Selfie Guidance System in Good Head Postures

Naihui Fang
Dept. of Computer Science
The University of Tokyo
fangnaihui88@hotmail.com

Haoran Xie
Dept. of Computer Science
The University of Tokyo
xiehr@acm.org

Takeo Igarashi
Dept. of Computer Science
The University of Tokyo
takeo@acm.org

ABSTRACT

Taking selfies has become a popular and pervasive activity on smart mobile devices nowadays. However, it is still difficult for the average user to take a good selfie, which is a time-consuming and tedious task on normal mobile devices, especially for those who are not good at selfies. In order to reduce the difficulty of taking good selfies, this work proposes an interactive selfie application developed with multiple user interfaces to improve user satisfaction when taking selfies. Our proposed system helps average users take selfies by providing visual and voice guidance interfaces on the proper head postures to achieve good selfies. Preprocessing through crowdsourcing-based learning is utilized to evaluate the score space of possible head postures from hundreds of virtual selfies. For the interactive application, we adopt a geometric approach to estimate the current head posture of users. Our user studies show that the proposed selfie user interface can help common users taking good selfies and improve user satisfaction.

Author Keywords

Selfie; head posture; crowdsourcing; user interface;

ACM Classification Keywords

D.2.3. Design Tools and Techniques: User Interface

INTRODUCTION

The selfie is a type of self-portrait that usually consists of a photograph taken by mobile camera at an arm's length. It is one of the most direct ways to show oneself to others as well as to record one's daily life for one's own benefit. Selfies became more and more popular with the development of social media. Now, it plays an important role in our daily life. Along with its popularity, how to take a good selfie has become an urgent issue to be solved. People want to look good on social media but many of them encounter trouble when taking selfies. They often spend long time and take a bunch of selfies trying to get one good selfie that makes them look attractive and different from others. In particular, it is difficult to find a good head posture for a selfie. In addition, it's not easy to have the same head posture for the next one.

SymCollab '18, March 11, 2018, Tokyo, Japan

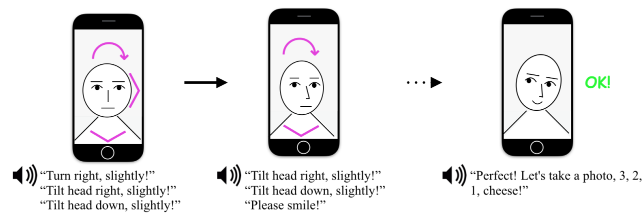

Figure 1: Our guidance system includes visual and voice user interfaces which guide the common user to achieve satisfying selfie based on crowdsourcing results.

With the development of computer-vision technology, especially face recognition and facial-feature extraction approaches, various selfie applications have come out to help the average user take better selfies on mobile platforms. For example, photo corrections and enhancements from a selfie can be provided in a post-process way to specific area of the faces [7]. Recently researchers proposed an approach to enhance the selfies by suggesting expression [16] and face geometries [15] of the users in real-time. However, a user interface to guide common users to take good selfies is still absent in all these previous works.

We propose an interactive selfie user interface to give clear suggestions for users to take selfies based on crowdsourcing results on mobile platforms. There are many factors in taking a good selfie, such as the light condition, background, user expressions, and so on. In contrast to a recent selfie system considering face position and light condition [10], we chose to study head posture (orientation), which is the most important and controllable factor for selfie-taking in this work. Note that the other factors share commonalities in the design of user interfaces and can therefore benefit from this work.

In this work, we first conducted crowdsourcing tasks to define good head postures. In order to avoid factors other than head posture, we generated many virtual selfies using 3D human models to receive the selfie score from crowd workers. The virtual selfie receives higher scores if the users considered it as more attractive. Based on the crowdsourcing results, we developed a real-time user interface to be able to extract the facial features from each frame and estimate the head postures of the users in the real world. A geometry model approach [6] was adopted to implement the real-time head-pose estimation with the facial feature information extracted from each frame. The system can calculate the closest candidates to good head postures from crowdsourcing results, so that the system can instruct a user to achieve good selfie under the guidance of a reference human head. This work is expected to reduce the difficulty of taking good selfies, thus improving user satisfaction while taking selfies.

The main contributions of this work include: We proposed a novel approach to support taking selfies by providing real-time suggestions on head postures; we collected data of head postures in selfie-taking from crowdsourcing results obtained by using hundreds of virtual selfies from 3D models. We implemented multiple ways of interaction between the user and the proposed interfaces as illustrated in Figure 1.

RELATED WORK

Selfie Systems. Creamcam [11] and YouCam [4] to the facial features or add makeup effect onto the human face based on facial feature extraction algorithms. Smile shutter function of Sony DSC T300 provides the support on facial expressions. An interactive selfie application was proposed to suggest face size, face position and lighting direction [10]. In contrast to these works, our system focuses on this important selfie factor, head postures, in real time.

