UP2U: Program for Raising Awareness of Phubbing Problem with Stimulating Social Interaction in Public using Augmented Reality and Computer Vision

Kawin Metsiritrakul[1], Nuttapol Puntavachirapan[2], Thananop Kobchaisawat[3], Sangsan Leelhapantu[4], Thanarat H. Chalidabhongse[5]
Department of Computer Engineering
Faculty of Computer Engineering, Chulalongkorn University
Bangkok, Thailand
kawin919@gmail.com[1], nutzaichu@gmail.com[2], thananop.k@student.chula.ac.th[3], sangsan_momo@hotmail.com[4]
thanarat.c@chula.ac.th[5]

***Abstract*— 'UP2U' project has an objective to create awareness of phubbing behavior in today society. 'UP2U' is an interactive installation application which is designed to be set up in an indoor waiting area. The system has an objective to raises people awareness of their phubbing behavior and encourage them to have more interpersonal communication with each other. It was implemented using computer vision and augmented reality techniques. By conducting the experiments in both the lab and the real indoor public waiting area, we found the system demonstrates high accuracy of head tracking, face direction estimation, and gender classification. Moreover, the system has promising social impact to the phubbing society.**

Keywords—phubbing; Kinect; computer vision; augmented reality;

I. Introduction

In the age that social network plays a key role in people's daily life, the pictures of people looking at their phones, while sitting in a restaurant, taking public transportation, or walking in the street, become common things. This kind of society is called a 'Phubbing Society'.

Phubbing can be described as an individual interacting with his or her mobile phone while ignoring surrounding people, i.e., dealing with the phone and escaping from interpersonal communication. There are several studies on the phubbing behavior and virtual addictions. Karadag, et al.[1] investigated the determinants of phubbing behavior based on gender, smartphone ownership, and social media membership and found that phubbing behavior is determined by four major addictions, i.e., games, The Internet, social media, and phones. These addiction behaviors lead to negative consequences, reported in various researches. The study by Roberts, J. A., et al.[2] stated that people who attend to their cell phones in the presence of their partners are less satisfied in their relationships and report more symptoms of depression. In other words, the phubbing behavior tends to ruin relationships.

However, there are attempts to solve this problem. One of them is 'Stop Phubbing' campaign [3] which visualize phubbing statistic throughout the world and provide anti-phubbing services, such as anti-phubbing posters to be downloaded. Another attempt is from IKEA [4], who came up with a smartphone-powered-steamboat that people need to place their devices into the base of the machine to power the steamboat. The result was that everyone had made more interpersonal communication than looking at their own devices.

We would like to be another attempt to relieve the phubbing issue. Thus, we have developed an application named 'UP2U' to reflect the phubbing behavior of people in the society. The objective of the application is to raise people awareness on their phubbing behavior, and to encourage them to have more interpersonal communication with others. The system is a kind of interactive installations implemented using computer vision and augmented reality techniques.

The name 'UP2U' has 2 connotations. First, it means keeping your face up and look at others instead of your device. Second meaning is it is your choice to choose between your friends or your phone.

The rest of this paper is organized as follows: In Section II the related works on phubbing issues are presented. The proposed system overall process is illustrated in Section III. In Section IV, the detail of how the system works behind the scene is described. Then, followed by Section V which both technical and social experimental results are reported and analyzed. Finally, we conclude the work and suggests further improvements in Section VI.

II. Related Works

There is a mobile application called Forest [5], developed by ShaoKan Pi, that also takes a part in the anti-phubbing campaign. Forest is an application that you can plant a seed in your smartphone which will gradually grow into a tree unless you leave this application to use your smartphone. This application helps people to put down their phone and focus on what is more important in their life.

Another anti-phubbing mobile application is RestEye [6], developed by Forwen, this is a simple application that reminds you when you have put too much attention to your phone. The application also uses background location services to allow you to monitor the location of your phone usage.

'UP2U' has some differences from the applications above. While Forest and RestEye are aimed for users' concentration on

978-1-5090-2033-1/16/$31.00 ©2016 IEEE

978-1-5090-2033-1/16/$31.00 ©2016 IEEE

their tasks, 'UP2U' is focusing on users' interpersonal communication, by using computer vision and augmented reality techniques. In addition, people need to know Forest and RestEye and must download it into their devices. In contrast, 'UP2U' doesn't require this kind of awareness of the existing of the system. The users can gain benefits of the system without knowing it before.

