Kinect device is proposed, on which a game of harvesting three kinds of fruit can be played for food and agricultural education. The Kinect is used to capture users' motion images, the Unity3D is used as the game engine, and the Kinect SDK is used for developing programs, to implement the tasks of face detection and tracking, hand-gesture recognition, and body-model matching and tracking involved in fruit-harvesting activities. AR-based photos of the harvest result can be taken and downloaded as souvenirs. The system was exhibited and observations of the users' performances as well as interviews with experts and the users were conducted. The collected opinions were used to evaluate the effectiveness of the system, reaching the following conclusions: 1) the interactive experience of using this system is simple and intuitive; 2) the use of body movements for man-machine interaction is given positive reviews; 3) the introduction of somatosensory interactive education can arouse participants' interest, achieving the effect of edutainment; and 4) the experience of taking commemorative photos can achieve the publicity and promotion effect of food and agricultural education through sharing on social media.

1. INTRODUCTION

1.1. BACKGROUND AND MOTIVATION

The issue of agricultural development is always one of the main concerns of human beings in the world. The Food and Agriculture Organization (FAO) of the United Nations pointed out that agriculture not only has to provide people with food and clothing, but also needs to maintain sustainable operations while meeting the social objectives of food safety, nutrition supplies, and health promotion [1], [2], [3]. A good way to advocate this concept to the public

is via *food and agricultural education* [4], [5]. Especially, food and agricultural education helps children grow up and cultivate their abilities to choose food correctly and have good eating habits [6]. For adults, food and agricultural education can enhance their sense of the local agriculture culture, and promote their healthy life via the selections of appropriate agricultural products as daily food to eat [7].

With the progress of technology, many digital media and related devices have been developed as tools for assisting educations in various knowledge fields. These devices can be used to simulate the experiencing situation related to the educational content, and bring interactive experiences into the learning process [8], [9], [10], [11]. At present, somatosensory interaction techniques, such as skeleton detection, gesture recognition, face detection, facial expression analysis, eye tracking, etc., have been developed, which can be used in various human-machine interaction applications [12], [13], [14], [15]. In particular, Hsu [14] mentioned that the application of computer vision in education has great potentials, not only bringing learners interesting experiences of interactive learning, but also enhancing their willingness to participate in related activities.

The aim of this study is to utilize computer vision and augmented reality (AR) techniques to design an interactive system for food and agricultural education, on which a game of fruit harvesting can be played via the use of the Kinect device with the “human” as the interface. Through game playing, the system can offer users interactive experiences of agricultural activities conducted in fruit farms. In particular, human motions and body joint points are detected by the Kinect device as features for the man-machine interaction and process control in the game process.

1.2. LITERATURE REVIEW

1.2.1. FOOD AND AGRICULTURAL EDUCATION

Food and agricultural education was promoted by many countries in the world [16], [17], [18], [19]. Asakao [20] advocated that food and agricultural education not only should integrate experiences related to “agriculture” and “diet,” but also needs to include the concept of environmental education, so as to make food and agricultural education a holistic education.

Japan’s “Basic Law on Food Education” [48] defines the content of food and agricultural education to include six items: 1) promoting a healthy diet; 2) being grateful for food; 3) understanding the significance of food through participations in various experiencing activities; 4) considering regional characteristics to adjust the content of food and agricultural education; 5) maintaining traditional food culture; and 6) understanding the relevant knowledge of food safety. Watanabe et al. [49] classified the school’s food and agriculture education as well as the learning contents into seven types, including: 1) local agriculture; 2) agricultural experience-based learning; 3) agricultural knowledge-based learning; 4) life habits and diseases; 5) nutrition knowledge; 6) diet actions and habits; and 7) traditional modern food culture.

The National Association of Agricultural Educators (NAAE) in the United States regards *learning by experiencing* as an important way for food and agricultural education [21]. Also, the previous survey of food and agricultural education in various countries show two common focuses, namely, *food experience* and *agricultural experience*, where the former emphasizes on the use of meals and ingredients while the latter on the understanding of agricultural knowledge and production. In this study, the latter is chosen as the research theme, aiming at constructing an interactive teaching aid for food and agricultural education in the form of a game which may be played by the public to gain contextual experiences of agricultural production activities.

1.2.2. HUMAN-MACHINE INTERACTION

Human-machine interaction is a research field about communication between “humans” and “machines.” With the advance of communication and computer technologies, human-machine interaction has been developed in many ways to create connections between technology and human beings, providing users with new types of experience and fun [22], [23], [24].

Jaimes and Sebe [25] mentioned that some input modes of human-machine interfaces correspond to the human sensory systems; with proper peripherals, computers can be used to simulate human sensory systems and create

many interactive devices for different applications. Integrating proper interfaces for users to interact in practical situations and providing them with good system-performance experiences is importance in the field of *interactive design* [26]. A good interactive design should follow the principles of “3e indicators,” namely, “effectiveness,” “easiness,” and “enjoyment,” as suggested by Yeh [24], which are elaborated in the following:

effectiveness – corresponding to “functionality,” meaning that the work must effectively guide users to complete tasks, solve problems, or achieve goals;

easiness – corresponding to “usability,” meaning that the design of the work needs to help users reduce memory, physical, visual, and comprehension works, and the use of the work must be easy for the users; and

enjoyment – corresponding to “pleasantness,” meaning that users must be able to enjoy the process while playing the work, which can be subdivided into four levels: physical, social, psychological, and ideological pleasures.

According to the surveys conducted by Sharma, et al. [50], Preece et al. [51], Jaimes and Sebe [25], Gibbon et al. [52], and Turk [53], the various forms of human-machine interfacing are summarized in this study as shown in Tables 1 and 2 for the input and output parts of the interface.

Table 1: Input types of human-computer interfacing.

Form	Input device	Input information	User behavior
<i>Physical interface</i> (Tangible)	<ul style="list-style-type: none"> • Mouse • Keyboard • Physical object • Sensors such as switches 	<ul style="list-style-type: none"> • Forming a loop or not • The sensed value 	The user actually manipulates the physical interface.
<i>Touch interface</i> (Touchable)	<ul style="list-style-type: none"> • Capacitive touch screen • Capacitive switch • Conductive material and the development board 	<ul style="list-style-type: none"> • Touching the interface • Touched position • Touched area 	The user touches the interface with skin or other conductive materials.
<i>Perceptual interface</i> (Perspective)	<ul style="list-style-type: none"> • Camera • Infrared sensor • Eye Tracker 	<ul style="list-style-type: none"> • Existence of Target • Target location • The state of the target 	The user moves within the sensing range, allowing the system to recognize the target's state.
<i>Wearable interface</i> (Wearable)	Electronic devices that can be worn on the body, such as <ul style="list-style-type: none"> • watches, • bracelets, • glasses, etc. 	<ul style="list-style-type: none"> • The location of the device • The movement status of the device • Information transmitted by the device 	When the user wears the device for activities, the values obtained during the period can be transmitted to the system.
<i>Voice interface</i> (Speech)	<ul style="list-style-type: none"> • Microphone 	<ul style="list-style-type: none"> • Voice 	As the user speaks into the microphone, the system recognizes the voice.
<i>Tactile interface</i> (Haptic)	<ul style="list-style-type: none"> • Pressure sensor 	<ul style="list-style-type: none"> • Pressure value 	The user presses the device for a short period of time to allow the system to sense the pressure.
<i>Brain-computer interface</i> (Brain-Computer)	<ul style="list-style-type: none"> • Brainwave measurement equipment, such as electroencephalograph 	<ul style="list-style-type: none"> • Brainwave signal 	The user wears the brainwave measurement device to transmit the brainwave signal to the system.

Table 2: Output types of human-machine interfacing.

Sensory stimulation	Output device	Feedback form	Description
Vision (sight)	<ul style="list-style-type: none"> • Screen 	<ul style="list-style-type: none"> • Text • 2D/3D image • Video/motion picture • Augmented reality 	The output is composed of multimedia components, which are the most common form of feedback, and whose level sometimes is enhanced by the integration of virtual and real entities.
	<ul style="list-style-type: none"> • Projector 	<ul style="list-style-type: none"> • Text • 2D/3D image • Video/motion picture • Projection mapping 	
	<ul style="list-style-type: none"> • Head-mounted display 	<ul style="list-style-type: none"> • Augmented reality • Mixed reality 	The user puts on a head-mounted display and sees a 360° panoramic image to enhance the sense of immersion during experiencing.
	<ul style="list-style-type: none"> • Physical device such as motors and other electronic media 	<ul style="list-style-type: none"> • Movement produced by physical devices 	Through interaction, the mechanism of the physical device produces movement.
Vision (Sound)	<ul style="list-style-type: none"> • Audio electronic system • Headset • Buzzer 	<ul style="list-style-type: none"> • Digital music/sound effects • Electronic sound 	The output device generates feedbacks of music or sound effects.
	<ul style="list-style-type: none"> • Physical device such as motors and other electronic media 	<ul style="list-style-type: none"> • Sound produced by physical devices via collision 	Through interaction, the physical device is moved, and sound is generated via collision.
Tactile (Touch)	<ul style="list-style-type: none"> • Vibration switch 	<ul style="list-style-type: none"> • Vibration 	The physical device is triggered to vibrate.
	<ul style="list-style-type: none"> • Electric fan • Water mist making machine • Electronic heater 	<ul style="list-style-type: none"> • Wind • Water mist • The temperature 	The user feels a special touch.
Smell (Smell)	<ul style="list-style-type: none"> • Fragrance Machine 	<ul style="list-style-type: none"> • Smell 	The fragrance machine creates the smell.

1.2.3. COMPUTER VISION FOR HUMAN-MACHINE INTERACTION

Computer vision technology, when applied to human-machine interaction, is mainly to locate objects in images [27]. Jaimes and Sebe [25] divided this process in images by computer vision into four stages: motion segmentation, object classification, tracking, and interpretation, for which various techniques have been proposed [28], [29], [30].

In addition to requiring low-cost hardware, computer vision technology has the advantage of supporting human-machine interactive activities in *wide-range fields* [54]; the uses of matching appropriate scenarios and using computer vision techniques to implement perceptual human-machine interfaces can allow users to break hardware limitations, explore freely in the environment, and create natural experiences [11], [55]. Therefore, computer vision technology is often employed by interactive designers as a powerful design tool; participants using this technology can realize abstract interaction between the intangible and the tangible [56].

Based on Crowley, et al. [57] and Turk and Kölsch [58], the various computer vision techniques related to human beings' sensory capabilities for human-machine interaction are summarized in this study to be as shown in Table 3. From the table, it can be found that the aspects of interactive interfaces implemented by the computer vision techniques can be divided into two types, namely, *object* and *human*.

In this study, the "human" is adopted as a *unit* of the interface, for which the following capabilities with their details listed in Table 3 can be carried out to realize the function of human-machine interaction: 1) *determination of the existence and position* (of the human); 2) *decision of the body posture* (of the human); and 3) *detection and recognition of the hand movement* (of the human).

As to the case of adopting the "object" as the unit of the interface based on the *tangible interfacing* concept [31], [32] the following capabilities can be carried out: 1) *determination of the existence and position*; and 2) *recognition of information*.

Table 3: A list of computer vision techniques related to human beings' sensory capabilities.

Capability No.	Goal	Computer vision technique
Determination of the existence and position	<ul style="list-style-type: none"> Does any human or object appear in the environment? Where is the target human or object? How many humans appear in the environment? Where is the human or object located (from 2D or 3D viewpoint)? 	<ul style="list-style-type: none"> Face detection Segmentation and extraction Color detection Head and body tracking
Recognition of information	<ul style="list-style-type: none"> Who are they? What are these objects? How is the silhouette of the object? 	<ul style="list-style-type: none"> Face recognition Gait recognition Color detection Feature analysis Boundary representation
Understanding of human expressions	<ul style="list-style-type: none"> Is the human smiling, frowning, or speaking? 	<ul style="list-style-type: none"> Facial feature tracking Expression modeling and analysis
Detection of the visual focus	<ul style="list-style-type: none"> What is the direction of the human's visual focus? 	<ul style="list-style-type: none"> Head/face tracking Eye gaze tracking
Decision of the body pose	<ul style="list-style-type: none"> How is the current pose of the human's entire body? How is the current state of the human's body movement? 	<ul style="list-style-type: none"> Region segmentation and extraction Motion detection Body modeling and tracking
Detection and recognition of the hand movement	<ul style="list-style-type: none"> How is the momentum of the human's limb movement? How is the human's current gesture? How is the silhouette of the human's hand? How is the fingertip of the human's hand? 	<ul style="list-style-type: none"> Motion detection Gesture recognition Hand articulated capturing and analysis

On the other hand, the use of computer vision techniques with the *human* as the interface can achieve interesting and free interactions, so that users are no longer limited to the use of physical button interfaces, and can freely use their body postures to explore during the experiencing process.