Photo Editing. An interactive photomontage system was proposed to create a composite set of images for the same objects, resulting in a single image for subsequent enhancement [1]. A similar approach was proposed to obtain good flash images using a pair of flash and non-flash images [2].

Aesthetics Feedback. Aesthetics feedback was given for photos taken by mobile users from compositions, color combinations and the aesthetic rating [14]. An aesthetic evaluation approach utilized a peer-rated online photo-sharing website as the input datasets based on machine learning [5].

Assistive Technology. There are other researches that considering about the assistive technology to support the interaction between users and user interface. Voykinska et al. studied about the way that blind people use to interact with social network [13]. Researchers also provided approach to ask remote paid workers to help the blind people with some camera devices and the internet [3]. Similar research asked the crowd to answer the visual questions to help blind people [9].

FRAMEWORK

The framework of this work consists of both offline and online parts as shown in Figure 2. For the offline part, we explored good head postures by crowdsourcing results to define good head posture. We generated hundreds of virtual selfies with different head postures (two types of virtual selfies used in this paper are created based on the 3D models from Unity3D). Among these virtual selfies, we defined "good" head posture as selfies which got higher than 7. For best selfie position, we generated virtual selfies with different distances to the camera to evaluate the attractiveness and chose the distance with the highest score.

For the online part, we extracted the user's facial features for each frame and generated a 3D facial model to calculate the head posture of users using a geometric approach and in real time. Our system calculates the closest good head posture to the user's current head posture and suggests it as the ideal pose

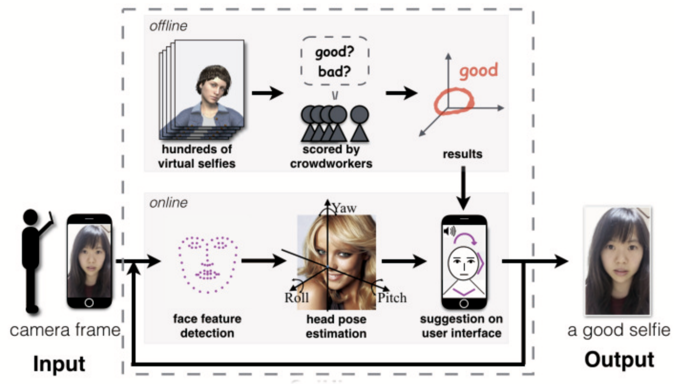

Figure 2: The framework of our selfie guidance system includes offline crowdsourcing tasks and online selfie guidance system.

for the user. To communicate this selfie suggestion, our system shows a visual guide in the user interface to direct the motion of the user's head to achieve the ideal head posture. Voice instructions are also proposed to help the user to move their head using the proposed system. After the user achieves the ideal head posture, the proposed system provides a voice notification and takes the selfies automatically. This helps the user to take a good selfie more easily, thus reducing the difficulty of taking good selfies and enhancing user satisfaction.

Crowdsourcing Tasks for Head Posture

We designed two crowdsourcing tasks to explore good head postures on CrowdFlower. In the first task, the participants were asked to score each selfie, and the average score of each selfie was used to represent its attractiveness. In the first task, we ask the crwodworkers to score each selfie according to its attractiveness from 1 to 10 where 10 is stronly attractive while 1 represents not attractive at all. For each selfie, 5 participants are asked to score it and we adopt the average score of each selfie to represent its attractiveness. The second task is to ask users to evaluate the attractiveness of the selfie with different distances from camera. We adopted the 3D virtual model instead of real selfies for evaluation due to the accurate control of head posture that is possible with virtual selfies; meanwhile, it is difficult to obtain real selfies with exact head postures. The usage of virtual selfies can reduce other factors influencing the selfies in order to specifically evaluate the attractiveness of selfies. To construct a relatively comprehensive selfie database that includes various head postures, we created a database using virtual selfies, which was comprised of 486 different head postures with pitch angle ranging from (-40, 40), roll angle ranging from (-40, 40) and yaw angle ranging from (-30, 20). We sample virtual selfies in every 10 degrees for each dimension.

The results of the crowdsourcing task are shown in the right of Figure 3. The green points represent the head postures that were rated higher than 7, the red points represent the head postures rated higher than 5 but lower than 7 and the grey points represent the head postures rated lower than 5. In the

Figure 3: Virtual selfies obtained good, medium, and bad scores from crowdsourcing results.

first crowdsourcing task, among 486 selfies, the 58 with the highest score were defined as good head orientations. We described the head postures that got scores higher than 7 as "good", 5-7 as medium and under 5 as bad. The proposed selfie guidance system aims to take "good" selfies; thus, we chose the "good" head posture to set the system. We show the selected samples of the selfies which got good, medium and bad scores in Figure 3. From these results we found that the selfies with the front face achieved the highest scores from the crowdsourcing results. However, the common users would not be satisfied with single head pose, and more choices of head poses are preferred.

