Effective Interface Design Using Face Detection for Augmented Reality Interaction of Smart Phone

Young Jae Lee

Dept. of Multimedia, Jeonju University
#45, Backma-Gil, Wansan-Gu,Jeonju, Jeonbul, 560-759, Korea
leeyj@jj.ac.kr

Abstract

This study proposes effective interface by using face-detection information, which comes from input image of the camera installed on a smart phone, and also by using movement information from reliable characteristics of the object. To do this, augmented reality interface has been constructed by putting the virtual object onto the camera image of the real space. The gamer and the virtual object, which are put in two separate spaces, engage in various real-time interactions such as attack, defense and alert. Difference image per frame of the tracked object information was employed to check if the hand has touched the interested area. Its effectiveness has been proven through experiment. Infusion technology, which combines TTS (text to speech) and vibrating sensor installed on a smart phone, was applied to provide emotional information to help the game participant to get immersed in the game. The proposed method could be applied as the basic material for AR interface.

Keywords: Augmented Reality, interface, interaction, face detection, smart phone

1. Introduction

Advancement of the augmented technology in recent years has made it available for various areas including game, education, distant medical diagnosing, broadcasting, architectural design, etc. Especially, there is heightened interest in services related to mobile augmented reality and the technology required, with the changing mobile computing environment [1, 9]. Mobile augmented reality uses various contexts to register virtual content onto real space and enables interaction, whose applications are extensive. Gartner identifies augmented reality as one of the strategic technologies for the next generation.[2] Google Trends also shows steady increase in search rates of information regarding augmented reality worldwide [3]. Mobile augmented reality service could vary depending on constructing methods; sensor-based pseudo augmented reality service, marker-based augmented reality service, and vision based augmented reality service, for whose realization many key technologies are being developed including object detection, recognition, tracking, rendering, interaction, etc [4].

With widespread use of mobile smart phones, augmented reality could augment appropriate virtual content to the environment and the object the user deals with, heightening usability. The earliest version of augmented reality service involved heavy equipment which hampered easy use for ordinary users [5, 6]. However, with rapid growth of multi sensors-mounted smart phones, users are able to use augmented reality service anywhere, anytime in their work and daily lives. Therefore, without separate markers, and with only a camera mounted on a smart phone, this study infuses TTS function and vibrating sensors and proposes AR environment construction and AR interface in a smart phone. For this purpose,

the smart phone game was developed based on face and hand recognition interface; background was constructed with smart phone camera images; real-time face recognition was conducted by using Android API; and AR environment was constructed by making 3D enemy character appear. The enemy character attacks the face continually, and TTS and mask are used to identify the situation and realize defense. Input images are both still image and moving. The vibrating sensor makes explosion on impact more real.

2. Mobile Augmented Reality Service and Contents

Mobile augmented reality services available so far could be classified, based on technologies of object tracking, into pseudo augmented reality, marker based augmented reality service, and vision based augmented reality service [6, 9].

First, pseudo augmented reality service uses sensors applicable on mobile gadgets to identify augmented location on screen and visualize related content. Information is augmented with GPS and digital compass installed on mobile gadgets. Easy realization is possible even without the need for complicated technology, but the drawback is that we cannot expect precise information augmentation on a particular location.

Second, marker based augmented reality service utilizes visual markers attached to the object to identify and locate the object. Related information is augmented to provide augmented reality service. The advantage is that you can easily identify the object only with the image from the camera without need for a compass or GPS sensor. But the markers attached to the object are sensitive to light changes and irritate one's eyes.

Third, vision-based augmented reality service captures main characteristics from the camera images and identifies and tracks the object to augment related content. The setback is that it requires abundant resources because of accompanying real time image processing and content augmentation. But the advantage is that it is highly mobile. This study is based on the third method, the vision based augmented reality service, which is not bound in space and highly mobile.

3. AR Game Design

Input images of camera installed on a smart phone are used to make the background for the game. If a person exists on the input image, the face is detected. The virtual character waits to attack until the face is detected. When face detection occurs, TTS delivers the warning message to the game participant, which alerts the game participant to move fast to avoid the attack. If the game participant fails to avoid the attack, the face explodes as a result. The gamer has the option to attack the enemy character with a missile while it is getting ready to attack the game participant. Or the gamer can defend against the attack by using the mask. That is, interface of attack and defense on the virtual space can be realized by using face detection, mask, and missile.