Computer vision technology with the "human" as the interface is based on the sensing of the user's body posture to generate various interaction activities. Jaimes and Sebe [12] proposed the concept of "human-centered vision," which suggests the combination of computer vision techniques with image depth sensing and other technologies to

track the user’s body posture by using 1) the contour of the grasping posture; 2) the appearance of the grasping posture features such as skin color, face, etc.; and 3) the real-time body model generated with components such as cylinders and spheres as illustrated by Figure 1 [67]. By analyzing the shape, contour, and movement of the body, or analyzing the user’s body structure model, computer vision can be utilized to capture and analyze the human posture for various applications, like person recognition, movement analysis, etc. [65], [66].

Wang and Wu [33] derived four possible *manifestation forms* of computer vision with the *human* or *object* as the interface for computer vision-based interaction situations as shown in Table 4, where the red dotted arrows in each figure represent the image-taking directions of the cameras. In this study, an interactive game for food and agricultural education is proposed *with the human as the interface*, i.e., the interaction of the game is based on the human-machine interfacing realized by computer vision techniques.

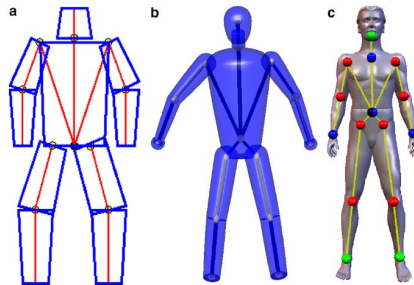


Figure 1: Human shape models with kinematic model [67].

Table 4: Manifestation forms of computer vision with “the human or object” as the interface [33].

Manifestation	Illustration	Location of the camera	Sensing by computer vision
Interactive wall		<ul style="list-style-type: none"> Put on the table, facing forward Hung right above the user on the ceiling, looking downward Hung on the wall, slightly tilted down to face the user Hung on the ceiling at a side, looking downward 	Sensing the overall posture and movement of the participant, or focusing on tracking certain parts of the body or the positions of the fingertips.
Interactive table		<ul style="list-style-type: none"> Hung right above the user Put underneath the table, looking upward 	Sensing the outer characteristics of the object on the table, followed by tracking and real-time localization of it.
Interactive floor		<ul style="list-style-type: none"> Hung right above the user, looking downward 	Sensing the presence, location, overall outline or movement of the participant.
Interactive space		<ul style="list-style-type: none"> Hung on the ceiling, right above the user Hung on the wall, slightly tilted down to face the user 	As a composite application of the interactive wall and the interactive floor, sensing the presence, location, overall outline or movement of the user based on the location of the camera.

1.2.4. INTERACTIVE EXPERIENCE AND EXISTING WORKS WITH THE “HUMAN” AS THE INTERFACE

In human-machine interaction activities, the interactive experience is based on the user’s cognition, emotion, and feeling, with the goal to promote the user’s thinking, action, and pleasure [34]. Pine and Gilmore [35] think that a “pleasant experience” should cover four states, namely, *esthetic*, *entertainment*, *educational*, and *escapist*. With the popularization of technology, it has become a trend for museums to carry out educational exhibitions that offer interactive experiences. Mitchell et al. [36] suggested that if museum education in the 21st century could integrate educators, professional knowledge, and digital media tools to form cross-field cooperation, it would be able to effectively attract more crowds to the exhibition. By simulating the situation of the teaching theme and introducing interactive experiences, exhibition education can effectively convey knowledge and enhance the interactivity between visitors and exhibitions.

In recent years, education aided by digital media technology with the “human” as the interface to provide interactive experiences has also been implemented for use in various fields [59], including several works in the form of *interactive walls*. Some examples are introduced in the following.

“TABEGAMI SAMA” (Figure 2) [60]

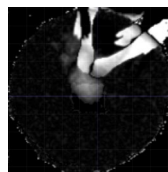
The projected interactive-wall work “TABEGAMI SAMA” (Eating God in Japanese) [60] can be explored in a darkened, immersive space to see Japanese-style food ingredients that grow up during the four seasons. By stacking rice grains into a mountain, and setting up a camera on the top to detect the rhythm and position of the participant’s hand turning the rice grains, dynamic particle effects and contour lines can be generated and projected in real-time to integrate the experience of vision, touch and smell which brings the people close to rice agriculture.

“A Nong’s Fantastic Adventure” (Figure 3) [61]

“A Nong’s Fantastic Adventure” is an interactive wall of agricultural experience. Through the use of computer vision and image synthesis technology, participants can experience the fun of planting rice and farmers’ hard farming works. In this work, Kinect is placed in front of the wall, which can sense the positions of multiple participants’ faces and feet, and synthesize their appearances wearing hats and rain boots on the front projection wall. The participants only need to simulate planting seedlings within the scope of the interactive wall, and the seedlings will be planted in front of the participants in the image shown on the wall screen.



(a)



(b)

Figure 2: “TABEGAMI SAMA” [60]. (a) A user is turning the rice grain mountain. (b) An image taken by the camera hung on the ceiling.



(a)



(b)

Figure 3: “A Nong’s Fantastic Adventure” [62]. (a) The system can synthesize the appearances of the users wearing hats and rain boots. (b) Another case.

“Fire & Ice” (Figure 4) [62]

“Fire & Ice” is a public interactive art installation with its interactive wall made up of eight screens with two opposing elements, “fire” and “ice,” allowing two participants to act as the characters of fire and ice, respectively. Two Kinects were set up on the wall, facing the body posture of the two participants at a slight oblique angle above; and computer vision techniques are used to track the participants’ skeleton nodes, and analyze their gestures and movements. When the participants stand in front of the interactive wall and begin to condense “magical energy,” if they wave their arms at the other party, condensed “magical special effects” will be launched towards the other party, forming a visual experience of ice and fire.

“University of Dayton Interactive Wall” (Figure 5) [63]

“University of Dayton Interactive Wall” shows what kind of school life a student can have when he/she enters the University of Dayton. Due to the large horizontal range of this work, a total of 4 Kinects were set up on the ceiling to widen the sensing range of the cameras, allowing the cameras to perceive whether there are pedestrians passing by. If not, the interactive wall will show the wave effect of square bricks; when pedestrians enter the sensing range of the wall, the square bricks will produce a peeling effect of the same size as the upper-view contour of the participant in front of the participant’s position. After the tiles are peeled off, videos related to school life will be displayed.

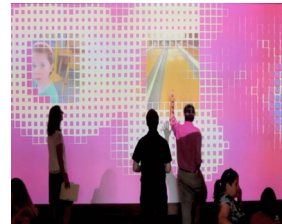


(a)



(b)

Figure 4: “Fire & Ice” [62]. (a) Two persons in actions. (b) A closer look at the person of the “fire” side



(a)



(b)

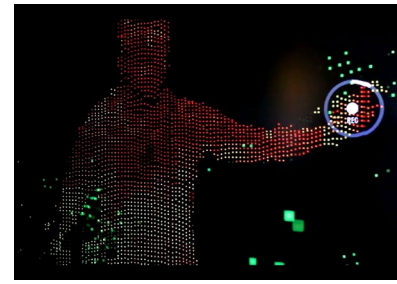
Figure 5: “University of Dayton Interactive Wall” [63]. (a) A wider view. (b) A closer view.

“NikeFuel Station” (Figure 6) [64]

“NikeFuel Station” is an interactive experience wall built in the NIKE store on which hangs a Kinect facing the participant to sense his/her movement posture. When the participant stands in front of the interactive wall, a 3D body contour consisting of particles is generated in real-time, and a video recording function can be turned on by touching the virtual button with the body contour. When the participant continues to move his/her body, the 3D outline will gradually change from red to green, conveying the idea that exercise can lead to a healthy life. After the experiencing process is over, the participant can obtain the videos recorded on their mobile phones.



(a)






(b)

Figure 6: “NikeFuel Station” [64]. (a) Generating in real-time a 3D body contour consisting of particles. (b) When the participant continues to move his/her body, the 3D outline will gradually change from red to green.

1.3. DEVICES FOR COMPUTER VISION APPLICATIONS

Computer vision technology mainly uses cameras as sensors. With the increasing demand for human-computer interaction in recent years, cameras with the additional properties of measuring depth data have been widely developed. In this study, three types of cameras used by computer vision applications are identified with their types and functions shown in Table 5.

Table 5: Types of cameras used for computer vision applications.

Camera type	Illustration	Introduction
Webcam		The cameras used in general computer video conferences which can capture color images.
Infrared filter (IR-Pass) camera		By adding an infrared filter (IR-pass filter) between the photosensitive element of the webcam and the lens to filter the visible light, the camera can only see infrared light, simplifying the image captured by the camera.
Motion sensing camera		With the 2nd-generation Kinect as an example, in addition to the built-in color camera, this type of camera also has an infrared emitter and a depth sensor, which can obtain the depth information in the taken image.

1.4. BRIEF DESCRIPTION OF THE PROPOSED SYSTEM

The above literature survey provides reviews of various concepts and case studies about food and agricultural education, human-machine interaction, and computer vision technology. Accordingly, relevant principles for designing a system for food and agricultural education were derived. The principles were followed to construct a real system on which a game can be played to learn knowledge about some fruit harvesting processes in a manner of high-freedom human-machine interaction. The system architecture is implemented by *computer vision and augmented reality techniques* utilizing an interactive wall with the ‘human’ adopted as the interface unit, as described in Table 6.

Table 6: Design of the proposed system using the “human” as the interface for food and agricultural education.

Work	Interface and interaction	Equipment	Computer vision techniques	Scenario
A game based on the use of an interactive wall, named “Fruit Picking Fun,” played on the proposed system	Experiencing by the “human” implemented by sensing the body and hand postures.	2 nd -generation motion sensing camera Kinect	<ul style="list-style-type: none"> • Face recognition and tracking • Hand posture recognition • Body model matching and tracking 	<ul style="list-style-type: none"> • Fruit picking • Fruit farm • Photo taking

2. METHODS

The methods used in this study are introduced here, including the prototyping, observation, and interview methods.

2.1. THE PROTOTYPING METHOD

Prototyping is a method for quick and low-cost evaluation of a system before it is formally constructed [42]. According to the prototyping process proposed by Naumann and Jenkins [43] and Eliason [44], an interactive prototype system for conducting food and agricultural education was constructed in this study which includes the following six major steps:

- 1) conducting a literature survey about related theories and existing systems;
- 2) deriving principles accordingly and follow them to design a prototype;
- 3) carrying out relevant experiments using the prototype;
- 4) evaluating the effectiveness of the prototype according to the users’ opinions;
- 5) improving the prototype to be a formal system; and
- 6) exhibiting the system in a public space for further testing.

2.2. THE OBSERVATION METHOD

The observation method is useful for qualitative analysis of the data collected from subtle observations of the users' performances from the perspectives of onlookers [45]. In this study, this method was adopted to collect and analyze the data collected from observations of the participants' performances of the proposed system from two aspects, "operation situation of human-machine interface" and "participant's behavior," with the detailed observation items listed in Table 7. The observation results were used for further improvement on the proposed system resulting from the prototyping method. The details will be presented later in this paper.

Table 7: The list of observation items about the participants' performances of the proposed system.

Aspect	Observation items
Operation situation of human-machine interface	<ol style="list-style-type: none"> 1) Whether the participant can understand the interactive way of the proposed system. 2) Whether the participant can operate the system smoothly.
Participant's behavior	<ol style="list-style-type: none"> 1) Whether the participants are interested in this system. 2) The feedback given by the participants in the operation of this system. 3) The participants' reaction when interacting with this system. 4) The participants' additional actions in operating this system.

2.3. THE INTERVIEW METHOD

In the interview method, invited persons are asked questions about the theme of the survey to collect objective facts from their answers [46]. This method was used in this study in two ways, namely, interviews with experts and interviews with users.