In the second crowdsourcing task, we generated 6 virtual selfies with different distances considering the length of the arm of human. We utilized the selfie scored as the highest to define the best distances. We accepted a larger tolerance for the distances for two reasons: first, we calculated the distance roughly by calculating the area of the triangle constructed by the far corners of the eyes and nose (this may be affected by different head postures, possibly rendering it inaccurate); from results of this task, we know that the distance doesn't affect the selfie as the head posture (selfies in a quite large range get relatively similar marks).

To obtain continuous assessment for head postures, we interpolated the crowdsourcing results to compute a score function for the head postures using trilinear interpolation. Figure 4 illustrates the results of the interpolation (We display it using several slices). The multi-colored progress bar represents the distribution of color from high score to low score of the head postures; the red region shows the high-scoring head postures, and the blue region shows the head postures that get relatively low scores. As a result, both datasets contain large portions of relatively good selfies as scored by the score function from the crowdsourcing results. Since both datasets are comprised of selfies or photographs that are considered relatively attractive, the experimental results verified that a good head posture itself has a significant impact on the aesthetic of a selfie.

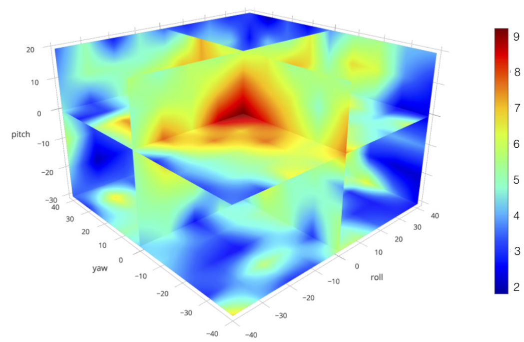

Figure 4: Interpolation of the results of crowdsourcing task.

Facial Feature Extraction Based on Face Landmark

In this work, facial feature extraction plays an important role in the estimation of the head posture of the user. Many researchers have made efforts to solve this problem in order to attain a high accuracy in the recognition of human faces and facial features. Also in this work, we extracted eyes, mouth and nose feature points based on dlib [8] which uses 68 points to describe human faces. This algorithm is a feature-based algorithm based on the HOG combined with a linear classifier and an image pyramid, and sliding window detection scheme.

Head Posture Estimation Based on Geometry Model

Head pose estimation is the key part in this work that can estimate the head orientation of the user based on the facial features we extracted. The head-movement suggestions provided in the user interface are all based on the estimation results. In this work, we estimate the head posture using a geometry model [12]. To introduce how we estimate the head posture, we will first introduce how we describe the head posture. Generally, we define it as the position and the orientation of the head. However, the position of the head can be calculated directly from the image; thus, we focus on how to calculate the orientation from a single image (frame). Head orientation can be described by roll, yaw and pitch angles. The three axes are orthogonal, so that we can calculate each angle respectively. The proposed geometry model assumes that the

Figure 5: Experimental results run on smart phone.

head is bilateral symmetry and the two far corner of eyes and the two mouth corners form a plane and also define a facial normal to show the direction of the nose. This facial normal shows the direction of the head. In contrast to other head-pose estimation algorithms (tracking approach, etc.), the series data of previous frames are not required by the proposed geometry approach to obtain the head pose. Another approaches is direct landmark fitting algorithm which is quite similar to our approach. However, by establishing a facial model, our approach enables relatively simple way to calculate each angle for the head pose which is enough to proof the concept in our research.

Facial Model Conduction

We define L_f to be the distance from the far corner of eyes to the corner of the mouth in the facial model, L_n to be the distance from the nose tip to the nose base, and L_m to be the distance from nose base to the mouth. L_e denotes the distances between two far corner of the eyes. Notice that l_f, l_n, l_m, l_e are the projected values. We calculate the following three ratios to define the facial model: R_m, and R_n. Here, R_m is calculated by L_m/L_f and R_n is L_n/L_f. Because all these ratios are quite stable with respect to different faces, constant values can be adopted to replace the ratios unless a very high accuracy of the orientation is required. For the typical face, $R_m = 0.6$, $R_n = 0.4$. The R_m and R_n are used to calculate the slant as shown in the next section, and the more detailed description can be found in [6].