A touch button installed on the virtual space enables the game participant to use the mask freely. Touching the button may mean lessened intuitiveness in maneuvering the game, but if only the camera is used without any markers, it could activate the physical activities of the game participant and also help with realistic interaction. To deal with variant inputs, both motionless and moving images are to be input. To deliver the explosion on impact emotionally to the gamer, the vibrating sensors installed realize the vibrating mode for 150ms. The game was designed in CAMERA_FACING_BACK which enables face detection both in motionless and moving images. Face detection API on Android OS is employed to detect face. The enemy character gets out of the face-detected area at the moment of explosion caused by

the face attack and enters into protective mode and then gets ready for the next attack by continuously checking out the face detection information. Samsung Galaxy Tab 10.1 is used for the test. Coding is done on Eclipse by using Android SDL 3.1 version.
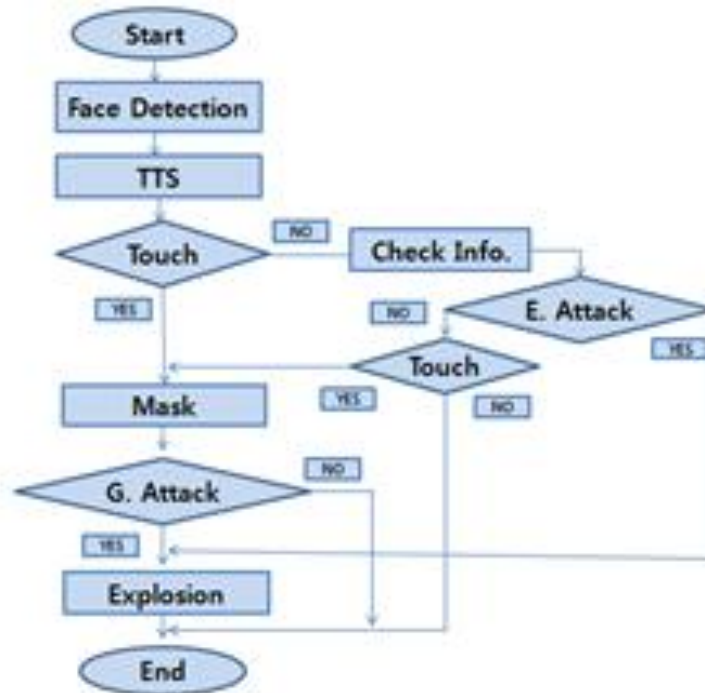
4. AR Game Diagram

Figure 1. AR Game Diagram

Figure 1 is the system diagram to test the AR game interface, which was applied in Experiment 5 (partially applied in Experiments 1-4). Image is input and face is detected. Then the warning message 'Face Detection. Be Careful' appears through TTS. The message is to alert the game participant about the attack by the enemy character. If the gamer touches the button on AR environment at this moment, the mask is worn on the face and protected from attack. If the gamer pushes the missile for attack (G. Attack), a missile is blasted at the enemy character and explodes the enemy. If the gamer fails to touch, the enemy character identifies the face detection coordinate and decides on the attack (E. Attack), resulting in face explosion. If the enemy does not attack, the gamer can once again decide to touch and move to the masked routine and continue with the diagram. But if it is not so, the gamer can move on to the end routine to finish the routine and restart.

5. Experiment

5.1 Experiment 1

Experiment 1 shows the motionless input image (Figure 2 (a)) and animated enemy character. Android API was used to detect the face (Figure 2 (a)). When the face detection is

successful, the enemy character should be located at the bottom. The enemy character uses the face detection information to attack the face at the optimal time. The attack causes an explosion on the face (Figure 2 (b)). Figure 2 (c) shows the image where the gamer destroys the enemy character with a missile before getting attacked upon face detection (Figure 2 (d)). Experiment 1 shows the enemy character, the virtual character on AR environment, being attacked and exploding, along with the gamer's attack for defense and the resulting explosion. From the resulting image in Experiment 1, we could verify that interface function for attack and defense using face detection information on AR environment is excellent and the size and location, and speed of animation or movement levels were properly designed.

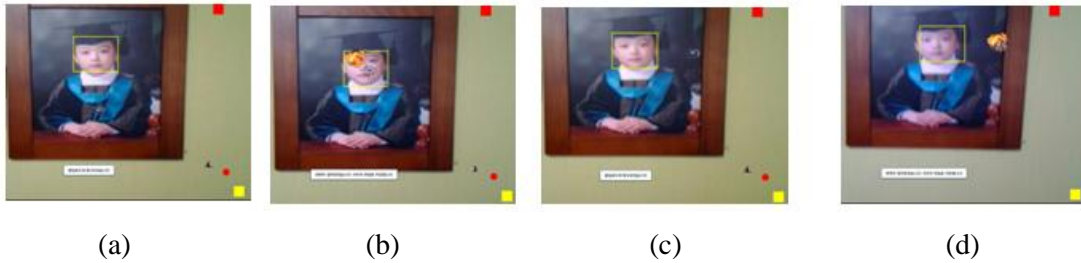| (a) | (b) | (c) | (d) |

Figure 2. Face detection, attack from enemy, explosion of the game participant, attack from the game participant for defense and the resulting explosion of the enemy

5.2 Experiment 2

In Experiment 2, the input image is still image (Figure 3 (a)). When face is detected, if you push the yellow square button on the bottom right before the enemy attack, the mask image appears as in Figure 3 (b) to defend the face. In this case, the enemy character attacks the face but the defense line built up by the mask does not allow any damage. The enemy character rapidly moves to another location for the next attack (Figure 3 (c)). At this moment, if you push the red round button to defend against the enemy attack, the missile next to it is launched to destroy the enemy character (Figure 3 (d)).