2.3.1. INTERVIEW WITH EXPERTS AS THE INTERVIEWEES

In this study, experts in related fields were invited to conduct in-depth interviews both before and after the users' experiencing activities using the proposed system. The first expert interview was conducted during the design stage of the proposed system (i.e., after the prototype was constructed and before the final system was completed). Three experts were interviewed to collect their comments about the contents for the theme of food and agricultural education and about the design for the interactive experiencing process, based on which the prototype system is improved. The second expert interview was conducted during the analysis and evaluation stage of the proposed system after the users conducted their performances of the system. Four experts were interviewed to collect their opinions on the usability, experiencing process, and education content of the proposed system.

2.3.1.1. RESULT OF THE FIRST EXPERT INTERVIEW

The result of the first expert interview is presented here, leaving that of the second interview to be described later in this paper. As shown in Table 8, the three experts invited for the first interview include an elementary school teacher, a founder of an enterprise, and a CEO of a design company. More information about their backgrounds and expertises can be found in the table as well.

Table 8: Experts interviewed during the design stage of the proposed system.

Code	Appointment Unit	Title	Expertise
P1	elementary school	teacher	child education, food and agriculture education, information technology
P2	food and agricultural education extension enterprise	founder	food and agricultural education, agricultural planting, experiencing activity of food and agricultural education
P3	interactive design company	CEO	interactive design, human-machine interaction, computer vision

Th experts were interviewed from three aspects of food and agricultural education: 1) the teaching content; 2) the experiencing activities; and 3) the introduction of computer vision and augmented reality technology. The opinions of the experts collected in the interview are listed in Table 9, from which the following conclusions can be drawn for prototype improvement:

- 1) the knowledge of food and agricultural education should be meaningful to the general public;
- 2) compared with textbook-style learning, the experience of food and agricultural education should add more than one sensory experience;
- 3) the “hands-on” sense of participation should be higher to effectively arouse the learners’ interest; and
- 4) livelier animation may be used to enhance the interactive context for people of any age to watch.

According to the above summary of the expert’s suggestions, special animation and sound effects were added into the prototype system to simulate the situations related to food and agriculture education, or specifically, related to fruit picking; and the Kinect device was adopted with the human as the interface to implement a gaming process with human-machine interaction, so as to make the interactive experience more vivid and to increase the degree of enjoyment of participants.

Table 9: The list of the opinions of the experts expressed in the first expert interview.

Aspect	Question	Opinions
Teaching content	What is the importance of learning in food and agricultural education?	<ul style="list-style-type: none"> • P1: food and agricultural education allows the public to understand the relationship between “food” and “agriculture,” and the importance of the land where food is produced. • P1 & P2: the concept of food and agricultural education is not just a slogan, but should allow the public to understand the hardships of farmers to produce food.
	How to arouse people's interest in the learning process?	<ul style="list-style-type: none"> • P1: the hands-on experiencing session can arouse interest in all age groups because there is a sense of actual participation. • P1: introduction of information technology, using mobile phones, tablets or computers to assist teaching, is also a way to arouse interest.
Experiencing activity	What is the learning effect of the experiencing activity?	<ul style="list-style-type: none"> • P2: the experience of food and agricultural education should help the user learn things that cannot be learned in textbooks, and enhance his/her autonomy, life skills, and expression capabilities.
	How to design the experiencing process?	<ul style="list-style-type: none"> • P2: more than one sensory experience can be added to the experience of food and agricultural education, and attention should be paid to the safety of the experiencing process.
Introduction of computer vision and augmented reality technology	What do you think about using digital media to simulate the experiencing process and adding interactivity?	<ul style="list-style-type: none"> • P1: this approach is at the forefront of the times and can achieve the effect of information education. • P2: The focus of the introducing interactive experiences is to interact with the user by game playing; it is necessary to consider what knowledge the user can obtain after playing.

	What details should be paid attention to when developing the interactive experiencing process by computer vision technology?	<ul style="list-style-type: none"> • P3: in the design of exhibition venues and installations, it is necessary to eliminate interference from light sources as much as possible and maintain the stability of the environment.
	For the current design of the system, are there any suggestions or what can be improved?	<ul style="list-style-type: none"> • P1: the system must show the user's game results, and add proper contents that allow the user to understand and reflect. • P2: animation can be used to present the interactive context which will be more lively for any age group to watch. • P3: it is possible to design the interactive experiencing process by use of things that modern people like.

2.3.1.2. RESULT OF THE SECOND EXPERT INTERVIEW

The four experts invited to accept the second interview are listed in Table 10, including an elementary school teacher and three university professors. They were asked questions of three aspects, namely, 1) man-machine interface operation; 2) experience content of the system; and 3) views on interactive experience of food and agricultural education, as listed in Table 11. The opinions of the four experts will be shown later in this paper when they, together with the participating users' comments obtained from a user interview, are used to evaluate the effectiveness of the proposed system.

Table 10: Experts accepting the second interview after using the proposed system.

Code	Appointment Unit	Title	Expertise
P4	an elementary school	teacher	child education, food and agriculture education, information technology
P5	dept. of digital media design at a national university	associate professor	digital learning, game design, information education
P6	dept. of digital media design at a national university	professor	situational interactive design, creative thinking, user experience
P7	dept. of creative life at a national university	assistant professor	design cognition psychology, social design, food value-added design

Table 11: The list of the questions of three aspects asked in the second expert interview.

Aspect	Question
Man-machine interface operation	<ol style="list-style-type: none"> 1) What is your opinion on the interface of the proposed system? 2) Do you think the whole experience process of this system is smooth?
Experience content of the system	<ol style="list-style-type: none"> 1) What is your opinion on the exhibition layout and digital content design of the system? 2) Do you think the system has an entertaining effect? 3) What do you think of the educational nature of the system on agricultural knowledge?
View on interactive experience of food and agricultural education	<ol style="list-style-type: none"> 1) What is your opinion on the application of "the experiencing form via the somatosensory interactive wall" of the system to food and agricultural education?

2.3.2. INTERVIEW WITH USERS AS THE INTERVIEWEES

During the exhibition of the proposed system to the public, 50 users of the system were randomly selected for interviews, aiming at collecting their comments for verifying the effectiveness of the proposed system. The “3e indicators” proposed by Yeh [24], namely, “effectiveness,” “easiness,” and “enjoyment,” as well as the “four states of pleasant experience” proposed by Pine and Gilmore [35], namely, “esthetic,” “entertainment,” “educational,” and “escapist” were adopted to design the questions asked in the interview process, resulting in a set of questions of three aspects: “operation situation of human-machine interface,” “operation experience,” and “views on interactive experience of food and agricultural education,” where

- 1) the first aspect of “operation situation of human-machine interface” comes from the two indicators of “easiness” and “effectiveness”;
- 2) the second aspect of “operation experience” comes from the “enjoyment” indicator and the “esthetic” and “educational” states; and
- 3) the last aspect of “views on interactive experience of food and agricultural education” is aimed at covering the remaining “entertainment” and “escapist” states.

The questions so designed are listed in Table 12 while the comments collected from the users accepting the interview will be presented later in this paper.

Table 12: List of the questions of three aspects asked in the interviews with the users.

Aspect	Question
operation situation of human-machine interface	<ol style="list-style-type: none"> 1) What do you think about the use of body movements to interact? 2) What is your opinion on the operation interface of this system?
operation experience	<ol style="list-style-type: none"> 1) What is your opinion on the layout of the exhibition or the design of the digital content of this system? 2) What do you think of the educational nature of this system on agricultural knowledge? 3) What are your thoughts and feelings while playing this work?
views on interactive experience of food and agricultural education	<ol style="list-style-type: none"> 1) What is your view on applying the experience form of somatosensory interaction to food and agricultural education?

3. RESULTS

The details about the construction of the proposed system are described in this section, including the design idea, architecture, hardware, software, and game-play process.

3.1. IDEA FOR DESIGNING THE PROPOSED SYSTEM

In the countryside, *fruit picking* in most sightseeing farms allows the public to participate in related agricultural activities, providing a good way of food and agricultural education. During such a kind of agricultural production experiencing process, in addition to being able to understand the agricultural process of fruit planting and harvesting, it can also leave a good memory of farmers’ daily tasks.

However, the traditional fruit picking activity is limited by the plant growth time and the available farm space, and so cannot be experienced by too many people at all time. Therefore, in this study, it is desired to simulate the situation of harvesting fruits in the farmland *by a somatosensory interactive game-play method*, hoping to break the limitations of time and space. Specifically, the activities of picking three kinds of fruit, namely, *banana, orange, and cantaloupe*, is implemented via computer vision technology in this study to offer an interesting interactive experiencing process, followed by the action of taking an augmented-reality photo, so as to keep the happy time in memory as well as to shorten the distance between the agricultural industry and the public.

3.2. THE USE OF THE INTERACTIVE KINECT DEVICE AND THE DESIGN OF THE EXPERIENCING PROCESS

The design of the proposed system with an interactive game called “Fruit Picking Fun,” as illustrated by Figure 7, is based on the uses of the second-generation somatosensory camera Kinect as the sensor and a series of related computer programs written in this study, allowing the participant to use various body postures to play the game, by which the participant can experience the food and agricultural education of simulated fruit-picking activities in farmlands as mentioned previously. Besides the Kinect device, the system also includes an LED display screen and a loud speaker, both connected to a computer for performing the tasks of image taking, computer vision processing, graphic displays, data transmission, message announcing, augmented-reality photo taking and downloading, etc.



Figure 7: Illustration of the design of the proposed interactive system with a game called “Fruit Picking Fun.”

3.3. DESIGN OF THE FRUIT-PICKING EXPERIENCING PROCESS

The game of the proposed system implemented in this study for interactive experiencing of fruit-picking activities is played in the following way.

Stage 1: initialization and AR-based dressing up

At the beginning, the system displays an initial screen, with a message of inviting the participant to select, by a *grabbing gesture* (extending a hand to grab a fruit in front the Kinect), one of the above-mentioned three kinds of fruit that they want to “harvest,” as shown in Figure 8(a). Then, the system displays a farm scene corresponding to the selected fruit with the participant appearing in the middle of a group of trees of the selected fruit, and a *hat-like object* (a fruit-tree leaf, a decorated hat, or a colored cap) is generated to appear over the participant’s head to dress him/her up to be like a farmer, like the example shown in Figure 8(b). The artificial object is *fixed* on the participant head even when he/she is moving around, and this *augmented reality (AR)* effect of dressing up the participant is realized by a *face detection and tracking* program written in this study.

Stage 2: fruit harvesting

The participant starts to harvest the selected fruit in this stage by making certain *body actions* and *hand operations*, as illustrated by the example shown in Figure 8(c) where the participant is harvesting bananas. The body actions and hand operations designed for this purpose simulate the *cutting*, *shaking*, and *picking* activities conducted by farmers in real fruit-harvesting situations. Graphics of such actions and operations of the participant are overlapped on the image of the background wall in the exhibition space by augmented reality techniques, and the results are shown on the display screen. For Figure 8(c), the body actions and hand operations include *moving the body* and *cutting by hands using knives*. Harvesting is judged to be successful after the body action is carried out for a certain number of times. Each type of body action is displayed according to certain *body models* and *hand gestures* with specific parameters, whose images can be recognized and tracked by the Kinect device using some written programs that carry out the operations of *body-model matching and tracking* and *hand-gesture recognition*.

Stage 3: AR photo taking and downloading

After the harvest is successful, the participant can take a digital photo of him/herself with the harvested fruit held by the two hands as shown by the example of Figure 8(d). This photo is generated by augmented reality techniques again and is kept in the computer. Furthermore, the photo may be downloaded to the user’s mobile phone

by scanning a QR code appearing on the display screen using the phone, as illustrated in Figure 8(e). This photo-download operation is carried out by a commercially-available QR code identification program.

Stage 4: game restarting

The user shows a “T-shaped pose” to trigger the system to go back to the initial screen, meaning that the game is started over again. Recognition of such a posture is carried out also by the *body-model matching and tracking* programs mentioned previously for Stage 2.

3.4. ARCHITECTURE OF THE PROPOSED SYSTEM

As shown in Figure 9, with the Kinect device used to capture the participant’s motion images, the game engine Unity3D is utilized in this study for multimedia integration and development for the game “Fruit Picking Fun” played on the the proposed system, and the Microsoft Kinect for Windows SDK (hereinafter referred to as the Kinect SDK) is used to develop the computer programs used in the proposed system.