Estimating of the User's Head Posture

We represent a head posture with τ, which is the angle between the symmetry axis and the facial normal (the projected one in the 2D image) and a slant σ, which is the angle between the normal of the image and the facial normal. The facial normal can be calculated as Eq. (1).

$$\hat{n} = [\sin\sigma\cos\tau, \sin\sigma\sin\tau, -\cos\sigma] \tag{1}$$

Slant σ can be calculated by the normal of the image $d = [x_{normal}, y_{normal}, z_{normal}]$ as $cos\sigma = \mid z^2_{normal} \mid$ (z_{normal} is a normalized value). z_{normal} can be defined using two values v_1 and v_2 which is shown in the Eq. (2) and the projection matrix $(I - dd^T)$.

$$z^2_{normal} = \frac{R_n^2 - v_1 - 2v_2R_n^2 + \sqrt{(v_1 - R_n^2)^2 + 4v_1v_2R_n^2}}{2(1-v_2)R_n^2} \tag{2}$$

Here $v_1 = \left(\frac{l_n}{l_f}\right)^2$ and $v_2 = \frac{y^2_{normal}z^2_{normal}}{(1-y^2_{normal})(1-z^2_{normal})}$.

Roll angle of the head can be simply calculated by the differences between the left eye and the right eye. Here we obtain the left eye and right eye coordinates by the facial feature extraction algorithm. We use a $arctan()$ function to turn the ratio of the differences of the coordinates to degrees to measure the roll angle.

$$\theta_r = \arctan\left(\frac{y_l - y_r}{x_l - x_r}\right) \tag{3}$$

Calculating yaw angle and pitch angle involves the normal we gained before and the x, y, z represents each component of facial normal vector which is defined in Eq. (1) ($x = \sin\sigma\cos\tau$, $y = \sin\sigma\sin\tau$, $z = -\cos\sigma$, $r = \sqrt{x^2+y^2+z^2}$).

$$\theta_y = \arccos\left(\frac{\sqrt{x^2+y^2}}{r}\right), \theta_p = \arccos\left(\frac{z}{r}\right) \tag{4}$$

These three angles describe the head orientation of the end user in the real world. Using this head-orientation information and the crowdsourcing results, the system can provide further suggestions for head-movement to the user in real time.

USER INTERFACE

We implemented the selfie guidance system on the iOS 9.2 using Swift and Objective-C to instruct the user to achieve the head posture that was defined to be good by the crowdsourcing tasks. The system applied the face-landmark algorithm to process the camera frame in question. We adopted the eyes, nose and mouth points to estimate the head posture. For the first frame, the system extracts the facial features to initialize the face-geometry model. It then calculates the head posture with the facial features from each frame and the face-geometry model. The whole system is described in Algorithm 1. The user interface of our selfie guidance system consists of a visual user interface and voice instruction. We combined these two methods for interacting with the user to provide new user experiences and reduce the difficulty for users in understanding the instructions when taking selfies. For pilot study, we implemented two selfie guidance systems: only visual user interface; both visual and voice user interfaces. From the user feedback, the participants reported that it is interesting and convenient to following the voice instructions.

Visual User Interface

The visual user interface is implemented as shown in Figure 6. To reduce the operational difficulty of head movement, the user interface provides suggestions for roll, yaw and pitch angles separately. Thus, the user can move their head one direction at a time. With the combination of these head movements, the user can easily achieve good head posture. The arrows shown in the top and bottom of the user interface suggest the pitch angle and ask users to adjust their head by tilting their head up or down slightly. The arrows in the right and left

Algorithm 1: Real-time processing for taking a selfie

input : Camera frame including a human face.
output : A good selfie stored in the album.

1. Define good head postures by crowdsourcing approach using 486 virtual selfies with different head postures. This results in 58 good head posture sets.
2. Extract the eyes, nose and mouth points from the frame based on the face landmark algorithm which uses HOG feature combining with a linear classifier to estimate a vector S (S contains the coordinates information of each facial point).
3. Use the facial features extracted from the frame to establish geometry facial model F.
4. Based on F and the facial feature information to estimate the head posture P_c from the current frame which is the current head posture of the user.
5. Calculate the Euclidean distances between P and each good head posture P_g (defined by crowdsourcing task) to find a minimum value which is defined as "the closest good head posture" P_{cg} (notice here the system calculates the good head posture once until the frame count f_c to be reset).
6. Compare each angle defined in P_{cg} and those of P_c to provide different suggestions
7. Go back to step 1 if the P_c doesn't equal to the P_{cg} (with a tolerance Δd).
8. Terminate the loop when P_c is close enough to the P_{cg} (with Δd); take photo and store to the album and reset the frame count f_c.

side of the screen instruct users to turn their head right or left slightly. The semicircle arrows instruct the user to tilt his or her head right or left. These arrows will keep showing in the user interface until the user's head achieves the right position.

