Received: 10 November 2021 | Revised: 10 May 2022 | Accepted: 23 May 2022

DOI: 10.4218/etrij.2021-0429

ETRI Journal WILEY

ORIGINAL ARTICLE

Virtual portraits from rotating selfies

Yongsik Lee[1] | **Jinhyuk Jang**[2] | **Seungjoon Yang**[3]

[1]Telecommunications & Media Research, Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea

[2]Artificial Intelligence Research, Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea

[3]Department of Electrical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Rep. of Korea

Correspondence
Seungjoon Yang, Department of Electrical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Rep. of Korea.
Email: syang@unist.ac.kr

Funding information
This work was supported by the U-K BRAND Research Fund of Ulsan National Institute of Science & Technology (UNIST) (1.210040.01) (Contribution Rate: 34%). This work was supported by Institute for Information and Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIP) (No. 2018-0-00198, Contribution Rate: 33%). This research was supported by 2022 Cultural Heritage Smart Preservation & Utilization R&D Program of Cultural Heritage Administration, National Research Institute of Cultural Heritage (Project Name: Development of AI based CAD conversion technology for traditional architecture drawing images, Project Number: 2022A02P03-001, Contribution Rate: 33%).

Abstract

Selfies are a popular form of photography. However, due to physical constraints, the compositions of selfies are limited. We present algorithms for creating virtual portraits with interesting compositions from a set of selfies. The selfies are taken at the same location while the user spins around. The scene is analyzed using multiple selfies to determine the locations of the camera, subject, and background. Then, a view from a virtual camera is synthesized. We present two use cases. After rearranging the distances between the camera, subject, and background, we render a virtual view from a camera with a longer focal length. Following that, changes in perspective and lens characteristics caused by new compositions and focal lengths are simulated. Second, a virtual panoramic view with a larger field of view is rendered, with the user's image placed in a preferred location. In our experiments, virtual portraits with a wide range of focal lengths were obtained using a device equipped with a lens that has only one focal length. The rendered portraits included compositions that would be photographed with actual lenses. Our proposed algorithms can provide new use cases in which selfie compositions are not limited by a camera's focal length or distance from the camera.

KEYWORDS
lens simulation, scene analysis, selfie, view synthesis

1 | INTRODUCTION

"Selfie" has become a common phrase as people take photos of themselves in interesting places and share them on social networking sites. The devices that people use to take selfies are usually equipped with a limited set of lenses. Furthermore, the distance at which selfies are typically taken is limited to the length of the arm. As a result,

selfie composition is limited to subjects in the center with a small amount of background in the field of view of a wide-angle lens. We present algorithms for rendering portraits with various compositions using multiple selfies via view synthesis. View synthesis has been studied to obtain free-viewpoint images and videos from multiple cameras [1–3]. In model-based synthesis techniques [4,5], geometric models of the scene are constructed. Furthermore, information on the scene, such as depth maps, is constructed in image-based synthesis techniques [6–9]. A view from a camera with arbitrary lenses at any location is rendered using the constructed model or information.

We analyze the scene from multiple selfies taken while a user is spinning around in this study. These selfies are used to create the scene's depth map. Then, a view from a camera with a longer focal length, at a distance farther from the subject, is rendered. A synthesized selfie is a portrait captured with a telephoto lens. Perspective and lens characteristics change as a result of the longer focal length and increased distance between the subject and camera. A view from a camera with a shorter focal length would necessitate images that include the upper and lower parts of the scene, which are absent from the selfies. Instead, a panoramic view of the scene is rendered, with a user's image placed at an arbitrary location. In contrast to view synthesis techniques that provide free viewpoints, our proposed method provides only a view from a camera moved in a straight direction away from a subject. The perspectives are broad enough to allow for portraits taken with telephoto lenses. We were able to avoid common difficulties in view synthesis, such as filling the newly exposed area around the subjects, because the viewpoints are limited and the selfies are captured while the user is spinning [6–9] or reconstructing a surface from skewed angles [4,5]. Our proposed algorithm is simple enough to run on mobile devices with limited computational power.