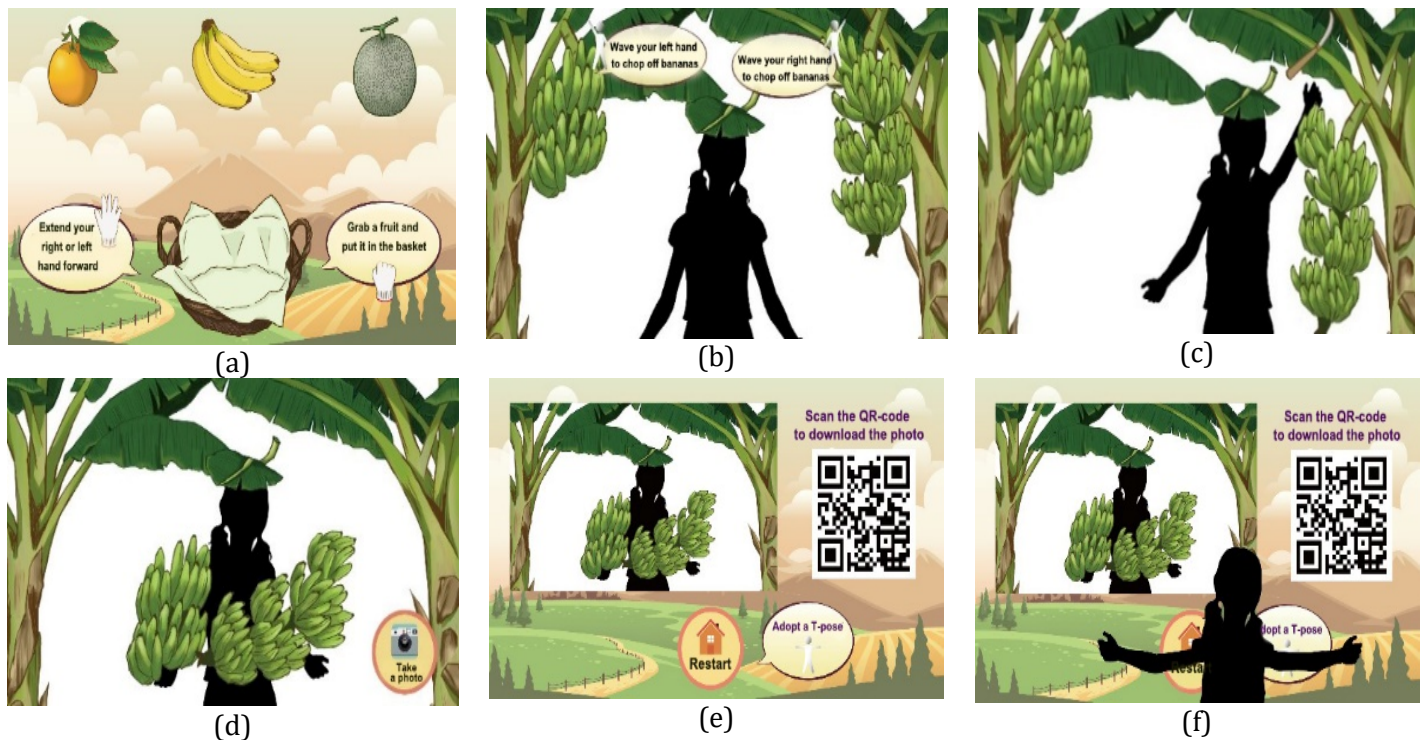


Figure 8: Illustration of the experiencing process of the proposed system with the game “Fruit Picking Fun.” (a) The initial screen for fruit selection. (b) Dressing up the user as a farmer with a hat-like object over the head. (c) Harvesting the fruit by body actions and hand operations. (d) Holding the harvested fruit for photo taking. (e) Inviting the participant to scan the QR code to download the “taken” photo into his/her mobile phone. (f) Showing a “T-shaped pose” to trigger the game back to the initial screen.

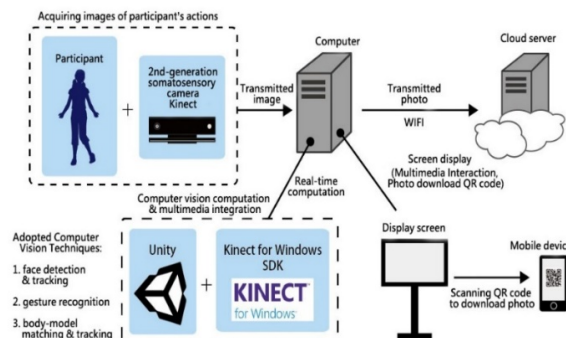


Figure 9: Illustration of the architecture of the proposed system with the game “Fruit Picking Fun.”

The second-generation Kinect as shown in the last row of Table 5 is used as a somatosensory camera, which is composed of a depth sensor, an RGB camera, and a microphone array with four units of microphones. By the sensors of the Kinect, color, 3D depth, and infrared images, as well as sound information of the target object, can be acquired.

Furthermore, via written programs of the Kinect SDK associated with the Kinect device, the functions of human body tracking and body skeleton identification can be implemented by use of the three-dimensional coordinates of up to 2525 joint points of the human body and fingers obtained from the acquired images. These functions can be used to implement the tasks of *face detection and tracking*, *hand-gesture recognition*, and *body-model matching and tracking* needed to implement the previously-mentioned four stages of actions involved in the fruit-picking activity.

After the participant has taken a digital photo after the experiencing process, the system will immediately transmit the photo to the cloud server via a wifi channel, and generate a QR code which includes a photo-download link. The participant can download the photo into his/her mobile phone by scanning the QR code with the phone.

3.5. IMPLEMENTED COMPUTER VISION TECHNIQUES WITH THE HUMAN AS THE INTERFACE

In this study, the Kinect device (shown in the third row of Table 5) and the Kinect SDK are used to implement the computer vision techniques used in the proposed system for playing the game of “Fruit Picking Fun” with the “human” as the interface. The Kinect device consists of a depth sensor, an RGB color camera, and a four-unit microphone array. These sensors can be used to obtain color images, 3D depth images, infrared images, and audio information. Combined with computer codes written in the Kinect SDK language, the Kinect device can be used to recognize and track many features of the human body and hands appearing in the color image taken by the Kinect. Specifically, it can extract up to 25 joint points of the human body and fingers and their three-dimensional coordinates.

In this study, the Kinect device and its related functions are used to implement the previously-mentioned functions of “face detection and tracking,” “hand-gesture recognition,” and “body-model matching and tracking” for implementing the experiencing process of “Fruit Picking Fun.” The details of the implementation results are described in the following.

3.5.1. FACE DETECTION AND TRACKING FOR AFFIXING A HAT-LIKE OBJECT OVER THE USER'S HEAD

In this study, the color and depth images acquired with the Kinect device itself are used to detect and track the participant’s face so as to implement the desired AR function of imposing a hat-like object over the participant’s head, as illustrated by the example of the banana leaf appearing in Figures 8(b)~8(d).

A program named F1 was written in this study to realize this AR function which is based on using the Kinect device to conduct the *detection* of the bone joint points of the head, followed by the *tracking* of them via the use of the location coordinates and rotation values of these points. As an example, the rectangular frame drawn in Figure 10 is the human face detected and tracked by this program F1. A more detailed pseudo-code algorithm describing this program is shown in Table 13. In the table, the italic characters like *detect* are used to specify the action conducted by the Kinect device, and the bold characters like **repeat** or **while** are commands used by the pseudo-codes.

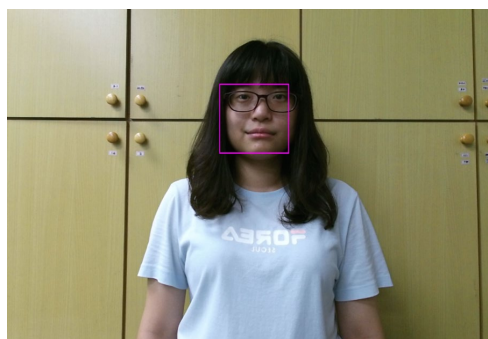


Figure 10: An example of face detection and tracking result carried out by the proposed system with the game “Fruit Picking Fun.”

Table 13: The algorithm (F1) for face detection and tracking to affix a hat-like object over the user's head.

Label & Function	Algorithm
F1: affixing a hat-like object over the user's head	<p>Input: (1) color image i_c and depth image i_d acquired by the Kinect device; and (2) the vertical offset v_{off} of the 3D position of the hat-like object h_o with respect to the user head.</p> <p>Output: the new 3D position p_n of the hat-like object h_o.</p> <p>Steps. Step 1: repeat <i>detect</i> the user's face in image i_c by the Kinect using the images i_c and i_d until detected. Step 2: <i>track</i> the user's face and compute the 3D position of the user's face as p_n. Step 3: if $v_{off} \neq 0$ then add $(0, v_{off}, 0)$ to p_n; end if. Step 4: exit with p_n as the new 3D position of the hat-like object h_o over the user's head.</p>

3.5.2. BODY-MODEL MATCHING AND TRACKING

Some procedures of the Kinect SDK are used in this study to track the user's body joint points and obtain the coordinate information of each joint point for use in the interactive activities of playing "Fruit Picking Fun" on the proposed system. Only the posture of the user's upper body is needed in this study, from which 10 joint points are detected and tracked as shown in Figure 11, including: 1) J1 – head; 2) J2 – spine shoulder; 3) J3 – right shoulder; 4) J4 – left shoulder; 5) J5 – right elbow; 6) J6 – left elbow; 7) J7 – right wrist; 8) J8 – left wrist; 9) J9 – right hand; and 10) J10 – left hand.

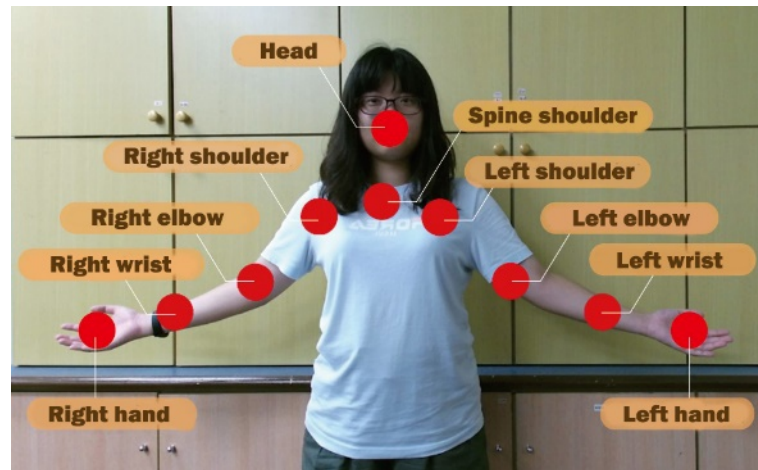


Figure 11: The user's body joint points detected and tracked in this study using Kinect SDK procedures for playing the game "Fruit Picking Fun" on the proposed system.

3.5.3. HAND-GESTURE RECOGNITION

By use of the Kinect SDK procedures, the human-body joint points can be detected using the Kinect device, and the 3D coordinates of these points can be acquired. A function of hand-gesture recognition has been implemented in this study as an algorithm using the relative 3D positions between these joint points computed in terms of the acquired 3D coordinates of these points. The algorithm is named G0 as described in Table 14, in which the set of parameters of the human-body joint points detected from the input images acquired by the Kinect device is matched

sequentially with the conditions of the various hand gestures listed in Table 15. The output of the algorithm is a *hand-gesture event* expressed currently by the user for use in other algorithms implemented in this study for the game “Fruit Picking Fun.”

It is noted by the way that the hand-gesture procedures G1 through G7 in Table 15 are originally built by use of the Kinect SDK for the Kinect device but with their conditional parameters modified to fit the applications of this study, while G8 through G10 are *new* hand-gesture procedures created in this study.

The above-mentioned hand-gesture recognition procedures, or simply hand-gesture events, are used in the interactions between the *game process control* and the *fruit picking scenes*. By tracking the coordinate values of the body joints, the effect of *touching objects* can also be achieved.

Table 14: The algorithm (G0) of hand-gesture recognition for use in the proposed system.