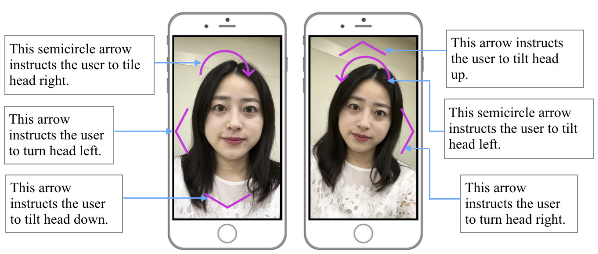

Figure 6: Visual user interface of the developed selfie guidance system.

Voice User Interface

In the voice instruction part, the selfie-supporting system provides voice instructions to help the user to understand how to move their head to match the good head posture. There are four types of voice instructions that indicate suggestions for proper head orientation and distance, and that check for the presence of a smile. For the head orientation suggestions, the system displays "tilt head right slightly!" and "tilt head left slightly!" to suggest the roll angle; "tilt head up slightly!" and "tilt head down slightly!" to suggest the pitch angle; and "turn left slightly!" and "turn right slightly!" for the yaw angle, and these suggestions correspond to the arrows on the visual user interface. When the user has the right posture, and is ready to take selfies, the system will display "Perfect! Let's take a photo, 3, 2, 1, cheese!" This notifies the user when to take the selfie and provides extra time for them to adjust their expressions.

USER STUDY

We designed a comparison user study in order to evaluate the validity of the developed selfie-supporting system. (We also designed pilot study to compare the system that only have visual instructions with the system that have visual and voice instructions). Of the 8 participants, 2 are female and others are male. The ages of all the participants are between 22 and 28. Some of them spend a great deal of time taking selfies while others take very few or no selfies. We mainly asked them to use our developed system and a normal camera to experience the difference and asked them to fill out a questionnaire to evaluate the system with regards to several aspects.

There were two tasks, and the participants were asked to complete the two tasks in a random order. Task 1 asked the participants to take several selfies using a normal camera without any suggestions during the process. We asked them to choose 5 selfies that they were satisfied with. They could take as many selfies as they wanted until they got 5 satisfactory selfies. Task 2 asked the participants to use the developed selfie-supporting system to take 5 or 6 selfies. (The reason some of the users take more than 5 selfies is that we were trying to avoiding multiple selfies with the same head posture). In this task, they followed the suggestions about their head posture and also heard voice instructions while taking selfies.

In the experiment, 4 of the participants were asked to complete Task 1 first and then complete the Task 2 while the other 4 users were asked to complete Task 2 first and then Task 1. In addition to this comparison user study, we also asked them to use our selfie-supporting system with the tripod settings to experience taking selfies with the visual- and voice-instruction user interface while at a distance further than human arms can reach. After all the tasks above, they were also asked to fill a questionnaire to evaluate their satisfaction with the system, the user interface and also the attractiveness of the head postures provided by this system.

RESULTS

We applied the score function according to the crowdsourcing results to the result of the user study. Figure 7 and Figure 8 illustrate examples in which the user used a normal camera to take selfies using our proposed system. From the result we can see that the selfies that taken by our system get higher scores on average. For example, for the second user, the average score of selfies taken by a normal camera was 6.94; meanwhile, the selfies taken by our system were scored 7.7 on average. Here we divided the user into two groups according to whether they were good at taking selfies or not. Figure 7 shows the examples of users who were not good at taking selfies. We can see from this result that the proposed system helps the users take better selfies overall. In these examples, the average score of selfies taken by our system increases around 30% compared to the selfies taken by normal camera. Figure 8 shows the examples of users who exhibited relatively

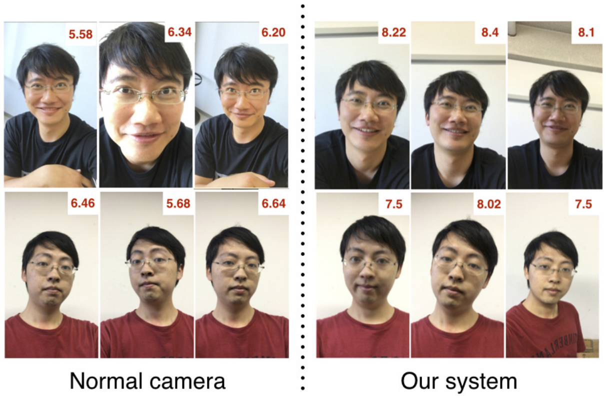

Figure 7: Scoring results of user's selfies taken by normal camera and our system for those who are not good at taking selfie.

better skill in taking selfies. They got higher scores even using a normal camera. However, our system still helped them get higher score on average. The proposed system can also support them by providing more head postures so as to give them options for different appearances.

