TouchPhoto: Enabling Independent Picture-taking and Understanding of Photos for Visually-Impaired Users *

Yongjae Yoo[1][0000−0002−5831−6492], Jongho Lim[1][0000−0001−9972−5703], Hanseul Cho[1], and Seungmoon Choi[1][0000−0002−5889−1083]

Pohang University of Science and Technology,
37673 Pohang, Gyeongbuk, South Korea
{dreamseed,lcdplayer,johanseul,choism}@postech.ac.kr
http://hvr.postech.ac.kr

Abstract. Photographs are a powerful medium for recording moments and sharing them with others. However, visually-impaired users have quite limited access to photograph's benefits. In this paper, we present an integrated system TouchPhoto, which provides visual-audio-tactile assistive features to allow visually-impaired users to take and understand photographs independently. A user can take photographs with auditory guidance and record several audio tags to aid recall of the photograph's content. For comprehension, a user can listen to the audio tags embedded in a photograph and also touch the photograph using an electrostatic friction display. The latter is done after salient features in the photograph, e.g., human faces, are extracted to facilitate tactile recognition.

Keywords: blind photography · assistive technology · multimodal interaction · electrovibration.

1 Introduction

Photography enables users to visually capture the moment, preserve it, and share it with others. However, approximately 250 millions of visually-impaired users hardly benefit from photo-related activities, although most of them are frequently exposed to such situations [3]. They have to seek help from sighted people, and if no help is available, they hold on or give up the task [10].

Several applications have been developed to support photo-related activities for visually-impaired users, mostly relying on sensory substitution. Their goals include assisting photo taking [9], recording ambient sounds and explanations with photos [3], and describing and reading the content of pictures [4]. However, attempts to directly transmit the graphical contents of photos by means of haptic exploration, as is done in tactile graphics, have not been made. This approach

* This project has been supported by an X-Project program (2016R1E1A2914792), and an IITP program (2017-0-00179), all funded by Korea Government (MSIP).

might turn out to be an effective modality for understanding the content of photos and reminding related memories and experiences.

In this paper, we present an integrated system of TouchPhoto, which is envisioned as an integrated system providing visually-impaired users with multimodal assistance to assist their independent undertaking of photo-related activities. TouchPhoto provides multimodal sensory feedbacks–visual, audio, and haptic stimuli–to deliver information. For haptic interaction, we use an electrostatic friction display for its wide applicability to regular touchscreen devices such as tablet PCs, as well as its convenience of use.

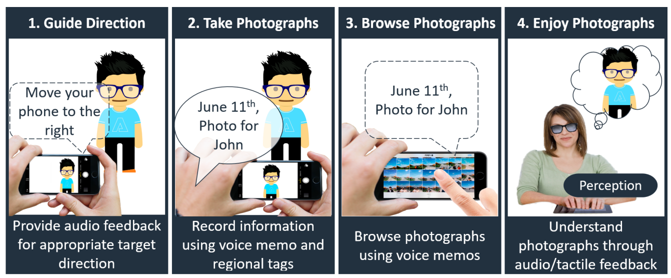

Fig. 1. Use scenario of Touchphoto.

2 Implementation of TouchPhoto

In this section, we present an overview and general design of TouchPhoto. At present, the main features of TouchPhoto support only photographs of people. This emphasizes the social and emotional function of photography, as a way to remember people and related events. Reviving such personal and emotional memories via audio-haptic multimodal interaction is our first goal, which also agrees to the general nature of the tactile sense [8]. Fig. 1 shows the use scenario of TouchPhoto.

2.1 Apparatus

As shown in Fig. 2, TouchPhoto uses a regular smartphone to take photographs. We made an application running on Android with the support of Google TalkBack. The application has two main functions, a camera and a photo album for visually-impaired users. The portrait photographs taken are processed to extract facial features of a nose, eyebrows, eyes, lips, and a face contour using an external API (Face++, Face Landmark) [2]. Each photograph is augmented with the coordinates of regional tags, and facial features, and audio data recorded by the camera app. The photographs are uploaded to a cloud storage service (Amazon S3; simple storage service).

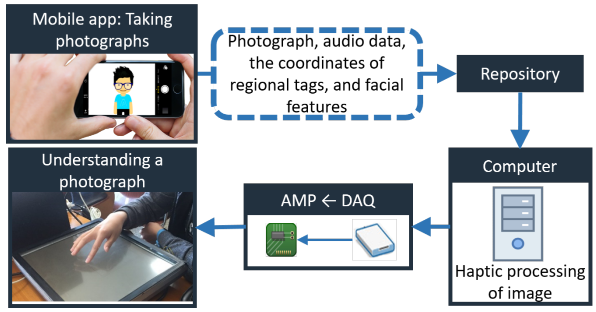

Fig. 2. The structure and data flow of Touchphoto.