Label & Function	Algorithm
G0: recognizing the hand gesture of the user	<p>Input: (1) the color image i_c and depth image i_d acquired by the Kinect device; and (2) conditions G1~G10 for the hand-gesture events shown in Table 15.</p> <p>Output: the gesture event g_e of the user.</p> <p>Steps. Step 1: <i>detect</i> the user’s body joint points by the Kinect device to form a set p_b. Step 2: repeat match the parameters of p_b with conditions G1 through G10; until a condition G_i is matched. Step 3: exit with the procedure corresponding to G_i in the 2nd column of Table 15 as the desired gesture event g_e.</p>

Table 15: Gesture event procedures implemented by the Kinect SDK for use in the game “Fruit Picking Fun.”











Label	Gesture event (Name of the procedure)	Illustration	Gesture condition
G1	Left/right hand grip (Hand Event Type. Grip)		Left & right hand states: five fingers clenched
G2	Left/right hand opening (Hand Event Type. Release)		Left & right hand states: the palm is open with the five fingers spread out
G3	T-shaped pose (Gestures. T pose)		(1) The heights of the right hand, right elbow, and right shoulder should not differ more than 10 cm. (3) The heights of the left hand, left elbow, and left shoulder should not differ more than 10 cm. (2) The distance seen from the front between the right hand and the right shoulder does not exceed 15 cm. (4) The distance seen from the front between the left hand and the left shoulder does not exceed 15 cm. (5) Conditions (1)~(4) are maintained for 0.3 seconds.
G4	Raising two hands over shoulders (Gestures. Psi)		(1) The right hand is more than 10 cm above the spine. (2) The left hand is more than 10 cm above the spine. (3) Conditions (1) and (2) are maintained for 0.3 seconds.

Table 15: (Continued).

G5	Raising up either hand and swiping (Gestures. Swipe Up)		<p>Stage 1 – either of the following two cases: (1) the right hand is lower than the left elbow for less than 15 cm; (2) the left hand is lower than the right elbow for less than 15 cm. Stage 2 (conducted within 1.5 sec.) – either of the following two cases: (1) the right hand is higher than the left shoulder by more than 5 cm, and the parallel movement of the right hand does not exceed 10 cm; (2) the left hand is higher than the right shoulder by more than 5 cm, and the parallel movement of the left hand does not exceed 10 cm.</p>
G6	Raising up the right hand (Gestures. Raise Right Hand)		(1) The right hand is higher than the left shoulder by more than 10 cm. (2) The left hand is lower than the left shoulder.
G7	Raising up the left hand (Gestures. Raise Left Hand)		(1) The left hand is higher than the right shoulder by more than 10 cm. (2) The right hand is lower than the right shoulder.
G8	Holding objects by two hands (Gestures. Raise Half Hands)		(1) The distance between the right/left hand and the right/left elbow is 15 to 80 cm. (2) The parallel movement of the right/left hand should not exceed 15 cm from the right/left shoulder. (3) The height of the right/left elbow from the right/left hand should not exceed 15 cm. (4) The right/left hand is higher than the hip.
G9	Raising the right hand flat (Gestures. Hold Right Hand Up)		(1) The height of the right shoulder should not exceed 10 cm from the right hand. (2) The height of the right shoulder should not exceed 10 cm from the right elbow. (3) The distance seen from the front between the right hand and the right shoulder should not exceed 15 cm. (4) The parallel movement of the left hand is not more than 20 cm away from the left shoulder. (5) The parallel movement of the left elbow should not exceed 20 cm from the left shoulder. (6) The distance seen from the front between the left hand and the left shoulder should not exceed 20 cm.
G10	Raising the left hand flat (Gestures. Hold Left Hand Up)		(1) The height of the left shoulder should not exceed 10 cm from the left hand. (2) The height of the left shoulder should not exceed 10 cm from the left elbow. (3) The distance seen from the front between the left hand and the left shoulder should not exceed 15 cm. (4) The parallel movement of the right hand is not more than 20 cm away from the right shoulder. (5) The parallel movement of the right elbow should not exceed 20 cm from the right shoulder. (6) The distance seen from the front between the right hand and the right shoulder should not exceed 20 cm.

3.6. INTERACTION PROCESS AND PROCESS CONTROL ALGORITHMS

The interactive process of the game “Fruit Picking Fun” is divided into 11 parts, including three farm scenarios of the fruits of banana, orange and cantaloupe. The interactive scenario process is shown in Table 16. Each step in the table can be implemented by the algorithms in Table 13, Table 17, and Table 18, where Table 13 has been presented before for user face recognition and tracking, and Tables 17 and 18 are presented subsequently with the former including the algorithms for flow control of the game “Fruit Picking Fun,” and the latter including the algorithms for interactions with the farm scenes.

Table 16: The illustration of the interactive scenario process of the game “Fruit Picking Fun.”





Stage	Scenario	Illustration	Explanation	Steps in the algorithm
1	Selecting the fruit in the initial screen		Initial screen: a farm scene with three types of fruit appearing on the top and a basket shown below. Messages: inviting the user to “extend a hand to grab a fruit” and put it into the basket to enter the experiencing farm scene of the selected fruit.	S1 S2
2-A	Experiencing a banana farm		Screen: banana trees appear around, and a banana leaf appears on the user’s head for use as a hat. Messages: inviting the user to “wave his/her left and right hands,” respectively, to chop off the bananas.	F1
2-B	Experiencing an orange farm		Screen: orange trees appear around, and a hat decorated with orange leaves appears on the user’s head. Message: inviting the user to “flick his/her arms up to touch the oranges to harvest them.”	F1
2-C	Experiencing a cantaloupe farm		Screen: cantaloupe scaffoldings appear around, and a cantaloupe-colored cap appears on the user’s head. Message: inviting the user to “wave hands to cut the cantaloupes to let them bounce for a fixed number of times to harvest them.”	F1

Table 16: (Continued).






<p>3-A</p>	<p>Harvesting bananas</p>		<p>Action: the user waves the right or left hand with a knife to cut off the banana bunch on the tree.</p> <p>How to pass this level: the harvest is successful when each of the left and right banana bunches is cut three times and falls.</p>	<p>S3</p>
<p>3-B</p>	<p>Harvesting oranges</p>		<p>Action: the user flicks his/her arms up to touch the oranges on the tree like shaking them off the tree.</p> <p>How to pass this level: the harvest is successful when all the oranges have been shaken off.</p>	<p>S4</p>
<p>3-C</p>	<p>Harvesting cantaloupes</p>		<p>Action: the user touches the cantaloupes to pick them from the vine, and makes the falling cantaloupes bounce.</p> <p>How to pass this level: the harvest is successful when the cantaloupes are picked off and touched to create more than 20 bounces.</p>	<p>S5</p>
<p>4</p>	<p>Returning to the initial screen</p>		<p>Action: the user shows a “T-shaped pose” with both hands held to the sides for a period of time.</p> <p>Result: the system returns to the initial screen.</p>	<p>C1</p>
<p>5</p>	<p>Showing the harvest result before photo taking</p>		<p>Action: the user stretches two hands forward for a period of time.</p> <p>Result: the system shows a graphic of the harvesting result over the user’s hands; and then enters a countdown stage for photo taking.</p>	<p>C2 S6</p>

Table 16: (Continued).


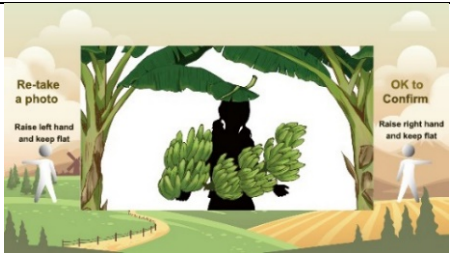
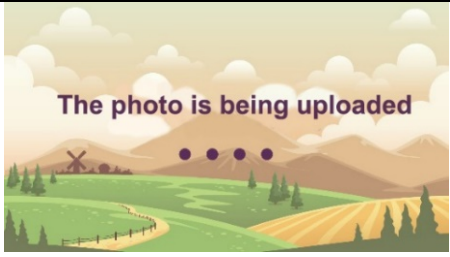


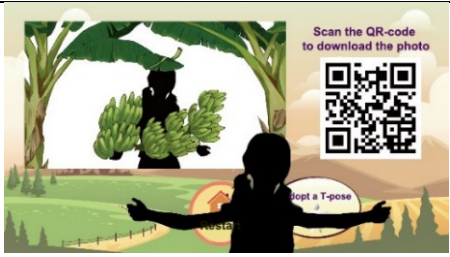
6	Countdown for photo taking		<p>Action: the user has five seconds to pose properly for photo taking with the harvested crops appearing over his/her hands.</p> <p>Result: the system takes a photo of the user's posture after a 5-second countdown.</p>	C2 S6
7	Retaking or confirming in the photo taking process		<p>Screen: a photo preview appears in the middle.</p> <p>Messages: inviting the user to select one of the two choices, "retake a photo by raising the left hand flat" or "OK to confirm by raising the right hand flat."</p>	C3 C4
8	Transmitting the photo to the cloud server		<p>Screen: a countryside with mountains far away.</p> <p>Message: "The photo is being uploaded."</p>	None
9	Inviting to download the photo		<p>Screen: the photo and a QR code appear.</p> <p>Messages: inviting the user to "scan the QR code to download the photo" using a mobile phone, and then "adopt a T-pose" to return to the initial screen (see Step 11).</p>	None
10	Downloading the photo		<p>Action: the user scans the QR code on the screen by a mobile phone to download the photo into the phone.</p> <p>Result: the system downloads the photo to the user's mobile phone.</p>	None
11	Going back to the initial screen		<p>Action: either the user shows a "T-shaped pose" or no human motion is detected for 10 seconds.</p> <p>Result: the system returns to the initial screen.</p>	C1 S7

Table 17: Algorithms of the process control for the game “Fruit Picking Fun” played on the proposed system.

Label & Function	Algorithm
<p>C1: returning to the initial screen – maintaining the T-pose action to return to the initial screen</p>	<p>Input: (1) color image i_c and depth image i_d acquired by the Kinect device; (2) the procedures described in Table 15. Output: the instruction for returning to the initial screen. Steps. Step 1: repeat <i>detect</i> the T-shaped pose of the user’s body by the Kinect device using the procedure G3: Gestures. Tpose as well as the input images i_c and i_d; until detected successfully for three times consecutively. Step 2: exit with the instruction for returning to the initial screen.</p>
<p>C2: triggering photo taking – maintaining holding objects in hand to trigger photo taking</p>	<p>Input: (1) color image i_c and depth image i_d acquired by the Kinect device; (2) the procedures described in Table 15. Output: the instruction for photo taking. Steps. Step 1: repeat <i>detect</i> the user’s posture of holding objects by two hands by the Kinect device using the procedure G8: Gestures.RaiseHalfHands as well as the input images i_c and i_d; until detected successfully for three times consecutively. Step 2: exit with the instruction for photo taking.</p>
<p>C3: triggering photo downloading – maintaining raising up the right hand or keeping it flat afterwards to trigger photo downloading</p>	<p>Input: (1) color image i_c and depth image i_d acquired by the Kinect device; (2) the procedures described in Table 15. Output: the instruction for photo downloading. Steps. Step 1: repeat <i>detect</i> the user’s posture of raising up the right hand or keeping it flat afterwards by the Kinect device using the procedure G6: Gestures.RaiseRightHand or the procedure G9: Gestures.HoldRightHandUp, respectively, as well as the input images i_c and i_d; until detected successfully for three times consecutively. Step 2: exit with the instruction for photo downloading.</p>
<p>C4: triggering photo re-taking – maintaining raising up the left hand or keeping it flat afterwards to trigger photo re-taking</p>	<p>Input: (1) color image i_c and depth image i_d acquired by the Kinect device; (2) the procedures described in Table 15. Output: the instruction for photo re-taking. Steps. Step 1: repeat <i>detect</i> the user’s posture of raising up the left hand or keeping it flat afterwards by the Kinect device using the procedure G7: Gestures.RaiseLeftHand or the procedure G10: Gestures.HoldLeftHandUp, respectively, as well as the input images i_c and i_d; until detected successfully for three times consecutively. Step 2: exit with the instruction for photo downloading.</p>

Table 18: Algorithms of the user’s interactions with the farm scenes for the game “Fruit Picking Fun.”