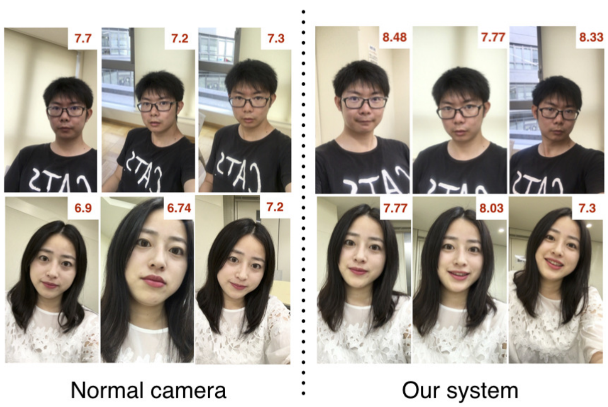

Figure 8: Scoring results of user's selfies taken by normal camera and our system for those who are good at taking selfie. Note that there are more options of head postures by using our proposed system than normal camera.

The questions in our questionnaire queried the user regarding system satisfaction, feedback on head posture provided by the system, the visual user interface and the voice instruction. For these statements, users were asked to score the statement from 1 to 5, where 1 represents "strongly disagree" and the 5 represents "strongly agree." The evaluation and analysis of the system have been done according to the result of the user study. Figure 10 (a) shows the results of the comparison of the normal camera and our system. As we can see in the Figure 10, 75% of users think that our system is more convenient and 25% of users think that the both systems are equivalent. Figure 10 (b) illustrates what users think of the head postures provided in our system. The results show that 87.5% of users agree that our selfie supporting system provide more choices for them when taking selfies (25% of users choose strongly agree; 62.5% of users choose agree); 12.5% of users neither agree nor disagree.

Figure 9 (a) shows the user feedback concerning the developed selfie-support system. We asked users about whether the application helped them take better selfies and if the function suggesting head posture was needed, in their opinion. We also asked about whether they thought it was easier to take good selfies with the supporting system. The chart shows the average score of each question and we can see that the users have quite high satisfaction with the system. Figure 9 (b) shows the average scores for evaluations about the head posture, which show that users confirm that the head postures provided from the application are good. Figure 9 (c) shows the average scores for evaluations about the visual user interface. In this part we asked users whether they could understand the meaning of the arrows and if they could realize which arrows suggested roll, yaw and pitch angle, respectively. From these results we find that the users can understand the meaning and how to follow the suggestions well.

Figure 11 shows the evaluations of the voice instructions. To evaluate whether the voice instructions work when the user is taking selfies, we asked the user about two aspects: whether they were aware of the voice instruction and whether they understood the meaning of the voice instructions. We also asked the users whether the voice instructions were interesting (Q24), whether they were helpful (Q25) and whether they were easy to understand (Q26). According to the results, users were able to notice and understand the voice instruction well.

CONCLUSION

In this work, we proposed a real-time selfie support system which provides suggestions for the head posture of the user. This proposed system aimed to help user to take good selfies. Two crowdsourcing tasks were conducted to explore good head orientations and the best distance from the camera (position). In each crowdsourcing task, we generated virtual selfies to create a database and asked crowdworkers to score each virtual selfie by its attractiveness. After the crowdsourcing task, we developed a selfie system which provides real-time suggestions for head postures. We implemented both the visual user interface, in which arrows suggest movements of the head as well as the voice instructions, which also give guidance to the users on how to move their head and in which direction. A user study was conducted where participants were asked to take selfies using two cameras (a normal camera and the camera with our head-posture support system). According to the study results, the proposed system can help users to take good selfies and improve user satisfaction.

In this work, we explored good head posture using a 3D virtual model and applied the crowdsourcing results to all the users. The aesthetic evaluation may vary due to different cultural backgrounds or knowledge differences, so this could be further

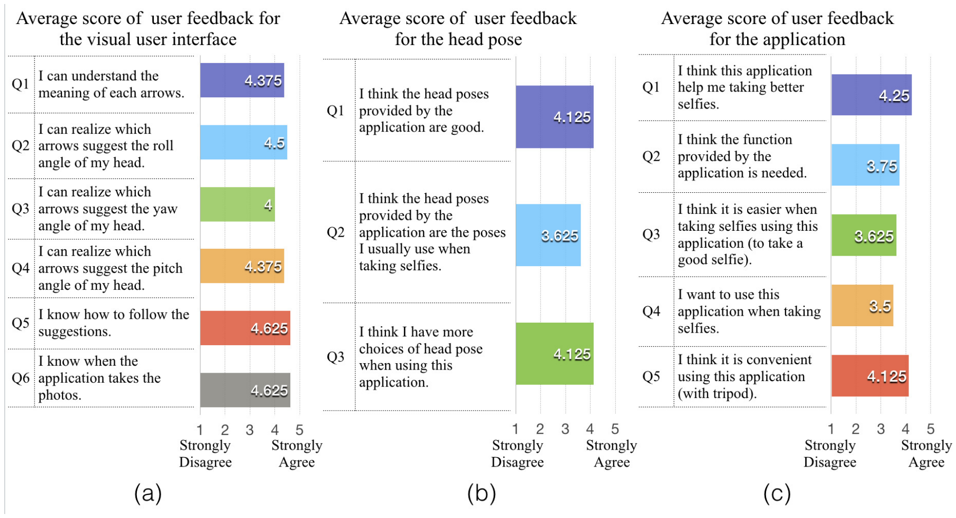

Figure 9: The average score of the user feedback for the application (a); the average score of user feedback for the head postures provided by the application (b); the average score of user feedback for the visual user interface (c).