We learned from visually-impaired users that they prefer a large workspace (15–17 inches; as large as a laptop PC) for the exploration of tactile graphics. Such large electrostatic display have not been adopted to commercialized products yet, so we built a large electrostatic display by attaching a transparent electroconductive film (3M, SCT3250) onto a 17-inch LCD monitor. An IR (infrared radiation) frame (E&T Tech Inc., Model T17) was mounted on the display for finger tracking. A PC downloads the photographs and controls the display using a data acquisition card (National Instruments; NI-6251) and a high-voltage amplifier (Piezodrive; MX200b).

2.2 Photo-taking

While taking a picture, the Android app of TouchPhoto detects human faces in the photograph using OpenCV (Fig. 3, the left panel). Then, it provides speech guidance to help the visually-impaired user move the detected face to the center of the screen, for example, "(Move your phone) to the right." During camera aiming, ambient sound was also recorded.

After taking a photograph, the user can record a voice message about the photograph. This voice memo augments the photograph with the user's own memories and experiences, which are valuable for the user's better recollection. The voice memo is also used as an auditory index in the photo album.

TouchPhoto supports another type of auditory tags, named as regional tags. By pinching on a certain region in the photograph, a user can set a regional tag and record a voice message specific to that region. Regional tags are designed to be added by a friend, but some low-vision users can do that by themselves. This is useful for pictures with a number of people or objects; later visually-impaired users can easily recognize the content of the photograph by touching the regional tags and listening to the voice memos. To our knowledge, regional auditory tags are a new feature to TouchPhoto.

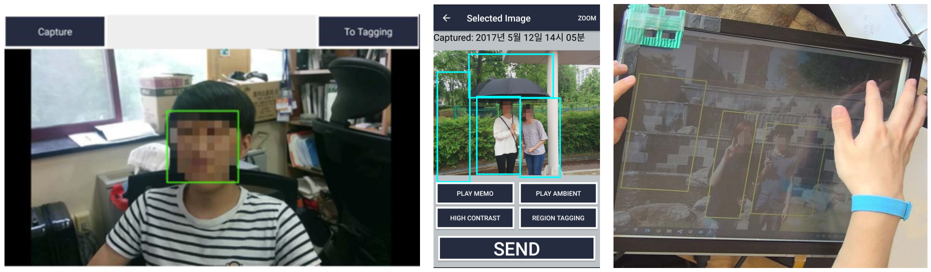

Fig. 3. Picture-taking, browsing, and understanding interface of Touchphoto.

2.3 Browsing and Understanding

Browsing and understanding the contents of photos can be done in both the Android app and tactile display. If s/he selects a photo in the album, menus for high-contrast image, regional tags, ambient sound, and sending the photo to tactile display are appeared on the screen. (Fig. 3, the middle panel) When the user selects 'Send' menu at the bottom of the screen, the contents would be sent to the tactile display controlled by a program running on a PC (Fig. 3, the right panel). The photograph sent from the Android app is displayed on the 17-inch LCD screen with rectangles representing audio-tagged regions.

The screen also shows two buttons. One button toggles regional audio tags. When the regional tag function is toggled on, rectangular representations are shown on the screen, and the user can find them by scanning on the screen and perceiving simple electrovibration rendered when a tagged region is touched. A regional audio tag is played back when the user taps on the area. The other button controls the magnifier function. While the magnifier function is toggled on, the region of the human face is zoomed for easier tactile exploration. The user can perceive the 3D geometries and textures of important face features, like a nose, eyebrows, eyes, lips, and face contour, by touch.

3 Tactile Rendering

3.1 3D Rendering

In graphics, 3D features are often presented on a 2D display. We tried the similar approach using a surface haptic display. Since a human face consists of several 3D features, height rendering might be essential for highly informative presentation.

The first step of our tactile rendering is to extract the 2D coordinates of facial features in the photograph. We use an external API (Face++, Landmark Analysis) [2], which finds the positions of facial features, such as a nose, eyebrows, eyes, lips and the face contour, in a photograph. Then we fit a 3D face model (CANDIDE-3 model [1]) to the extracted facial features to obtain the height coordinates. These 3D coordinates of the facial features are used for tactile rendering.

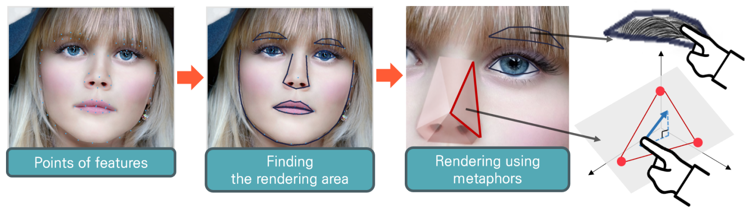

Fig. 4. Tactile rendering procedure.