Label & Function	Algorithm
<p>S1: selecting a fruit – detecting the hand gesture of clenching fingers, tracking the hand, and dragging the selected fruit</p>	<p>Input : (1) color image i_c and depth image i_d acquired by the Kinect device; (2) the procedures described in Table 15; (3) fruit objects $O_1, O_2,$ and O_3 of the three kinds of fruit and their original positions $P_1, P_2,$ and $P_3,$ respectively. Output: a new position P_i' of a selected object O_i. Steps. Step 1: repeat <i>detect</i> the joint point of the “left (or right) hand” (J9 or J10) as shown in Figure 11 by the Kinect device using the input images i_c and i_d; until detected. Step 2: if the left (or right)-hand joint point (J9 or J10) is moving forward then extract the coordinates (x, y) of the left (or right)-hand joint point; end if. Step 3: repeat <i>detect</i> the user’s posture of clenching five fingers by the Kinect device using the procedure G1: HandEventType.Grip as well as the input images i_c and i_d; until detected. Step 4: perform the following operations: 4.1 find the fruit object O_i close to the coordinates (x, y) of the left (or right)-hand joint point within a preset tolerable range. 4.2 <i>track</i> the joint point of the “left (or right) hand” by the Kinect device and extract its coordinates (x', y'). 4.3 compute the new position P_i' of fruit object O_i according to the coordinates (x', y') and the original position P_i of O_i. Step 5: exit with P_i' as the new position of the detected fruit object O_i shown on the display screen.</p>
<p>S2: displaying the farm scenes of the selected fruit – detecting opening of the user’s palm to display the farm scene of the selected fruit</p>	<p>Input : (1) color image i_c and depth image i_d acquired by the Kinect device; (2) the procedures described in Table 15; (3) a fruit object O_i selected by S1 above and its position P_i' computed by S1; (4) the basket object B. Output: the farm scene of the selected fruit object O_i or going back to the initial screen. Steps. Step 1: repeat <i>detect</i> the user’s posture of opening his/her palm with the five fingers spread out by the Kinect device using the procedure G2: HandEventType.Release as well as the input images i_c and i_d; until detected.</p>

	<p>Step 2: if the position P_i' of fruit object O_i is not close to the basket object B within a preset tolerable range then reset the position P_i' of fruit object O_i to its original value P_i and goto S1 in this table; else case based on O_i: case “O_1”: exit by displaying the farm scene of the banana fruit; case “O_2”: exit by displaying the farm scene of the orange fruit; case “O_3”: exit by displaying the farm scene of the cantaloupe fruit; end case; end if.</p>
<p>S3: cutting off the banana bunches – detecting the joint point of the left (or right) hand, judging if the banana bunch on either side has been touched for three times, and cutting the bunch off if so.</p>	<p>Input: (1) color image i_c and depth image i_d acquired by the Kinect device; (2) two banana bunches BB_{left} and BB_{right}; and (3) animations A_{left} and A_{right} of cutting off BB_{left} and BB_{right}, respectively. Output: instruction of cutting BB_{left} and BB_{right} off. Steps. Step 1: if both banana bunches BB_{left} and BB_{right} have been cut off then exit; end if. Step 2: if BB_{left} is not cut off then perform the following operations: 2.1 repeat <i>detect</i> the joint point of the left hand (J9) by the Kinect device using the input images of i_c and i_d; until detected; 2.2 repeat <i>track</i> the joint point P of the left hand (J9); until P is found to be close to BB_{left} for three times; 2.3 show the animation A_{left} of cutting off the left banana bunch BB_{left}; and goto Step 1; end if. Step 3: if BB_{right} is not cut off then perform the following operations: 3.1 repeat <i>detect</i> the joint point of the right hand (J10) by the Kinect device using the input images of i_c and i_d; until detected; 3.2 repeat <i>track</i> the joint point P of the right hand (J10); until P is found to be close to BB_{right} for three times; 3.3 show the animation A_{right} of cutting off the right banana bunch BB_{right} and goto Step 1; end if.</p>
	<p>Input: (1) color image i_c and depth image i_d acquired by the Kinect device; (2) the procedures described in Table 15;</p>

<p>S4: shaking off the oranges – detecting any posture that shakes off the oranges, and bouncing them if touched</p>	<p>(3) the set SO of oranges on the tree; and (4) animations $A_1, A_2,$ and $A_3,$ each of shaken-off 1/3 oranges in SO; and animation A_b of bouncing an orange. Output: animations A_1, A_2, A_3 and A_b. Steps. Step 1: set $i = 0$. Step 2: if all the oranges in SO have been shaken off then exit; else set $i = i + 1$; end if. Step 3: repeat <i>detect</i> all the user's 10 body joint points by the Kinect device using the input images of i_c and i_d; until detected. Step 4: repeat <i>track</i> the 10 joint points and <i>detect</i> the following user's postures by the Kinect device using the joint points and the procedures listed in Table 15: (a) raising two hands over shoulders by G4: Gestures.Psi; (b) raising up either hand and swiping by G5: Gestures.SwipeUp; (c) raising up the right hand by G6: Gestures.RaiseRightHand; and (d) raising up the right hand by G7: Gestures.RaiseLeftHand; until the above gestures are detected for three times. Step 5: perform the following operations: 5.1 show the animation A_i of shaking off 1/3 of the set SO of oranges on the display screen; 5.2 <i>track</i> the the joint point P of the left (or right) hand (J9 or J10); 5.3 if P touches any of the falling oranges, denoted as $O_i,$ then show the animation A_b of bouncing the orange O_i; end if; 5.4 goto Step 1.</p>
<p>S5: picking the cantaloupes – detecting the user's hand operations to pick and bounce the cantaloupes</p>	<p>Input: (1) color image i_c and depth image i_d acquired by the Kinect device; (2) two sets of cantaloupes CT_{left} and CT_{right} on the left and right sides of the screen; and (3) animations A_{left} and A_{right} of bouncing cantaloupes picked from CT_{left} and $CT_{right},$ respectively, and animation A_b of bouncing a cantaloupe. Output: animations A_{left} and A_{right}. Steps. Step 1: if both sets of cantaloupes CT_{left} and CT_{right} have been picked then goto Step 4; end if.</p>

	<p>Step 2: if CT_{left} is not picked then perform the following operations: 2.1 repeat <i>detect</i> the joint point of the left hand (J9) by the Kinect device using the input images of i_c and i_d; until detected; 2.2 repeat <i>track</i> the joint point P of the left hand (J9); until P is found to be close to a cantaloupe in CT_{left}; 2.3 show the animation A_{left} of dropping the cantaloups in CT_{left}; 2.4 goto Step 1. end if. Step 3: if CT_{right} is not picked then perform the following operations: 3.1 repeat <i>detect</i> the joint point of the right hand (J10) by the Kinect device using the input images of i_c and i_d; until detected; 3.2 repeat <i>track</i> the joint point P of the right hand (J10); until P is found to be close to a cantaloupe in CT_{right}; 3.3 show the animation A_{right} of dropping the cantaloups in CT_{right}; 3.4 goto Step 1; end if. Step 4: repeat (a) <i>track</i> the the joint points P of the left or right hand (J9 or J10); (b) if P touches any of the falling cantaloupes, denoted as T_i, then show the animation A_b of bouncing the cantloupe T_i; end if; until the number of touches to bounce the cantaloupes reaches 20. Step 5: exit.</p>
<p>S6: displaying the harvest result – tracking the left and right hands, and displaying the harvest result over the hands.</p>	<p>Input: (1) color image i_c and depth image i_d acquired by the Kinect device; (2) graphics of the harvest result of the selected type of fruit: (2.1) for bananas – a graphic G_{left} of a banana bunch for the left hand; and a graphic G_{right} for the right hand; (2.2) for oranges – a single graphic G_p of a pile of oranges; (2.3) for cantaloupes – a single graphic G_c of a group of cantaloupes. Output: the graphic of the selected fruit shown on the user’s hands. Steps. Step 1: repeat</p>

	<p><i>detect</i> the joint points of both the left and right hands (J9 and J10) by the Kinect device using the input images of i_c and i_d; until detected. Step 2: repeat (a) <i>track</i> the the joint points P_{left} and P_{right} of the left and right hands (J9 and J10), respectively; (b) case based on the type of selected fruit: case “banana”: (a) display the graphic G_{left} of the left banana bunch over the joint point P_{left}; (b) display the graphic G_{right} of the right banana bunch over the joint point P_{right}; case “orange”: display the graphic G_p of the orange pile over the middle point of P_{left} and P_{right}; case “cantaloupe”: display the graphic G_c of the cantaloupe group over the middle point of P_{left} and P_{right}; end case; until a 5-second countdown is ended. Step 3: exit.</p>
<p>S7: returning to the initial screen – when no user is detected for more than 10 seconds.</p>	<p>Input: (1) color image i_c and depth image i_d acquired by the Kinect device while the screen display shows the QR code; (2) a time counter T in unit of second. Output: display of the initial screen or doing nothing. Steps. Step 1: set time counter $T = 0$. Step 2: while no human motion is detected and $T < 10$ do (a) wait for a second and set $T = T + 1$; (b) <i>detect</i> all the user’s body joint points J_1 through J_{10} by the Kinect device using the input images i_c and i_d; (c) if none of J_1 through J_{10} is detected then decide “no human moiton is detected”; else exit; end if; end while. Step 3: display the initial screen and exit.</p>

3.7. PUBLIC EXHIBITION

The system was exhibited in a public space in a university for 10 days and visitors were invited to be participants in the study. The exhibition space is large enough to allow participants to interact with the system using body postures and conduct photo taking. Some pictures taken in the exhibition space are shown in Figure 12.

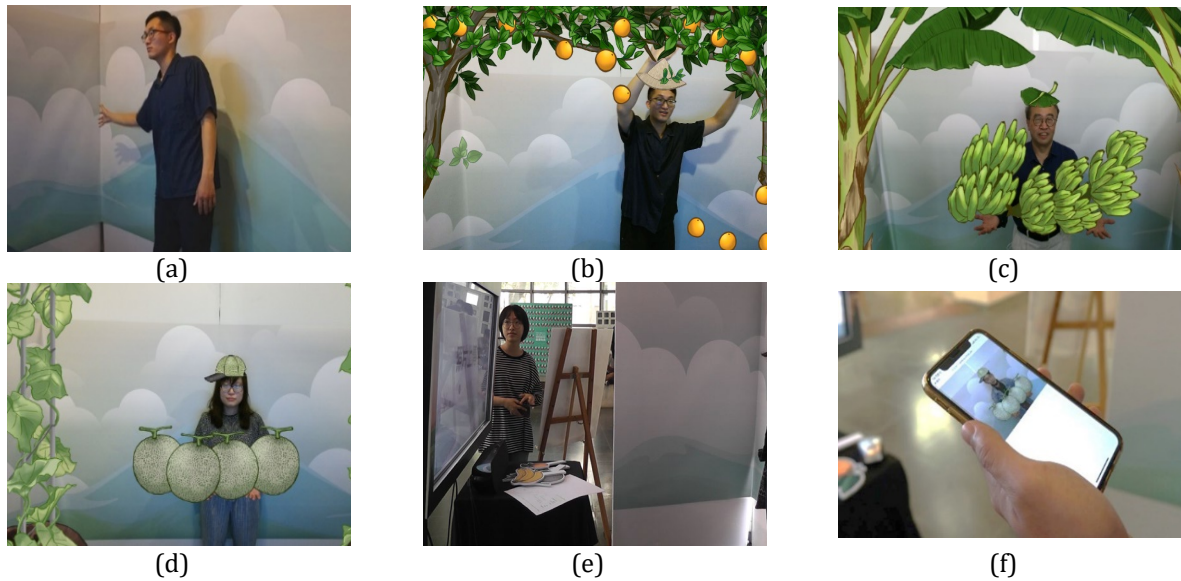


Figure 12: Some pictures taken in the exhibition space. (a) A participant is interacting with the system. (b) The participant of (a) was “shaking” the orange tree to harvest the oranges. (c) Another participant was waiting for photo taking with harvested bananas in two hands. (d) A third participant are waiting for photo taking with the harvested cantaloupes in hands. (e) A fourth participant was waiting for the taken photo to be uploaded to the server. (f) A photo has been downloaded to a participant’s mobile phone.

4. ANALYSIS

The proposed system was exhibited publicly in the design museum of a national university in May, 2020 for 10 days. People from the public older than the age of 18 and four experts were invited to experience the proposed system. Each participant’s experiencing activity lasts about 25 minutes, including 5 *minutes* to introduce the system to the participant, 10 minutes for carrying out the game-play process, and 10 minutes for an interview with the user. During the game-playing process, each user’s interaction was observed and recorded. Afterwards, the four experts and 50 randomly-selected users were invited further for interviews with their comments collected.