In our experiments, we used selfies photographed with a 28-mm focal length lens, typically found on mobile devices such as cellular phones. Portraits photographed with virtual lenses having focal lengths of 85 mm, 105 mm, 200 mm, and 300 mm were rendered from three selfies. Portraits photographed with the subject and camera moved away from the background were also rendered. The panoramic selfies were constructed from video selfies taken with the 28-mm focal length lens. The rendered images were compared to the images photographed using the same focal length lenses at the same rearranged distances to validate the performance. Our proposed method provided images of a scene viewed from a camera with a selected focal length for the subjects and camera at new locations accurately. The perspective and lens characteristic changes due to the lens and distance changes are also rendered accurately. Compared to digital zoom [10], our proposed method provided higher resolution images. Our proposed method can be used to provide portraits with a wide range of focal lengths. The proposed algorithms can provide new use cases wherein the composition of selfies is not limited by the focal length of a camera or the distance from a camera.

The contributions of this work can be summarized as follows:

(i) We presented a method for generating a portrait of a user from selfies taken with different focal length lenses at virtual camera locations. There is no need for a photographer to take the user's portrait from a long distance away. Images with long focal length lenses that the user's camera does not have can be synthesized. Long focal length lens images have no loss of resolution, unlike widely used digital zoom
(ii) We provided a method to synthesize a panorama from selfies. There is no need for a photographer to scan the scene away from the user to capture a panorama that includes the user. After that, the user's image in the panorama can be determined.

The remainder of this paper is organized as follows. Section 3 presents the first use case of rendering a view from a virtual camera from multiple selfies. Section 4 presents the second use case of rendering a panoramic selfie from a video selfie. Section 5 presents the experimental results. Finally, Section 6 provides the concluding statements.

2 | RELATED WORKS

View synthesis has been studied to obtain free-viewpoint images and videos from multiple cameras [1–3]. In model-based synthesis techniques [4,5], geometric models of the scene are constructed. Starck and Hilton [4] set up a multiple camera studio and simultaneously capture a moving person. Multiple images are used to reconstruct the scene geometry model. Lin and Xiao [5] display a novel view rendering with a 3D point cloud derived from a reconstructed 3D scene.

For 3D scene representation, suggest a multilayered variate-resolution sampling technique. Furthermore, in image-based synthesis techniques [6–9], information on the scene such as depth maps are constructed. Starck and Hilton [6] proposes the synthesis of novel views of people from multiple view videos. Depth-image-based rendering

(DIBR) is the process of synthesizing virtual views of a scene from the video. DIBR suffers from holes and artifacts. Liu and other [7] generate a better arbitrary view based on DIBR for filling the holes. Cheng and others [8] propose a new algorithm developed for recovering the large disocclusion regions in DIBR. Chen and others [9] present a virtual view synthesis approach based on asymmetric bidirectional DIBR. Chen and others [9] reduce rendering time and achieve significant hole reduction. A view from a camera with arbitrary lenses in any location is rendered using the constructed model or information.

Recently, view synthesis using deep learning techniques is being investigated [11–13]. Using neural light-transport field (NeLF), Chen [11] regenerates the face at an arbitrary point by learning with multiple face images. Freer and other [12] separate a person from a single photo and then regenerate the background and recompose with the person. Leimkühler and Drettakis [13] use generative adversarial networks to generate random viewpoint images from multiple face photos (GANs). There are mobile phone apps that simulate images from virtual lenses. For example, the Viewfinder Preview app [14] provides virtual focal length lenses by digital zooming. It is assumed, however, that a separate photographer is taking a photo with the virtual lens far away from a subject. Pitu app [15] simulates images of a telephoto lens by blurring the background. It will work with a selfie, but the perspective and composition will be different than when using the actual lens.

Based on the multiple geometries, the proposed method synthesizes a view from a virtual camera at various locations and settings. The most significant innovation is the use of selfies for the synthesis of portraits and panoramas with the user inside. Other methods typically assume that a separate photographer takes a portrait or a panorama with a separate subject inside. The proposed method assumes a scenario where a user takes selfies to synthesize a portrait of himself/herself or a panorama with himself/herself inside.