III. UP2U IN ACTION

The goal of this work is to develop an application which is able to encourage people to have more social interaction rather than attend to their phones.

'UP2U' is designed to be set up in an indoor waiting area, displaying by a large screen opposite to the seats as showed in Fig 1.

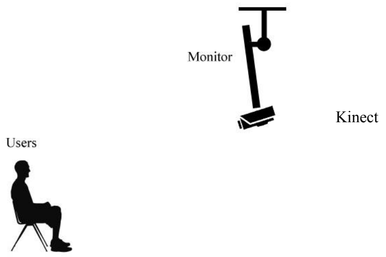


Fig. 1. Installation scenario of 'UP2U' application.

The application switches the display according to detected stage. There are 2 system stages: one is "empty scene" where the system detect no human in the scene, the other one is "people in scene" stage where one or more people are detected. When "empty scene", the system runs an anti-phubbing advertising VDO on the monitor. On the other hand, when "people in scene", they system will generate an avatar for each person based on gender and face direction of that person. The face direction consists of 4 directions; looking up, looking down, facing left, and facing right, as shown in Fig. 2.

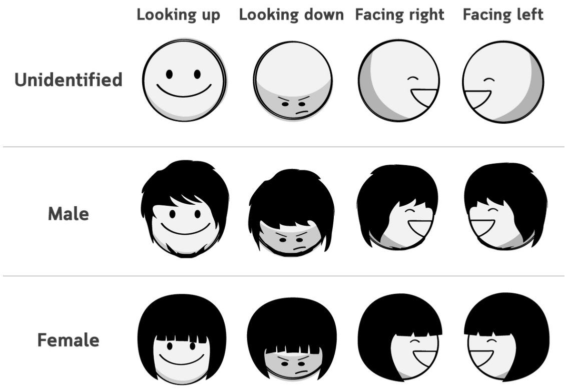


Fig. 2. All types of the avatar based on person's face direction and gender.

The avatar's color can also changes, depending on the face direction and interaction with others. They system works as follows: when the person looks up the avatar will turn yellow and when the person looks down the avatar will turn red. Meanwhile, if people turn their faces to each other, their face will become green.

In addition, the system also generates an interactive gauge that gradually increases when people turn their faces to each other and conversely it gradually decreases when they don't. If the gauge is maximized, the complimentary video will be displayed and the gauge will be reset. On the other hand, if any person looks down, the whole screen will gradually become red. When the red color is maximized, the warning video, illustrating phubbing negative consequences, will be displayed and the screen will be reset. All type of scenarios is showed in Fig. 3.

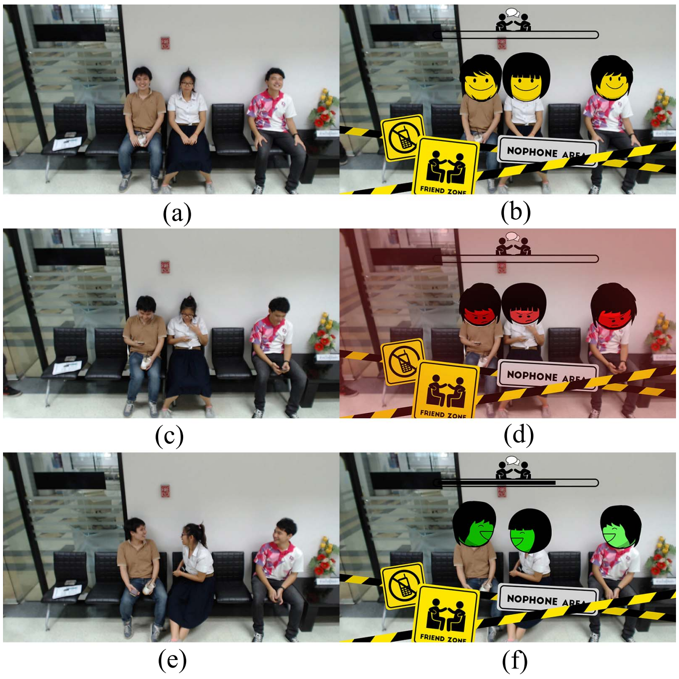


Fig. 3. Example of a various types of scenarios based on the number of people, face directions, and genders. (a)-(b) Fetched and augmented images for looking up scenario (c)-(d) Fetched and augmented images for looking down scenario (e)-(f) Fetched and augmented images for people looking at each other scenario.