4.1. ANALYSIS OF OBSERVED USERS’ BEHAVIORS FOR EVALUATING THE EXPERIENCING PROCESS

The actions of the participants who performed the proposed system were observed during the public exhibition period by video recording as well as by pens and paper. The observations were directed to two aspects: “operation situation of human-machine interface” and “participant’s behavior.” The results are listed in Table 19 and summarized as follows.

- 1) The interaction process of “Fruit Picking Fun” can attract participants to come and experience; and the participants found it interesting and wanted to play different farm scenes again and again.
- 2) The experience of taking and downloading photos is attractive to the participants, who took the initiative to pick up their mobile phones and scan the QR code.
- 3) The participants’ attention was on the display screen, so posture prompts and process guidance shown on the screen need be strengthened.
- 4) The height and angle of the Kinect device should take into account the average height of most participants; and it is necessary to avoid placing objects too close to the side to make the experience smoother.
- 5) It is necessary to eliminate the interference of other onlookers to make the sensing of the Kinect device more stable.

Table 19: Observation results of the users' performances of the proposed system with game "Fruit Picking Fun."

Perspective	Question	Summary of observations
Operation situation of human-machine interface	Whether the participants can understand the interactive modes of this system	<ol style="list-style-type: none"> 1) Participants watched the operating instructions of the system interface and can understand what kind of posture needs to be used to operate the system. 2) When the participants are not yet familiar with the interactive method of this work, they need explanations or action demonstrations to understand how to play the system.
	Whether the participant can operate the system smoothly	<ol style="list-style-type: none"> 1) Under the condition that the illumination angle of the Kinect device is the same, taller participants can operate the system more smoothly. 2) When there are other people watching, it sometimes interferes with the sensing of the Kinect device, causing participants to be unable to operate smoothly; on the contrary, if there is no other audience around, they can experience smoothly. 3) The location of the cantaloupe object is relatively nearby, and the participants cannot smoothly experience the harvest. 4) When the participant grabs the fruit and puts it in the basket in the initial screen, due to the illumination angle of the Kinect device, it is less sensitive to detect the hand gestures of grip and opening. 5) If the participants wear masks, it will affect the accuracy of face detection and body skeleton detection.
Participant's behavior	Whether the participants are interested in the system	<ol style="list-style-type: none"> 1) The participants are interested in the exhibition layout and music of the proposed system and take the initiative to visit it. 2) The participants look forward to the context and interaction of the three kinds of fruits on the initial screen, and experience each of the three scenes.
	The feedback given by the participants in the operation of this system	<ol style="list-style-type: none"> 1) The participants were curious about how the Kinect detects their actions. 2) The participants were curious about where the photos will be uploaded. 3) The participants feel that it is commemorative to be able to download photos to their mobile phones.
	The participants' reaction when interacting with this system	<ol style="list-style-type: none"> 1) The participants felt very happy during the process of interacting with the fruit objects. 2) The participants were surprised when they saw decorations on their heads or harvested fruits on their hands. 3) The participants did not know that they could change their actions when taking photos and counting down. 4) The participants' attention was focused on the screen in front of them, and were less likely to notice the operation instructions for harvesting and returning to the initial screen.

	The participants' additional actions in operating this system	<ol style="list-style-type: none"> 1) The participants wanted to take photos with other friends who were visiting together. 2) Some participants did not want to use the photo-taking function or scan the QR code to download photos and just wanted to experience the interaction with fruits.
--	---	--

4.2. ANALYSIS OF COMMENTS COLLECTED FROM INTERVIEWS WITH THE USERS

During the public exhibition period, as mentioned previously 50 participants were randomly selected to accept interviews of three aspects, namely, “operation situation of human-machine interface,” “operation experience,” and “views on interactive experience of food and agricultural education.” Some researchers of this study recorded the responses and counted the number of people who expressed each opinion. If more than 75% of the participants have expressed an identical opinion, the opinion is regarded as a majority suggestion; and if there are only five participants or less, it is regarded as a minority one. The collected opinions are listed in Table 20 and are summarized in the following.

- 1) The operation process of the proposed system is simple and intuitive.
- 2) The way to play the game on the proposed system was fun and interesting.
- 3) The participants have positive feelings about the feedback of getting photos immediately after the experiencing process.
- 4) The use of body movements for interaction, coupled with rich digital content design, can help the participants integrate into the interactive experiencing situation.
- 5) The participants gave positive comments on the experiencing form of introducing somatosensory interaction in food and agricultural education: in addition to attracting the public to understand the content of food and agricultural education, it also has an edutainment effect.
- 6) It is necessary to strengthen the part of prompting and guiding, so that participants can know what posture to interact with.
- 7) The height difference of participants and the placement of virtual objects affect the fluency of the operation.
- 8) More interactive actions, dynamic images, sound effects, or knowledge content feedbacks may be added into the system so as to create more educational effects.

Table 20: Results of interviews with the participants about their usages of “Fruit Picking Fun.”

Perspective	Question	Summary of opinions (* the digits in the parentheses specify the number of users making the comment)
Operation situation of human-machine interface	What do you think about the use of body movements to interact?	<ol style="list-style-type: none"> 1) Using body movements to interact is a novel and interesting way of experience. (42) 2) Using body movements to interact is a convenient and easy-to-understand way of experience. (28) 3) Using body movements to interact can be well integrated into the interactive experience situation, and have a sense of being immersed in the environment. (7)
	What is your opinion on the operation interface of this system?	<ol style="list-style-type: none"> 1) The overall operation interface of this system is simple and easy to operate. (34) 2) The gesture design of this system is intuitive and in line with the real situation. (8) 3) The objects in the screen are placed at the two sides so that if the gestures are not standard enough, the operation will become unsmooth. (15)

		4) This system needs to be strengthened in the prompt and guiding part in order to tell the user more clearly what to do next. (14)
Operation experience	What is your opinion on the layout of the exhibition or the design of the digital content of this system?	1) The digital content design of this work is rich and beautiful. (32) 2) The participants liked the AR visual effect of a farmer's hat-like object on their heads. (7) 3) More sounds or animation feedback related to agricultural knowledge may be added. (18)
	What do you think of the educational nature of this system on agricultural knowledge?	1) The participants can learn knowledge related to fruits by this system. (6) 2) This system can add more knowledge and enhance its educational effect. (16)
	What are your thoughts and feelings while playing this work?	1) The way to play this work is fun and interesting. (37) 2) Most participants liked to interact with the fruits in the scene. (39) 3) The participants liked the feedback of getting photos after the experiencing process. (13) 4) The participants can integrate into the interactive experience situation through posture and movement, and create a connection with the reality of fruit harvesting experience. (21) 5) The interaction of the cantaloupe scene was impressive. (16) 6) The feedback received after interacting with the fruit was insufficient, and the experience was a little insufficient because it ended too soon. (14) 7) Some mission levels or challenges can be added to increase the sense of accomplishment. (4) 8) The way of playing this system was boring and tedious. (9)
Views on interactive experience of food and agricultural education	What is your view on applying the experience form of somatosensory interaction to food and agricultural education?	1) It is appropriate to integrate the experience form of somatosensory interaction into food and agricultural education. (50) 2) Introducing the experience form of somatosensory interaction has the effect of edutainment. (26) 3) The somatosensory interactive experience can make people interested in actively learning about the content of food and agricultural education. (4)

4.3. ANALYSIS OF COMMENTS COLLECTED FROM SECOND EXPERT INTERVIEWS

In the interviews with the four invited experts (named P4 through P7 as seen in Table 10), questions of three aspects, namely, "operation situation of human-machine interface," "experiencing the *educational* content of the system," and "view on the interactive experiencing process," were asked. The collected comments are listed in Table 21 from which the following conclusions can be drawn.

- 1) The play experience of “Fruit Picking Fun” is both entertaining and educational, and receives positive reviews from the experts.
- 2) The somatosensory posture used by “Fruit Picking Fun” is in line with the context of fruit harvesting and the range of human capabilities, and the operation is simple and intuitive.
- 3) The series of the photo-taking experiencing process of “Fruit Picking Fun” is smooth.
- 4) When the participants encounter difficulties in operation, they must be given more prompts on the system interface.
- 5) Introducing somatosensory interaction in food and agricultural education is more attractive than general education using books.
- 6) Sharing the photos after the experiencing process through social media helps promote food and agricultural education.
- 7) It is suggested to further define the target user group and present more suitable content of food and agricultural education for the group.
- 8) A series of contexts of fruit harvesting can be explored further to enhance the depth of food and agricultural education.

Table 21: Results of interviews with the experts about the proposed system.

Perspective	Question	Summary of comments
Man-machine interface operation	What is your opinion on the interface of the proposed system?	<ul style="list-style-type: none"> • P4 and P7: the design of the somatosensory actions used in the system is intuitive and easy, in line with the scope of human capabilities. • P4: when participants encounter difficulties in operation, they must be given more prompts on the interface and more guidance to perform the system. • P6: somatosensory interaction is related to the user's body proportion, height, and behavioral ability; and the target group needs to be clearly defined.
	Do you think the whole experience process of this system is smooth?	<ul style="list-style-type: none"> • P4 and P7: the entire interactive experiencing process is very smooth. • P5 and P6: the Kinect is slightly delayed in gesture recognition due to hardware limitations, but the delay also depends on the operating context of the system; and sometimes such delays can be allowed.
Experience content of the system	What is your opinion on the exhibition layout and digital content design of the system?	<ul style="list-style-type: none"> • P5: it is good to create the feeling of having a good harvest by tracking the position of the hand and letting the fruit be held in the hands. • P6: it is necessary to allow the participants to experience the difference between traditional agriculture and digital experience to a certain extent; and it needs to think about how to deepen the impression of participants in the feedback design during the experiencing process. • P7: the same fruit can have images of different angles and sizes, so that the results of the interaction can have some changes.
	Do you think the system has an entertaining effect?	<ul style="list-style-type: none"> • P4, P5 and P7: the concept and game-play of the overall system are fun and interesting.

		<ul style="list-style-type: none"> • P5: because the participant’s hands must be used for picking fruits, his/her limbs will have relatively large movements, which can drive interest.
	What do you think of the educational nature of the system on agricultural knowledge?	<ul style="list-style-type: none"> • P4 and P7: more relevant knowledge content can be expanded for the system. • P6: it needs to conduct a more detailed investigation of the series of contexts behind the harvested fruits, in order to understand the needs of the target user group for food and agricultural education and to present the experiencing process in more relevant details.
View on interactive experience of food and agricultural education	What is your opinion on the application of “the experiencing form via the somatosensory interactive wall” of the system to food and agricultural education?	<ul style="list-style-type: none"> • P4 and P7: compared with traditional “flat” written education, through somatosensory interaction, food and agricultural education can be more attractive. • P5: the use of somatosensory interaction can create a situational atmosphere for food and agricultural education, and sharing photos through social media may attract people to experience the field. Such a virtuous cycle can help promote food and agricultural education.

5. CONCLUSIONS

Based on the design principles drawn from an extensive literature reviews of related human-machine interaction theories and existing cases of interactive devices, an interactive system with the Kinect as the core device using the “human” as the interface has been designed, on which a game named “Fruit Picking Fun” can be played for the aim of food and agricultural education. The interaction capability of the game is realized by the uses of computer vision and augmented reality (AR) techniques using the Kinect device as the sensor as well as a series of programs written in this study. The human-machine interaction is realized by these programs that implement the somatosensory functions of face detection and tracking, hand-gesture recognition, as well as body-model matching and tracking. The education content is taught in the way of playing the game to understand the harvesting processes of three typical types of fruit. The AR photo of the user with the harvested fruit held in hands may be taken by the system as a souvenir and downloaded to the user’s mobile phone.

During the public display of the system, the observation and interview methods were used to collect opinions from the participants and several invited experts. The effectiveness of the proposed system was evaluated according to these comments to reach the following positive conclusions.