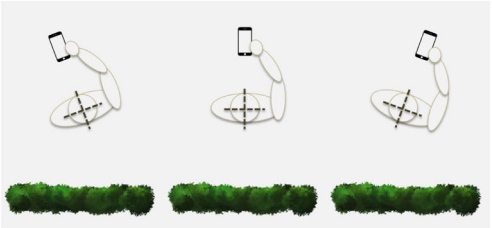

FIGURE 1 User scenario with three selfies: (A) first, (B) second, and (C) third shots are taken at the same location while a user is spinning around

3 | PORTRAIT FROM MULTIPLE SELFIES

3.1 | Scene analysis

Consider the following scenario: A user takes multiple selfies at the same location while spinning around. The user scenario is depicted in Figure 1. We assume that the subject is in the center of the photographs, with the background visible on both sides (left and right). The subject's locations in the three photographs do not have to be identical.

The selfie scene is analyzed with multiview geometry [2]. Specifically, the camera, subject, and background locations in three-dimensional (3D) space are estimated using two-view geometry. The three images obtained are denoted as I_1, I_2, and I_3, respectively, where I_2 is the center of the three views. For the scene analysis, two-view images are used. The steps below are illustrated with images $I1$ and I_1 and I_2; however, images I_2 and I_3 can be used to achieve the same results.

Let $p_1=[u_1,v_1,1]^{\mathrm{T}}$ and $p_2=[u_2,v_2,1]^{\mathrm{T}}$ be the homogeneous two-dimensional (2D) coordinates of the matching points in images I_1 and I_2. Then, points p_1 and p_2 satisfy

$$p_1{}^{\mathrm{T}}\mathbf{F}p_2=0, \tag{1}$$

where $\mathbf{F}\in\mathbb{R}^{3\times3}$ denotes the fundamental matrix [2]. The fundamental matrix $\mathbf{F}$ can be determined from a set of matching points. The feature points are obtained using the scale-invariant features transform (SIFT) [16], and the RANSAC algorithm [17] is used to obtain matrix $\mathbf{F}$.

The essential matrix is computed from the fundamental matrix by

$$\mathbf{E}=\mathbf{K}\mathbf{F}\mathbf{K}^{\mathrm{T}}, \tag{2}$$

where $\mathbf{K}\in\mathbb{R}^{3\times3}$ is the matrix with intrinsic camera parameters such as pixel sizes and the center coordinate [2]. Matrix $\mathbf{K}$ is prepared through camera calibration. The essential matrix can be used to obtain the extrinsic camera parameters. The relative rotation $R\in\mathbb{R}^{3\times3}$ and translation $t\in\mathbb{R}^3$ between the two cameras are found using the procedures in Hartley and Zisserman [2]. The camera matrixes $\mathbf{P}_1$, $\mathbf{P}_2\in\mathbb{R}^{4\times3}$ of the two cameras used to photograph image I_1 and I_2 are written as

$$\mathbf{P}_1=\begin{bmatrix}R\\ t^{\mathrm{T}}\end{bmatrix}\mathbf{K},\ \mathbf{P}_2=\begin{bmatrix}I\\ 0^{\mathrm{T}}\end{bmatrix}\mathbf{K}, \tag{3}$$

where $\mathbf{I}\in\mathbb{R}^{3\times3}$ is the identity matrix. The camera for image I_2 is in the center, and the camera matrixes express the relative rotation and translation of the camera for image I_1. The camera matrix is a projective transform from

point $[x, y, z, 1]^T$ in a 3D space to point $[u, v, 1]^T$ in a 2D space by the camera. As the same point in the 3D space is projected to the matching points p_1 and p_2 in image I_1 and I_2 by the camera matrix $\mathbf{P}_1$ and $\mathbf{P}_2$, respectively, we have

$$p_1 = \mathbf{P}_1 \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \text{ and } p_2 = \mathbf{P}_2 \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \tag{4}$$

The locations of the points in the 3D space, $[x, y, z, 1]^T$, can be calculated from the corresponding pairs of matching points, $p_1 = [u_1, v_2, 1]$ and $p_2 = [u_2, v_2, 1]$. The matching points within the subject and background provide the subject and background's locations in 3D space relative to the cameras.