In Experiment 2, the interface is designed in a way that the gamer can be proactive in dealing with the situation. When the still image picture-like image is input, the gamer identifies the enemy attack in advance, enters into the alert mode and launches the attack. Experiment 2 verified the functions of the entertaining AR interface which involves defense using mask, attack at the optimal timing, etc.

| (a) | (b) | (c) | (d) |

Figure 3. Face recognition, mask defense, attack from the gamer and the image of the enemy explosion

5.3 Experiment 3

Differing from Experiment 2, the input image is of a moving person. Face detection takes place (Figure 4 (a)) and the enemy character is located at the bottom. The enemy character uses the face detection information and launches the prompt attack at the optimal timing. The attack results in explosion in the face (Figure 4 (b)). Figure 4 (c) shows the image where the enemy character explodes upon being attacked with the missile by the gamer (Figure 4 (d)). The gamer attacks before the enemy starts to attack upon face detection. From the resulting image in Experiment 3, we could know that the attack and defense interface works properly when a moving image is input on AR environment.
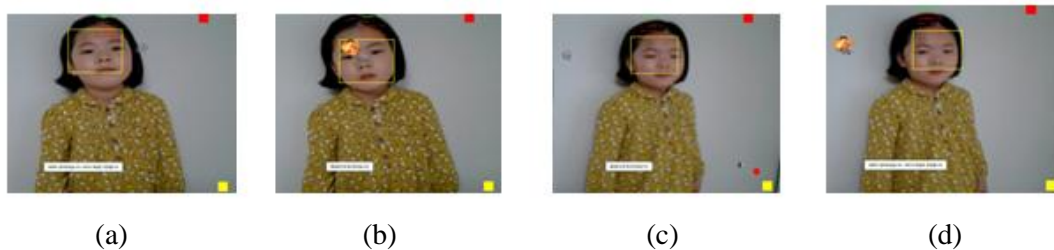

(a)   (b)   (c)   (d)

Figure 4. In the case of moving image input, face detection, enemy attack, mask for defense, gamer's attack and enemy explosion

5.4 Experiment 4

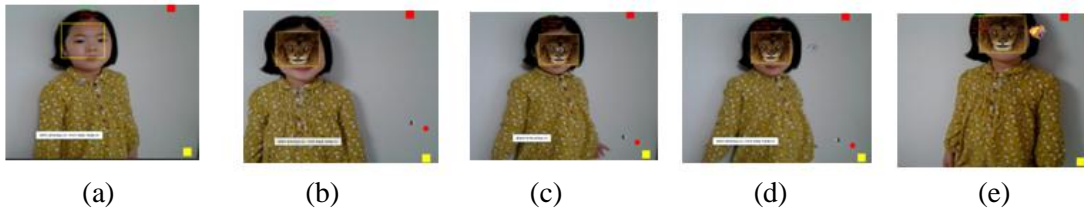(a)   (b)   (c)   (d)   (e)

Figure 5. In moving image input, face detection, mask defense, attack from the gamer for defense, enemy explosion

In Experiment 4, as in Experiment 3, moving image is input((Figure(a)). Face detection occurs, which is recognized through TTS. The yellow square button on the bottom right is pushed before the enemy attack and the mask image defends the face as in Figure(b). The enemy character attacks but the mask acts as defense with no damage occurring(Figure(c)). Then the enemy character moves fast to another location for the next attack(Figure(d)). To defend against the enemy attack, the red round button is pushed and the missile next to it is blasted to the enemy character, destroying it(Figure(e)).

Experiment 4 is to test the interface enabling proactive gamer response in which the gamer identifies the enemy attack in advance, enters into alert, and conducts a preemptive attack. AR interface was proven for its entertaining function through the use of the mask, attack at the optimal timing, etc.