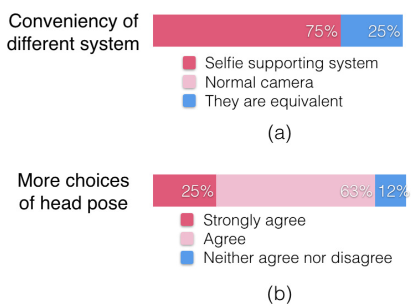

Figure 10: The user feedback for the comparison of the normal camera and the developed selfie supporting system (a); the user feedback for whether they have more choice of head postures with our selfie supporting system (b).

studied by considering more parameters for certain users. The integration of other selfie factors is not difficult to implement in our selfie guidance system. In the future, it is worth exploring the relationship between the substantial amounts of selfie data and methods for subjective crowdsourcing evaluations using deep learning to achieve satisfying selfies. An alternative way to evaluate the quality of selfie can be explored by example-based approach to compare the reviews from social network or ask crowdworkers to score the selfies taken by our system and other work with functionality. The crowdsourcing-based evaluation approach can also be applied to other areas such as evaluating the accuracy of face recognition.

ACKNOWLEDGMENTS

We thank all participants for joining our case studies, and Morph 3D for sharing the head model in Figure 3. This work was supported by JST CREST Grant Number JPMJCR17A1 and JSPS KAKENHI Grant Number JP17H06574, Japan. Haoran Xie is funded by Epson International Foundation.

REFERENCES

1. Aseem Agarwala, Mira Dontcheva, Maneesh Agrawala, Steven Drucker, Alex Colburn, Brian Curless, David Salesin, and Michael Cohen. 2004. Interactive Digital Photomontage. *ACM Trans. Graph.* 23, 3 (Aug. 2004), 294–302. DOI: http://dx.doi.org/10.1145/1015706.1015718

2. Amit Agrawal, Ramesh Raskar, Shree K. Nayar, and Yuanzhen Li. 2005. Removing Photography Artifacts Using Gradient Projection and Flash-exposure Sampling. *ACM Trans. Graph.* 24, 3 (July 2005), 828–835. DOI: http://dx.doi.org/10.1145/1073204.1073269

3. Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, and Tom Yeh. 2010. VizWiz: Nearly Real-time Answers to Visual Questions. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A) (W4A '10)*. ACM, New York, NY, USA, Article 24, 2 pages. DOI: http://dx.doi.org/10.1145/1805986.1806020

4. Perfect Corp. 2015. YouCam Perfect. (2015). http://www.perfectcorp.com/stat/product/CyberLink_app/Perfect_Corp/enu/design/perfectCorp/index.jsp

5. Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. 2006. Studying Aesthetics in Photographic Images Using a Computational Approach. In *Proceedings of the 9th European Conference on Computer Vision - Volume Part III (ECCV'06)*. Springer-Verlag, Berlin, Heidelberg, 288–301. DOI: http://dx.doi.org/10.1007/11744078_23

6. Andrew Gee and Roberto Cipolla. 1994. Determining the gaze of faces in images. *Image and Vision Computing* 12, 10 (1994), 639 – 647. DOI: http://dx.doi.org/10.1016/0262-8856(94)90039-6