The literature showed the feasibility of height rendering on an electrostatic friction display using gradient-based lateral force modulation [5, 7]. We apply this method to render the 3D features in a face. If a user's finger scans the surface of tactile display, the position of contact was tracked by the infrared touchscreen frame that is attached on the display in real time. When the position of contact was inside the region of facial features, we calculated a 2D vector $\overrightarrow{v} = (x_2 - x_1, y_2 - y_1)$, by differentiating the two consequent input coordinates of (x_1, y_1) and (x_2, y_2). The height difference, $\Delta z = z_2 - z_1$, obtained from the fitted 3D face model, were calculated by interpolations in the plane which is being touched. (Fig. 4) The tangential component, $\frac{\Delta z}{||v||}$ decided the output voltage. If it exceeds 2.0 $(\tan(63.5^{o}))$, it was clipped to 2.0, due to the amplifier's voltage limitations. The amplitude of output signal was calculated as $A = 20 \times \frac{\Delta z}{||v||} + 10.0$ (V), where 10.0 is the approximated AL point of our display's electrovibration. Thus, the output waveform was $A \times \sin(2\pi f t) + A$ or $A \times sgn(\sin(2\pi f t)) + A$, where A is the DC offset component. The frequencies (f) for the corresponding facial landmarks were determined by iterative procedures (Table 1).

Table 1. Tactile rendering parameters for textures.

	Eyebrow	Eye	Nose	Lip	Face contour
Target texture	Hair	Glass	Skin	Soft	Embossed carving
Waveform	Square	Sine	Sine + Square	Sine	Square
Frequency (Hz)	40	550	100	50	100
Amplitude (V)	80	80	40	60	80

3.2 Texture Rendering

Changing the frequency and amplitude of vibration provides the sensations of textures with different roughnesses [6]. For better identification of facial features, we render different textures for the major facial landmarks using the parameter values shown in Table 1. The parameter values are tuned to build analogies to the features. For example, eyes were rendered by high-frequency, smooth vibrations, while lips were by a bit rougher vibrations, albeit not too much. Hairy parts such as eyebrows were rendered using low-frequency, rough textures. The tactile rendering procedure is illustrated in Fig. 4.

4 Conclusions

In this paper, we have presented TouchPhoto, an integrated system to assist visually-impaired users to take, manage, and understand photographs independently. Visually-impaired users can take photographs with auditory guidance, as well as record several audio tags to aid recall of the photograph's content. For comprehension, a user can listen to the audio tags embedded in a photograph and also touch the photograph using an electrostatic friction display. We also implemented a tactile rendering algorithm of human faces that allows users to perceive height and texture using gradient-based lateral force modulation. We envision that the outcomes of this study could contribute to a better design of accessible photography apps for visually-impaired users. As a future work, we will improve haptic devices and rendering methods for understanding photographs, while not sacrificing the small and handy device form factor.

References

1. Ahlberg, J.: CANDIDE-3–an updated parameterised face. Tech. rep., Linkŏping University (2001)
2. FacePlusPlus Inc.: Face++ cognitive services (2015), http://www.faceplusplus.com/
3. Harada, S., Sato, D., Adams, D.W., Kurniawan, S., Takagi, H., Asakawa, C.: Accessible photo album: Enhancing the photo sharing experience for people with visual impairment. In: Proceedings of the 31st SIGCHI Conference on Human Factors in Computing Systems (CHI '13). pp. 2127–2136 (2013)
4. Jayant, C., Ji, H., White, S., Bigham, J.P.: Supporting blind photography. In: Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'11). pp. 203–210 (2011)
5. Kim, S.C., Israr, A., Poupyrev, I.: Tactile rendering of 3d features on touch surfaces. In: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST'13). pp. 531–538 (2013)
6. Klazky, R.L., Lederman, S.J., Hamilton, C., Grindley, M., Swenesen, R.H.: Feeling texture through a probe: Effects of probe and surface geometry and exploratory factors. Perception & Psychophysics **65**(4), 613–631 (2003)
7. Osgouei, R.H., Kim, J.R., Choi, S.: Improving 3d shape recognition with electrostatic friction display. IEEE Transactions on Haptics p. (Preprint) (2017)
8. Paterson, M.: The Senses of Touch: Haptics, Affects and Technologies. Bloomsbury Academic (2007)
9. Vázquez, M., Steinfeld, A.: An assisted photography framework to help visually impaired users properly aim a camera. ACM Transactions of Computer-Human Interaction (TOCHI) **21**(5), 25:1–25:29 (2014)
10. Voykinska, V., Azenkot, S., Wu, S., Leshed, G.: 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). pp. 1584–1595 (2016)