3.2 | Matting

Matting refers to the process of extracting a foreground object from an image. A given image I is modeled by

$$I = \alpha I_f + (1 - \alpha) I_b, \tag{5}$$

where I_f and I_b are the images of a foreground object and background, respectively, and the multiplications are element-wise. The alpha matte α denotes the foreground opacity, which is the linear combination's pixel-wise weights. We use matting to separate the foreground subject and background in the three selfies. The alpha matte can be obtained by

$$\min \alpha^T \mathbf{L} \alpha \quad \text{s.t.} \quad \mathbf{D}_s \alpha = \alpha_s, \tag{6}$$

where $\mathbf{L}$ is the matting Laplacian matrix, α_s is the alpha matte for the known foreground and background, and $\mathbf{D}_s$ is a diagonal matrix that shows the pixels with the known alpha matte [18]. The matting Laplacian matrix $\mathbf{L}$ can be supplemented with two additional rows and columns containing the estimated relative probabilities of pixels belonging to the foreground and background [19].

The known alpha matte α_s is usually provided from a scribble by a user. In our problem, we have more than one image available. We can generate α_s from the three images, which are photographed while the user is rotating. The images of the user in the center have less displacement than the background image, which has shifted due to the user's rotation. As a result, pixel displacements provide information about the subject and background locations in an image. We estimate displacement d between the two image, I_1 and I_2, using optical flow [20]. Subsequently, the scribble is obtained by

$$\alpha_s(u,v) = \begin{cases} 1, & \text{if } d(u,v) < \theta_1, \\ 0, & \text{if } d(u,v) > \theta_0, \\ 0.5, & \text{otherwise,} \end{cases} \tag{7}$$

with threshold $\theta_1 < \theta_0$. The color of pixels sampled in the regions

$$\begin{aligned} S_f &= \{I(u,v) | d(u,v) < \theta_1\}, \\ S_b &= \{I(u,v) | d(u,v) > \theta_0\}, \end{aligned} \tag{8}$$

are used to provide the relative probabilities used in the augmentation of the matting Laplacian.

The alpha mattes obtained for each image are used to separate the foreground and background images from the three images. The pixels in the foreground and background images in the 3D scene are located at the distances of the subject and background found in Section 3.1.

3.3 | View synthesis

A view from a virtual camera with a different lens at a new location can be synthesized from the scene analysis data. Our goal is to render an image with a longer focal length to simulate a portrait captured with a telephoto lens. Figure 2 shows an image captured by a camera with shorter and longer focal lengths with field of views θ_s and θ_l, respectively. If we change the lens with a wider field of view to a lens with a narrower field of view, the foreground subject becomes large. To maintain the same size of the foreground subject in a new image, the camera has to be moved away from the subject. Figure 3A shows how to position a virtual camera to provide a subject of the same size. Trigonometry is used to obtain the location of the virtual camera x as follows:

$$x = \frac{d_f \tan \theta_{s,f}}{\tan \theta_{l,f}}. \tag{9}$$

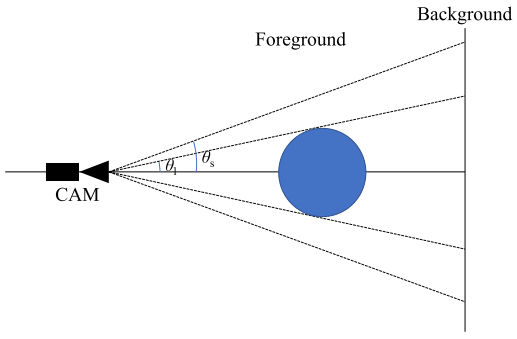

FIGURE 2 Sizes of the foreground subject for two lenses with different focal lengths