IV. UP2U BEHIND THE SCENE

As mentioned above, techniques behind 'UP2U' are both image analysis or computer vision (CV) part, and a computer graphic which generates augmented reality (AR) part. The overview of the image analysis part is shown in Fig. 4. The process composes of three main modules: head tracking, face direction estimation, and gender classification. First, the head tracking module is developed by using the skeleton tracking data received from the Kinect. Moving average and occlusion handling technique are applied to maximize efficiency of the module. After the head position is tracked, then the data is passed to face direction estimation module. By using skin and hair color segmentation on HSV color channels along with morphological filter, the areas of user's face and hair are extracted and enhanced quality to be used for the face direction calculation. Last, the gender of each user is identified by gender classification module. This module is done by requesting a FaceAPI service, provided by Microsoft Azure, via HTTP POST method for face analysis. Afterwards, all the data is gathered and sent to the display side, which generates user avatars, renders related graphic elements, and plays anti-phubbing video in real time.

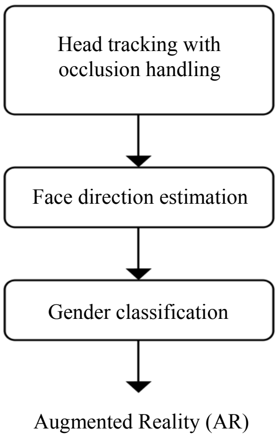

Head tracking with occlusion handling
Face direction estimation
Gender classification
Augmented Reality (AR)

Fig. 4. Overall image analysis process of 'UP2U'.

Head tracking: The head tracking module is implemented by using a robust 3D skeleton tracking middleware developed by PrimeSense [7] for OpenNI. The middleware's algorithms utilize the depth, color, and IR information received from the Kinect, which enable us to perform skeleton tracking and return positions of body parts: head, neck, shoulders, hands, etc., of each appeared subject in real time. In 'UP2U', we use only head position. However, the head position obtained from OpenNI's skeleton tracking is not stable, thus the moving average is applied to smoothen the head positions over time.

Occlusion handling: From our observation, we found that the head positions obtained from Kinect+OpenNI as mentioned above is quite sensitive. Sometimes, the Kinect is possibly unable to detect skeletons in some situations, such as when someone passes through the scene and occlude the tracked subjects. This problem is fixed by adding a delay for skeleton vanishing. Each skeleton has its own id number. When the skeleton disappears, the system holds the last skeleton position and still displays the avatar of that skeleton for a certain period of time. If the skeleton has disappeared exceed that period of time, the system will remove the avatar out of the scene.

Face direction estimation: After locating the head position of each subject in scene, our system will proceed with a process as shown in Fig.5. Starting by creating a square boundary, covering each subject's head region, as shown in Fig. 6. The size of the bounding box is calculated as a function of input image resolution, estimated human head size, and distance between subject and the camera, based on the perspective projection transformation.

$$B = f(r, s, d)$$

where B is size of bounding box
r is input image resolution
s is estimated human head size
d is distance between subject and the camera

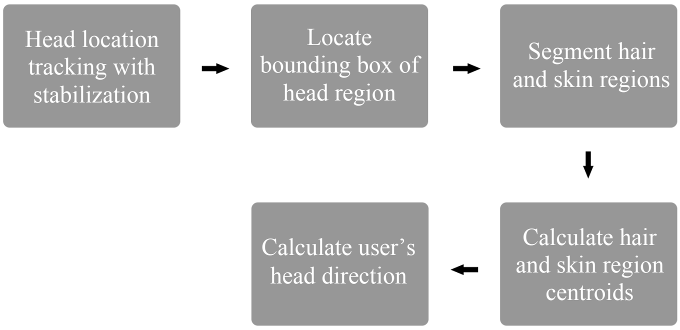

Head location tracking with stabilization
Locate bounding box of head region
Segment hair and skin regions
Calculate hair and skin region centroids
Calculate user's head direction