5.4 Experiment 5
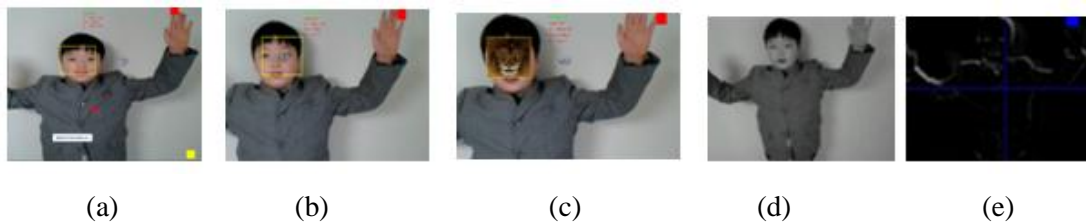

(a)   (b)   (c)   (d)   (e)

Figure 6. Face detection, enemy attack, explosion of the gamer, button touch for defense, mask, attack from the gamer and enemy explosion

Experiment 5 shows moving image input(Figure(a)). After face detection, TTS information is output and the gamer, unlike in the previous experiment, pushes the red square button on the upper left. Then the change in hand location of the gamer is identified to check if the button is touched or not. If the gamer touches the button, mask is output on the face area to defend the face.

In Experiment 5, where the moving image is input, the interface allows face-protecting mask triggered by the touch of the button in the AR environment when the face is detected. To realize the interface, the color of the input image is changed to grey(Figure (d)), and the difference image for each time frame is calculated. This is the method to check if the button was touched or not by checking the differences in movement information of the object depending on the reliable characteristic pixel changes of the location of the blue button on the virtual space. When we install the button to be touched by hand, intuitiveness of game maneuvering is lessened. But the interface was identified to be effective in activating the physical activities of the gamer and for realistic interaction when only the camera was used without markers. In addition, it was possible to build the interface which allows various jobs in AR environment.

6. Conclusion

Interface is an important technology to enhance the value of information gadgets. This study proposes effective interface for AR interactive contents like high value-added games in a smart phone. No markers are used and only the input images of camera installed on a smart phone are employed to get the movement information form reliable characteristics of the object. Face detection takes place either from still image picture-like image or from moving images. Experiments of various interfaces for attack and defense are presented by using an animated enemy character. The information of object tracking is identified for touch by using difference images for each frame whose effectiveness was verified through experiments. In addition, TTS(text to speech) and vibrating technology using vibrating sensors could provide situational information as well as emotional information.

References

[1] C. Shin, Y. Oh, Y. Seo, U. Uh, "Trends in Mobile Augmented Reality Service & Prospects for Sustainable Contents Ecology", The Korean Institute of Information Scientist and Engineers Society,06, pp43-49, (**2010**).

[2] Gartner, http://www.gartner.com/, accessed (**2010**) May 20.

[3] Google Trends, http://google.com/trends/, accessed (**2010**) May 20.

[4] F. Zhou, H. B. Duh and M. Billinghurst, uc0½rends in Augmented Reality Tracking, Interaction and

Display:A Review of Ten Years of ISMAR", IEEE International Symposium on Mixed and Augmented Reality, pp. 193-202, **(2008)**.

[5] T. Hölerer, S. Feiner, T. Terauchi, G. Rashid and D. Hallaway, "Exploring MARS: Developing Indoor and Outdoor User Interfaces to a Mobile Augmented Reality System", Computers and Graphics, 23(6), Elsevier Publishers, pp. 779-785, **(1999)** December.

[6] S. Dieter, S. Gerhard, W. Daniel, B. Istvá, G.Reitmayr, N. Joseph and F. Ledermann, "Managing Complex Augmented Reality Models", IEEE Computer Graphics and Applications, Vol. 272, Number 1716, pp. 32-41, **(2007)** July/August.

[7] Y. Jang, U. Uh, D. Kim and C. Shin, "Technology Trends in Mobile Augmented Reality", Featured Article on Mobile Internet, Open Standards and Internet Association, Vol. 38, Number 1, pp. 41-52, **(2010)**.

[8] Y. J. Lee, G. H. Lee, "Augmented Reality Game Interface Using Effective Face Detection Algorithm", International Journal of Smart Home, Vol. 5, No. 4, pp. 77-88, **(2011)** October.

[9] D. R. Kisku, P. Gupta, J. K. Sing, "Face Recognition using SIFT Descriptor under Multiple Paradigms of Graph Similarity Constraints", International Journal of Multimedia and Ubiquitous Engineering, Vol.5, No.4, pp. 1-18, **(2010)** October.

[10] T. Yamabe and T. Nakajima, "Possibilities and Limitations of Context Extraction in Mobile Devices: Experiments with a Multi-sensory Personal Device", International Journal of Multimedia and Ubiquitous Engineering, Vol.4, No.4, pp. 37-52, **(2009)** October.

Authors

Young Jae Lee received the B.S. degree in Electronic Engineering from Chungnam National University in 1984,the M.S. degree in Electronic Engineering from Yonsei University in 1994, and the Ph.D. in Electronic Engineering from Kyung Hee University in 2000. He ispresently an Associate Professor in the College ofCulture and Creative Industry at Jeonju University. His research interests include Smart Media, Computer Vision, AR(Augmented reality), and Computer Games.