- 1) *The interactive experience of this work is simple and intuitive* — the design of gestures is simple, intuitive, and in line with the real situation; it is easy to connect with the actual fruit harvesting experience; and the experience flow of the overall system is smooth.
- 2) *The use of body movements for interactive experiencing is given positive reviews* — using body movements to interact is a novel and fun way of experience, which can be well integrated into the context of fruit harvesting and create a sense of personal experience. The posture of holding the fruits after playing the game also has the concept of realistic harvest, allowing users to have the joy of harvest like the farmer.
- 3) *The introduction of somatosensory interactive food and agricultural education can arouse the interest of the participants and achieve the effect of edutainment* — the somatosensory interaction offered by the system is quite suitable for use in food and agricultural education, which is more attractive than ordinary book education, and can achieve the effect of edutainment.
- 4) *In addition to being commemorative, the experience of taking AR photos can achieve the effect of publicity and promotion of food and agricultural education through sharing on social media* — being able to scan the QR code

to get the AR photos after the experiencing process is commemorative, and sharing through social media is also a way to publicize and promote food farmer education.

The three types of fruit used in the game “Fruit Picking Fun” are just examples; other agricultural products may also be included in the future. Furthermore, the interactivity of the game “Fruit Picking Fun” may be increased, and more special animation effects and sound feedbacks related to the knowledge of food and agricultural education can be added. Finally, the target groups to use the system may be extended, and the knowledge content of food and agricultural education may be improved to be richer.

SOURCES OF FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The author have declared that no competing interests exist.

ACKNOWLEDGMENT

None.

REFERENCES

- [1] Food and Agriculture Organization of the United Nations (2014). Biodiversity for Food and Agriculture: Contributing to Food Security and Sustainability in a Changing World. <http://www.fao.org/family-farming/detail/en/c/284748/>. Accessed 23 June 2020.
- [2] Boccaletti, S. Environmentally responsible food choice. *OECD Journal: General Papers*. 2008;2:117-152.
- [3] Tscharntke T, Clough Y, Wanger TC, Jackson L, Motzke I, Perfecto I, Whitbread A. Global food security, biodiversity conservation and the future of agricultural intensification. *Biological Conservation*. 2012;151:53-59.
- [4] Konuma H. Status of world food security and its future outlook, and role of agricultural research and education. *Journal of Developments in Sustainable Agriculture*. 2016;10:69-75.
- [5] Kimura AH. Food education as food literacy: privatized and gendered food knowledge in contemporary Japan. *Agriculture Human Values*. 2011;28:465-482.
- [6] Morita T. Background and situation of dietary education □ In connection with the Basic Food Education Bill. *Survey and Information (調査と情報)*. 2004;457:1-10 (in Japanese).
- [7] Uenaka O. Significance and issues of local production for local consumption in food and agriculture education. *Educational Studies Review (教育学論究)*. 2013;7:47-53 (in Japanese).
- [8] Malik S, Agarwal A. Use of multimedia as a new educational technology tool □ A study. *International Journal of Information and Education Technology*. 2012;2:468-471.
- [9] Jeng T, Lee CH, Chen C, Ma, Y. Interaction and Social Issues in a Human □ Centered Reactive Environment. In: *Proceedings of 7th International Conference on Computer Aided Architectural Design Research in Asia (CAADRIA)*, Cyberjaya, Malaysia, Apr. 18-20, 2002; 258-292.
- [10] Crowley JL, Coutaz J. Vision for Man Machine Interaction. In: *Proceedings of IFIP International Conference on Engineering for Human-Computer Interaction (EHCI'95)*, Grand Targhee, Wyoming, USA, Aug. 1995; 28-45.
- [11] Turk M. Computer vision in the interface. *Communications of the ACM*. 2004;47:60-67.
- [12] Jaimes A, Sebe, N. Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding*. 2007;108:116-134.
- [13] Wilson AD. PlayAnywhere: A Compact Interactive Tabletop Projection-Vision System. In: *Proceedings of 18th Annual ACM Symposium on User Interface Software and Technology (UIST '05)*, Seattle, WA, USA, Oct. 23-27, 2005; 83-92.

- [14] Hsu HMJ. The potential of Kinect in education. *International Journal of Information and Education Technology*. 2011;1:365-370.
- [15] Geller T. Interactive tabletop exhibits in museums and galleries. *IEEE Computer Graphics Applications*. 2006;26:6-11.
- [16] Fujii Y. Sagen Ishizuka's dietary education and dietary method: A study on the intellectual framework of nutrition therapy 11. *Bulletin of Faculty of Human Life Studies, Fuji Women's University*. 2014;51:25-38 (in Japanese).
- [17] National Chengchi University Aboriginal Studies Center. Food farmers education in the United States. *Aboriginal Education World*. 2018;81:74-77 (in Chinese).
- [18] Petrini C. *Slow Food Nation: Why Our Food Should Be Good, Clean, and Fair*. New York: Rizzoli Publications; 2013.
- [19] Powell D, Agnew D, Trexler C. Agricultural literacy: Clarifying a vision for practical application. *Journal of Food and agricultural education*. 2008;49:85-98.
- [20] Asaoka N. *Practice of New Environmental Education*. Tokyo: Kobundo; 2005 (in Japanese).
- [21] NAAE. What is Food and agricultural education? <https://www.naae.org/whatisaged/>. Accessed 26 July 2020.
- [22] Kantowitz BH, Sorkin RD. *Human Factors: Understanding People-System Relationships*. John Wiley & Sons Inc.: Hoboken, NJ, USA; 1983.
- [23] Deng L, Wang G, Yu, S. Layout design of human-machine interaction interface of cabin based on cognitive ergonomics and GA-ACA. *Computational Intelligence and Neuroscience*. 2016;2016:1-12.
- [24] Ye J. *Introduction to Interactive Design*. Taipei: Artist; 2010 (in Chinese).
- [25] Jaimes A, Sebe, N. Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding*. 2007;108:116-134.
- [26] Ahamed MM, Bakar ZBA. Triangle model theory for enhance the usability by user centered design process in human computer interaction. *International Journal on Contemporary Computer Research*. 2017;1:26-32.
- [27] Szeliski R. *Computer Vision: Algorithms and Applications*. New York: Springer; 2010.
- [28] Hu W, Tan T, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. *IEEE Trans. on Systems, Man, Cybernetics, Part C (Applications and Reviews)*. 2004;34:334-352.
- [29] Ojha S, Sakhare S. Image Processing Techniques for Object Tracking in Video Surveillance-A Survey. In: *Proceedings of 2015 International Conference on Pervasive Computing (ICPC)*, Pune, India, Jan. 09-10, 2015; 1-6.
- [30] Ragland K, Tharcis P, Wang L. A survey on object detection, classification and tracking methods. *Engineering Research & Technology*. 2014;3:622-628.
- [31] Iraola AB. *Skeleton Based Visual Pattern Recognition: Applications to Tabletop Interaction*. PhD Dissertation, the University of the Basque Country, Leioa, BI, Spain, 2009.
- [32] Crowley JL, Coutaz J, Berard F. Things that see. *Communications of the ACM*. 2000;43:54-64.
- [33] Wang CM, Wu TD. A new investigation of technology art - A study of applying computer vision techniques to interactive context. *Journal of National Taiwan College of Arts*. 2005;76:113-130.
- [34] Hassenzahl M, Diefenbach S, Göritz A. Needs, affect, and interactive products - Facets of user experience. *Interacting with Computers*. 2010;22:353-362.
- [35] Pine BJ, Gilmore JH. Welcome to the experience economy. *Harvard Business Review*. 1998;76:97-105.
- [36] Mitchell A, Linn S, Yoshida H. A tale of technology and collaboration: Preparing for 21st-century museum visitors. *Journal of Museum Education*. 2019;44:242-252.
- [37] ReactJ (2009). ReactJ - ReactTable Trance Live Performance #2. <https://www.youtube.com/watch?v=Mgy1S8qymx0>. Accessed 5 July 2020.
- [38] TeamLab. (2013). A Table Where Little People Live. <https://www.teamlab.art/w/kobitotable>. Accessed 9 July 2020.
- [39] TeamLab. (2015). Worlds Unleashed and Then Connecting. <https://www.teamlab.art/w/worlds-unleashed-restaurant/>. Accessed 9 July 2020.
- [40] TeamLab. (2017). Connecting! Block Town. <https://www.teamlab.art/w/block-town/>. Accessed 9 July 2020.
- [41] Rumu Innovation (2018). Happy Farmer. <https://www.rumuinno.com/happy-farmer>. Accessed 9 July 2020.
- [42] Buchenau M, Suri JF. Experience Prototyping. In: *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, New York, NY, USA, 2000; 424-433.

- [43] Naumann JD, Jenkins AM. Prototyping: the new paradigm for systems development. *MIS Quarterly*. 1982;3:29-44.
- [44] Eliason AL. *System Development: Analysis, Design, and Implementation*. Northbrook, IL, USA: Scott Foresman & Co.; 1990.
- [45] Lidwell W, Holden K, Butler J. *Universal Principles of Design, Revised and Updated: 125 Ways to Enhance Usability, Influence Perception, Increase Appeal, Make Better Design Decisions, and Teach Through Design*. Beverly, MA, USA: Rockport; 2010.
- [46] Ye Z, Ye L. *Research Methods and Essay Writing*. Taipei: Shangding Culture; 1999 (in Chinese).
- [47] Yoo H, Kim H. A study on the media arts using interactive projection mapping. *Contemporary Engineering Sciences*. 2014;7:1181-1187.
- [48] Ministry of Agriculture, Forestry and Fisheries, Japan (2011). *Food Education Basic Law*. <http://www.maff.go.jp/j/syokuiku/kannrennhou.html>. Accessed 16 Apr. 2019.
- [49] Watanabe M, Nakamura O, Miyazaki A, Akinaga, Y. Current status of food education in school education. *Nagasaki University Comprehensive Environmental Research*. 2006;8:53-60 (in Japanese).
- [50] Sharma R, Pavlovic VI, Huang TS. Toward multimodal human-computer interface. *Proceedings of the IEEE*. 1998;86:853-869.
- [51] Rogers Y, Sharp H, Preece J. *Interaction Design: Beyond Human-Computer Interaction*. New York: Wiley; 2002.
- [52] Gibbon D, Mertins I, Moore RK. *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*. Berlin: Springer Science & Business Media; 2012.
- [53] Turk M. Computer vision in the interface. *Communications of the ACM*. 2004;47:60-67.
- [54] Murthy G, Jadon R. Computer Vision Based Human Computer Interaction. *Journal of Artificial Intelligence*. 2011;4:245-246.
- [55] Bobick AF, Intille SS, Davis JW, Baird F, Pinhanes CS, Campbell LW, Wilson A. The KidsRoom: A perceptually-based interactive and immersive story environment. *Presence*. 1999;8:369-393.
- [56] Levin G. Computer vision for artists and designers: Pedagogic tools and techniques for novice programmers. *AI & Society*. 2006;20:462-482.
- [57] Crowley JL, Coutaz J, Berard F. Things that see. *Communications of the ACM*. 2000;43:54-64.
- [58] Turk M, Kolsch M. (2004). Perceptual interfaces. In: Medioni G, Kang SB eds. *Emerging Topics in Computer Vision*. Englewood Cliffs, NJ, USA: Prentice Hall; 2004.
- [59] Su YR. Explore, experience, and interaction: The learning field of the museum for learning and playing together - Take the "New Farm Organic Fun and Fun Special Exhibition" in the South Gate Park of the Taiwan Expo as an example. *Taiwan Museum Quarterly*. 2016;35:42-49 (in Chinese).
- [60] Moment Factory (2017). *TABEGAMI SAMA*. <https://momentfactory.com/work/all/all/tabegami-sama>. Accessed 3 July 2019.
- [61] xXtralab (2016). A Nong's fantastic adventure- The New Farming and Organic LOHAS Exhibition. http://www.xxtralab.tw/tw/projects_post.php?id=39&nowTag=FEATURED#35. Accessed 5 July 2020.
- [62] Cinimod Studio (2017). *FIRE & ICE*. <https://www.cinimodstudio.com/fire-and-ice>. Accessed 6 July 2019.
- [63] Hush Studio (2012). *University of Dayton Interactive Wall*. <https://vimeo.com/28178841>. Accessed 9 July 2019.
- [64] Onformative (2012). *Nikefuel Station*. <https://onformative.com/work/nike-fuel-station>. Accessed 7 July 2019.
- [65] Moeslund TB, Hilton A, Kruger V. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*. 2006;104:90-126.
- [66] Shotton J, Sharp T, Kipman A, Fitzgibbon A, Finocchio M, Blake A, Moore R. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*. 2013;56:116-124.
- [67] Poppe R. Vision-based human motion analysis: An overview. *Computer Vision and Image Understanding*. 2007;108:4-18.