Fig. 5. Face direction estimation process

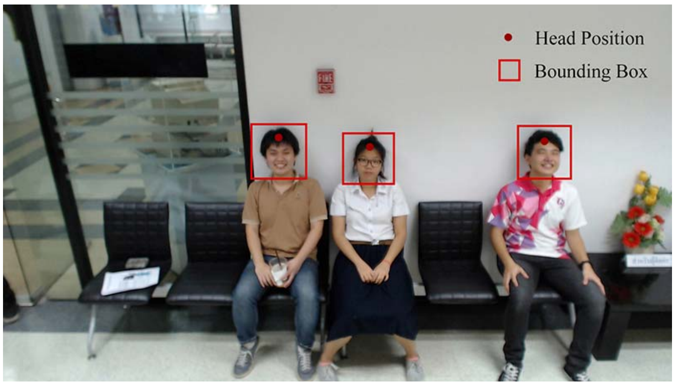


Fig. 6. Tracked subjects' head region localization.

After the bounding box of head regions are located, the system performs skin segmentation, using human skin tone color in each boundary to find the subjects' face areas. The skin tone color is detected when it is in the specific HSV color range as follows: from H 0 S 40 V 60 to H 39 S 150 V 255, according to the study of K. Sobottka and I. Pitas [8]. The hair segmentation is done by the same method with the color range from H 0 S 0 V 0 to H 255 S 255 V 80 for native Asian hair color, defined by Atlas of Human Hair [9]. Subsequently, the system uses a morphological filter to improve the quality of the segmentation. Firstly, the segment is morphed by dilation following by erosion to close all holes inside. Secondly, the segment is morphed by erosion following by dilation to remove the noises. The segmentation results for each step are illustrated in Fig. 7.

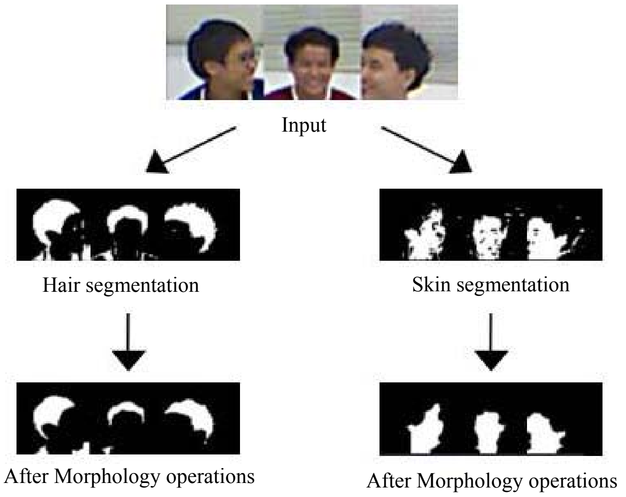


Fig. 7. The skin and hair segmentation process.

However, the system is possibly confused due to an error, affected by the clothes color or the background color which are in the same range of skin or hair color, as shown in Fig. 8. These unwanted areas are filtered by size and relative location of the blobs.

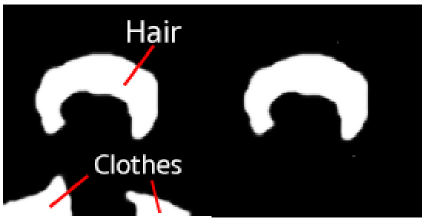


Fig. 8. Before and after remove unwanted areas.

After the system gets the area of face and hair, it calculates the direction of the user's head (see Fig. 9) using the following equation:

$$Angle = \tan^{-1}\left(\frac{x_f - x_h}{y_f - y_h}\right) \qquad (1)$$

where (x_f, y_f) is face's centroid, and
(x_h, y_h) is hair's centroid

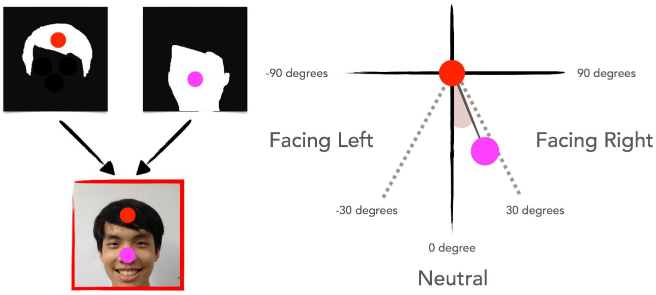
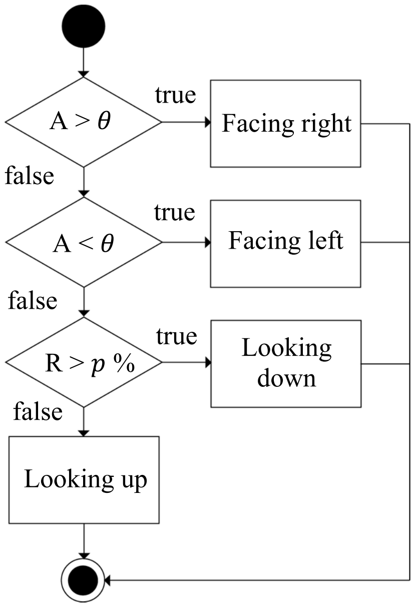


Fig. 9. Face direction is estimated based on relative angle between hair and face centroids.

If the angle is greater than θ degrees, the user's head is considered "facing right". On the other hand, if the angle is less than$-\theta$ degrees, the user's head is considered "facing left". But if the angle is between $-\theta$ to θ degrees, there are 2 possible cases. First, if the hair area is more the p percentage of the whole head area, the subject is considered "looking down". Otherwise, the subject is considered "looking up". The flowchart of this process is illustrated in Fig. 10. Note that the threshold values, θ and p, are varied depending on setup environments. Thus, they are configurable.



where A is face and hair relative angle
R is percentage of hair per whole head area

Fig. 10. Face direction estimation flowchart.

Gender Classification: For gender recognition, we employed the FaceAPI service provided by Microsoft Project Oxford [10], which is an API for detecting human faces in an image and returns face locations, ages, and genders. We have implemented a sub-system named 'WhatGender', to send a neutral upright face image of tracked subject to the FaceAPI, wait for the results, and return value to 'UP2U' system. 'UP2U' will calls 'WhatGender' only when it is able to extract a neutral upright face image of the detected subject. Once the gender is identified, the tracked subject will be labeled with the specified gender, 'MALE' or 'FEMALE', throughout the session. For unidentified gender subject, 'UP2U' will label it with 'UNDEFIED'. The 'UP2U' will repeatly call 'WhatGender' sub-system every time it can detect and extract a neutral upright face image of unidentified gender subject. Once the tracked subject's gender status changes from 'UNDEFINED' to 'MALE' or 'FEMALE', the 'UP2U' will stop the request process for that subject.

Display: For every frame, the computer vision (CV) part sends extracted information, which are the head position, face direction, and gender of each subject in the scene, to the augmented reality (AR) part. The AR will display the avatars over the users' head on the monitor, render relevant graphic, and play anti-phubbing video, related to the result we have got from the system in real time, currently we can run 60 fps on Mac Mini.

V. Experiments and Results

The system was set up at several venues such as in our lab, in a student lounge, in an exhibition hall, and at a public waiting area as shown in Fig. 11. This is to learn what factors that affect the system and how to change the configuration so that the system is robust to various environments.

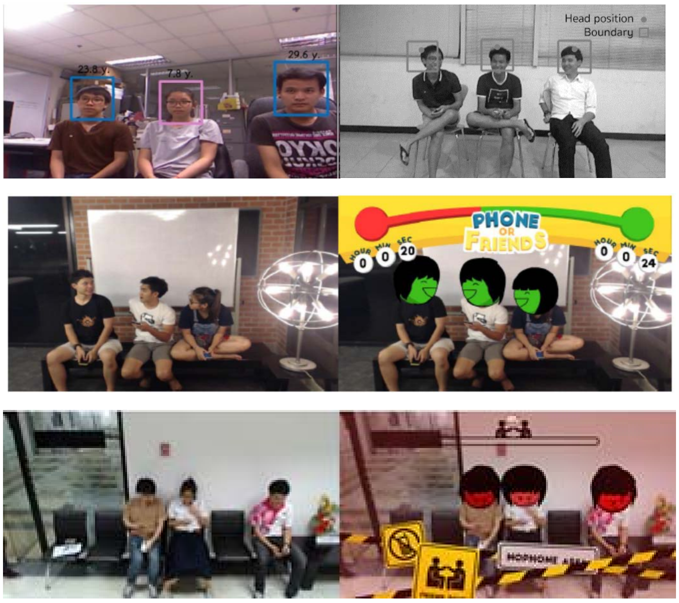


Fig. 11. The system was set up and tested in various indoor environments.