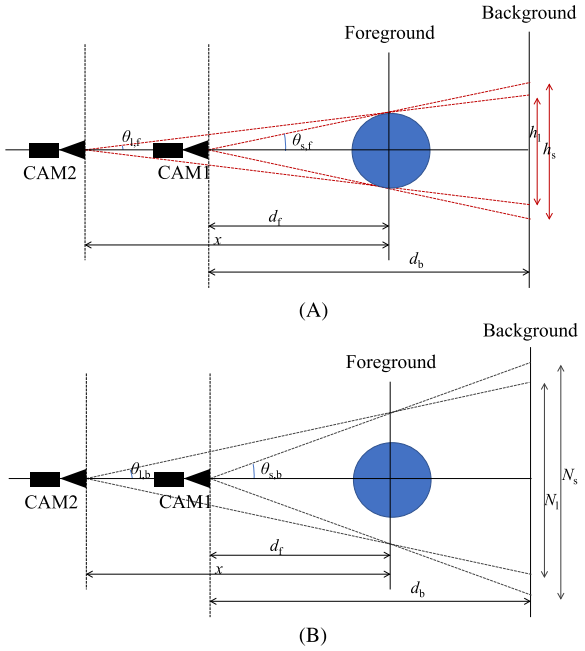

FIGURE 3 View captured by a virtual camera with a telephoto lens at a new location, considering the (A) foreground subject and (B) background

In Figure 3A, h_l and h_s are the regions obscured by the foreground subject when viewed with the cameras with longer and shorter lenses, respectively, at two different locations. Region h_s is larger than region h_l. When the scene is viewed by the virtual camera with a longer focal length lens at a new location, parts of the regions that were previously hidden by the foreground subject are revealed. When a scene is rendered using a virtual camera with a longer focal length lens, pixel data in the uncovered region are not available. Because the selfies are taken while the user is rotating, the uncovered region that is not visible in one image is visible in the others. By constructing a panoramic background from the backgrounds of the three selfies, we solve the problem of unavailable pixel data in the uncovered background.

Figure 3B shows how much of the background is included in the two views. N_l and N_s are the regions of background included in the scene of the cameras with longer and shorter focal length lenses at the two different locations, respectively. Region N_l is smaller than N_s. The smaller region fills the sensor with the same pixel counts as the larger region. As a result, the portion of the panoramic background image must be enlarged by the ratio N_s/N_l. In trigonometric terms, this is as follows:

$$N_s = 2d_b \tan \theta_{s,b}, \tag{10}$$

$$N_l = 2(x - d_f + d_b) \tan \theta_{l,b}, \tag{11}$$

where $\theta_{l,b}$ and $\theta_{s,b}$ denote the halves of the field of view of the lenses. As the size of foreground is maintained to be the same, we have

$$\frac{h_l}{N_l} = \frac{h_s}{N_s} = \frac{\tan \theta_{l,f}}{\tan \theta_{l,b}} = \frac{\tan \theta_{s,f}}{\tan \theta_{s,b}}. \tag{12}$$

Ratio N_s/N_l can be obtained from (12).

When the virtual camera's telephoto lens is focused on the foreground subject, out-of-focus blur blurs the background. The diameter of the blur kernel for the out-of-focus blur is given by

$$\mathrm{CoC} = \left| A \frac{F(d_s - d_b)}{d_b(d_s - F)} \right|, \tag{13}$$

where A denotes the aperture, F is the focal length, d_s is the distance to the subject in focus, and d_b is the distance to the blurry background [21,22]. The panoramic background is blurred by the circular blur kernel with the radius given in (13) to simulate the out-of-focus blur.

Because selfies are typically taken with a short focal length lens, the foreground image has barrel distortion. Images captured with a telephoto lens typically exhibit pincushion distortion. Because of the geometric distortion caused by perspective changes, facial images captured with different focal length lenses show different