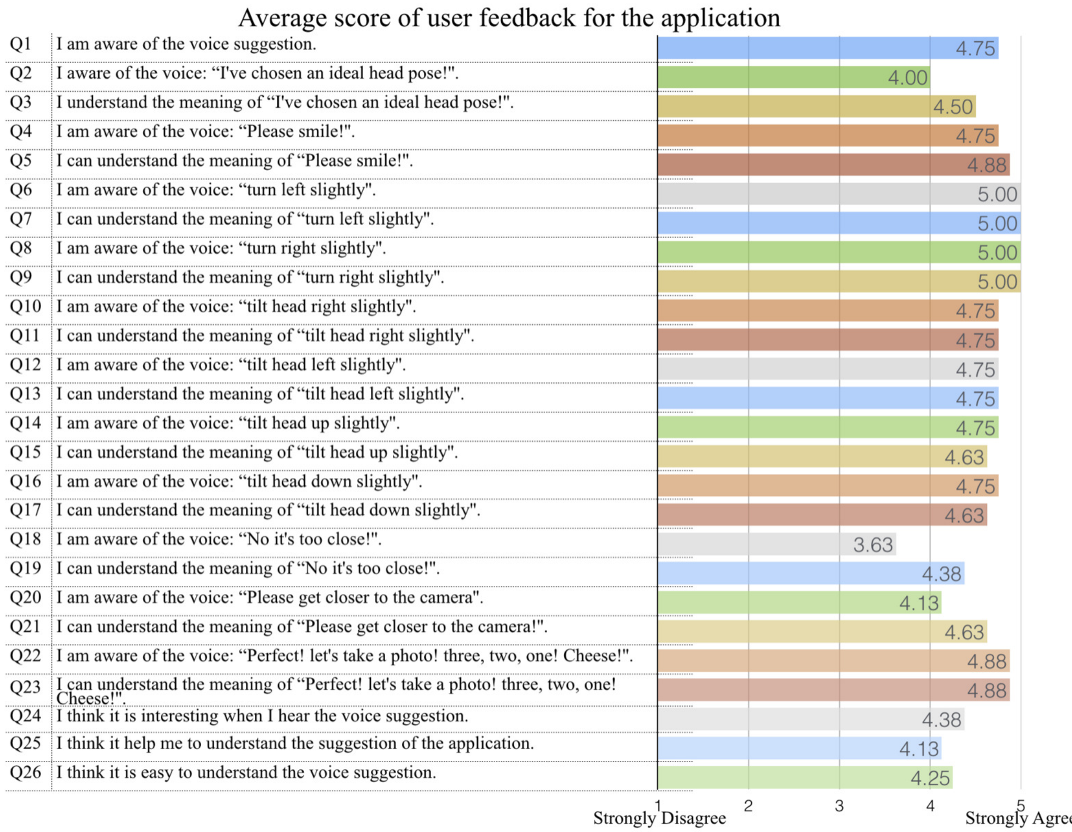

Figure 11: The user feedback about the voice instructions.

7. Neel Joshi, Wojciech Matusik, Edward H. Adelson, and David J. Kriegman. 2010. Personal Photo Enhancement Using Example Images. *ACM Trans. Graph.* 29, 2, Article 12 (April 2010), 15 pages. DOI: http://dx.doi.org/10.1145/1731047.1731050

8. V. Kazemi and J. Sullivan. 2014. One millisecond face alignment with an ensemble of regression trees. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 1867–1874. DOI: http://dx.doi.org/10.1109/CVPR.2014.241

9. Walter S. Lasecki, Phyo Thiha, Yu Zhong, Erin Brady, and Jeffrey P. Bigham. 2013. Answering Visual Questions with Conversational Crowd Assistants. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS ’13)*. ACM, New York, NY, USA, Article 18, 8 pages. DOI:http://dx.doi.org/10.1145/2513383.2517033

10. Qifan Li and Daniel Vogel. 2017. Guided Selfies Using Models of Portrait Aesthetics. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS ’17)*. ACM, New York, NY, USA, 179–190. DOI: http://dx.doi.org/10.1145/3064663.3064700

11. LoftLab. 2016. CreamCam Selfie Filter. (2016). http://creamcamapp.com

12. Baback Moghaddam and Alexander P. Pentland. 1994. Face recognition using view-based and modular eigenspaces. *Proc. SPIE* 2277 (1994), 12–21.

13. Violeta Voykinska, Shiri Azenkot, Shaomei Wu, and Gilly Leshed. 2016. How Blind People Interact with Visual Content on Social Networking Services. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW ’16)*. ACM, New York, NY, USA, 1584–1595. DOI: http://dx.doi.org/10.1145/2818048.2820013

14. Lei Yao, Poonam Suryanarayan, Mu Qiao, James Z. Wang, and Jia Li. 2012. OSCAR: On-Site Composition and Aesthetics Feedback through Exemplars for Photographers. *International Journal of Computer Vision* 96, 3 (2012), 353–383.

15. Mei-Chen Yeh and Hsiao-Wei Lin. 2014. Virtual Portraitist: Aesthetic Evaluation of Selfies Based on Angle. In *Proceedings of the 22Nd ACM International Conference on Multimedia (MM ’14)*. ACM, New York, NY, USA, 221–224. DOI: http://dx.doi.org/10.1145/2647868.2656401

16. Jun-Yan Zhu, Aseem Agarwala, Alexei A Efros, Eli Shechtman, and Jue Wang. 2014. Mirror Mirror: Crowdsourcing Better Portraits. *ACM Transactions on Graphics (SIGGRAPH Asia 2014)* 33, 6 (2014).