We evaluated the system in two aspects; one is the technical part and the other is the social impact evaluation. For the technical evaluation, we measured the accuracy of our 3 modules: head tracking with occlusion handling, face direction estimation, and gender classification. The Kinect was set up 3 meters high from the ground with approximated 45 degrees of depression angle and the distance between the bench and the Kinect was about 2 meters. The experiment consisted of 6 participants, 3 males and 3 females, between the ages of 21-23.

A. Head tracking evaluation

To evaluate the head position tracking, each participant was asked to walk in and out the scene for about 5 seconds to capture 300 frames. While the subject was moving, we compared the tracked head position with the actual head position in each frame. If the distance between the tracked and actual positions is less than 60 pixels, which is the about the width of the face in this particular environment, the frame will be counted as a True Positive (TP) with Correct Head Position (CHP). If the distance is more than 60 pixels apart, the frame will be counted as TP with Incorrect Head Position (IHP). If the system loses the tracked object, the frame is counted as either False Positive (FP) or False Negative (FN). The result is shown in Table I which shows that our proposed smoothened head tracking module with occlusion handling yields 87.67% accuracy, as oppose to the original OpenNI's Kinect skeleton tracking that yields 72.23%. The accuracy is computed by (CHP+TN)/Total Frames.

TABLE I. Head position Tracking evaluation

	Total Frames	**False (FN+FP)**	**TP**		**TN**	**Accuracy**
			IHP	**CHP**		
OpenNI's Kinect skeleton tracking	1800	462	36	**1302**	**0**	**72.23%**
Our head tracking w/ occlusion handling	1800	198	24	**1518**	**60**	**87.67%**

FN = False Negative, FP = False Positive, TP = True Positive, TN = True Negative, IHP = Incorrect head position, CHP = Correct head position. Accuracy = (CHP+TN)/Total Frames.

B. Face direction estimation evaluation

To evaluate the face direction estimation, each participant was asked to sit down on the bench and turning his or her face in a specific direction for about 5 seconds to capture 300 frames. Then, we compared the computed face direction with the actual ground truth. If the result was matched, then the frame is a correct frame. The accuracy was defined by the percentage of correct frames to total frames. The result in Table II showed that the average accuracy for the face direction detection module is 89.17% on average.

TABLE II. Face Direction estimation evaluation

Face direction	**Accuracy**
Looking up	98.00%
Looking down	98.33%
Facing left	80.67%
Facing right	79.67%
Average	**89.17%**

The face direction estimation experimental results show that the neutral faces with in both “looking up” and “looking down” have high accuracy, while the faces with “facing left” and “facing right” give lower accuracy. Because when the people are turning left or right, their hair may cover their faces and cause the hair and face centroid calculation is inaccuracy.

C. Gender classification evaluation

As previously mentioned, we did not do anything about gender classification except calling a service from MS Azure FaceAPI. But we wanted to evaluate some situations that might affect its performance. Thus, we conducted 2 experiments; one is sending neutral upright face images to the FaceAPI, and the other one is sending a number of random direction of face images. We captured 10 images for each participant, so we obtained 60 images for each experiment. After compared the gender classification result with the ground truth, we found that the gender classification performs so well when receiving nice neutral upright face images that achieves an accuracy of 100%. For the random face direction images, the classification accuracy drops to 73.33 %. The results are shown in Table III.

TABLE III. GENDER CLASSIFICATION EVALUATION

Azure FaceAPI Gender Classification	**Unknown**	**Known**		**Accuracy**
		False	**True**	
Neutral faces	0	0	60	100%
Random face direction	10	6	44	73.33%

From the results above, we found the accuracy of FaceAPI performs very well with neutral face images. Thus, in our system, we will analyze the captured face image and wait until the neutral face is captured before calling the “WhatGender” module. With this mechanism, our system can render the avatars that quite accurate gender to the tracked subjects.

D. Social impact evaluation

Our motivation behind the development of this system is to create awareness to the society of the existing “Phubbing” problem and hope to alleviate it. In order to validate our idea, we conducted the social experiments by installing the ‘UP2U’ system in a public waiting area. The Kinect was attached beneath an existing 40-inches digital signage LCD screen in front of an office. The opposite of the screen is a bench as shown in Fig. 12. The experiment was conducted for 4 hours from 1.00 P.M. to 5.00 P.M. of one working day. There were 22 persons involved; 14 males and 8 females, between the ages of 18-58 years old. We observed the behavior of people while they were interacting with the system, and asked them to answer a questionnaire as well as interviewed them for their opinion about phubbing society nowadays, and how ‘UP2U’ could raise awareness and impact to their phubbing behaviors.

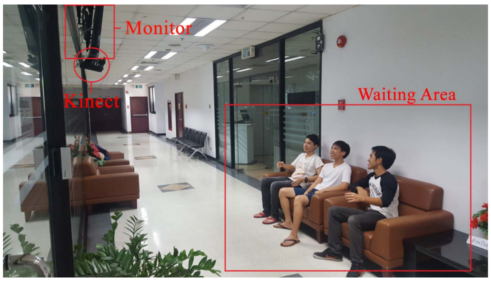


Fig. 12. The set up at public waiting area to observe the behavior of the participants and collect the social impact data.

The result shows the participants gained more awareness of their phubbing behavior at moderate level, on average 3.64 out of 5. The participants felt more enjoyable to interact with other persons rather than their phones at moderate level, on average 3.95 out of 5. And, the participants believed that ‘UP2U’ has potential to change the phubbing society at above derate level, on average 3.77 out of 5. Although the numbers are not extremely high, but from the responses, it has shown that ‘UP2U’ has some impact to them, at least reflect their phubbing behavior and build awareness.

VI. CONCLUSION

We have proposed the development of ‘UP2U’ system, which is a creative & interactive installations. The system provides a collaborative environment that encourages social interaction by employing computer vision (CV) and augmented reality (AR) techniques. The system composed of CV module, that performs human head tracking, face direction determination, and gender classification, and AR module that rendering avatars into the screen according to analyzing results of CV. The system was implemented based on Windows using C++ with OpenCV, OpenNI, and FaceAPI. It performs at about 60 fps rate on Mac mini.

We have evaluated the system in both technical part and social impact part, as well as installing the system in various venues to make sure it performs well and robust. The result is shown positive impact to people in creating awareness of phubbing behavior and encouraging them to have more interpersonal communication to other people rather than interacting with the phone and ignoring surrounding people.

ACKNOWLEDGMENT

This work was supported in part by Thailand Junior Science Talent Project (JSTP) and Thailand National Software Contest (NSC).

REFERENCES

[1] Karadag E., Tosuntas S. B., Erzen E., Duru P., Bostan N., Sahin B. M., Culha I., Babadag B. (2015). Determinants of phubbing, which is the sum of many virtual addictions: A structural equation model. Journal of Behavioral Addictions, 4(2), 60–74.

[2] Roberts, J. A., & David, M. E. (2016). My life has become a major distraction from my cell phone: Partner phubbing and relationship satisfaction among romantic partners.*Computers in Human Behavior, 54*, 134-141.

[3] "Stop Phubbing", *Stopphubbing.com*, 2016. [Online]. Available: http://stopphubbing.com/. [Accessed: 30- Mar- 2016].

[4] R. Lee, "IKEA Phubbing | SoyaCincau", *Soyacincau.com*, 2016. [Online]. Available: http://www.soyacincau.com/tag/ikea-phubbing/. [Accessed: 30- Mar- 2016].

[5] S. Pi, *Forest: Stay focused, be present*. ForestApp inc., 2014.

[6] *RestEye - Stop Phubbing*. 2014.

[7] OpenNI/OpenNI,” GitHub. [Online]. Available: https://github.com/OpenNI/OpenNI. [Accessed: 06-Mar-2016].

[8] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," Signal Processing: Image Communication, 1998.

[9] Zaher Hamid Al-Tairi, Rahmita Wirza Rahmat, M. Iqbal Saripan, and Puteri Suhaiza Sulaiman, "Skin Segmentation Using YUV and RGB Color Spaces," Journal of Information Processing Systems, vol. 10, no. 2, pp. 283~299, 2014. DOI: 10.3745/JIPS.02.0002.

[10] “Face API.” [Online]. Available: https://www.microsoft.com/cognitive-services/en-us/face-api. [Accessed: 30-Oct-2015]