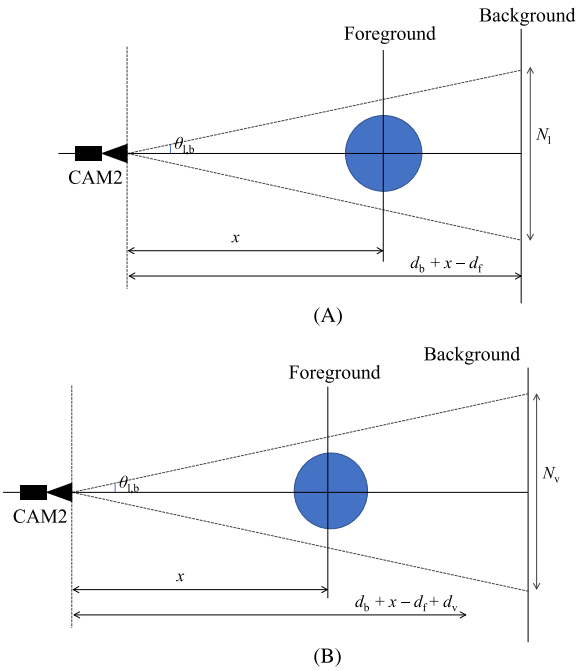

FIGURE 4 View captured by a virtual camera with a telephoto lens according to the distance of background: (A) original and (B) longer distances

facial features [23]. The foreground facial images can be processed by a perspective-aware manipulation method such as in Fried and others [24] for correction. We add pincushion distortion on purpose to simulate the pincushion of a telephoto lens [25].

The virtual view is synthesized by

$$I_v = \alpha_2 I_{2f} + (1-\alpha_2) I_{\mathrm{p}}, \tag{14}$$

where α_2 is the alpha matte for image I_2, I_{2f} is the pincushion introduced foreground of image I_2, and I_{p} is the blurred scaled-up panoramic background.

With the camera, subject, and background positions identified by the scene analysis, a new composition can be synthesized. For example, we can shift the camera and subject positions away from the background to increase the out-of-focus blur effect. Figure 4 illustrates the virtual composition with the camera and subject located away from the background. The new virtual composition includes a larger region of the background in its view. There is enough panoramic background in a new view of the virtual composition because the panoramic background is prepared from rotating selfies.

4 | PANORAMIC SELFIE SYNTHESIS FROM A SELFIE VIDEO

4.1 | Scene analysis

We consider a scenario wherein a user takes a video selfie at the same location while spinning around. The user scenario is depicted in Figure 5. In comparison with the previous section's first user scenario, the rotation covers a wider view angle, and the images include a panoramic view of the background behind the user. The video selfies are used in such a way that the video frames contain images with smaller geometric differences. If three images are used, as in the previous scenario, the structures that appear in the three images may differ significantly.

The user appears in the center of the image while taking a selfie, with the background appearing on the left and right sides of the user. We use the left and right backgrounds to construct two background images. Let $I(u, v, k)$ be the image of the kth frame of the acquired video. The left and right parts of the image are denoted as I_{L} and I_{R}. We obtain displacement $(\hat{d}_u, \hat{d}_v)$ between the adjacent frames by

$$\max_{d_u,d_v} \phi(I_{\mathrm{LB}}(u,v,k-1), I_{\mathrm{L}}(u+d_u, v+d_v, k)), \tag{15}$$

where $I_{\mathrm{LB}}(u, v, k-1)$ is the left-side background constructed using up to the $k-1$th frame, $I_{\mathrm{L}}(u, v, k)$ is the left-side background image of the kth frame, and function ϕ measures the similarity between the two images. The left-side background image is initialized with $I_{\mathrm{LB}}(u, v, 0) = I_{\mathrm{L}}(u, v, 1)$.

The background image is updated using the two images: background image $I_{\mathrm{LB}}(u, v, k-1)$ and displaced image $I_{\mathrm{L}}(u+\hat{d}_u, v+\hat{d}_v, k)$. For the pixels wherein the two images overlap, we use a smooth transition as follows:

$$\begin{aligned} I_{\mathrm{LB}}(u,v,k) &= (1-w(v)) I_{\mathrm{LB}}(u, v, k-1) \\ &\quad + w(v) I_{\mathrm{L}}(u+\hat{d}_u, v+\hat{d}_v, k), \end{aligned} \tag{16}$$

where $w(v)$ increases linearly from zero to one in the overlapped region. The left background image after all the frames in the video is denoted as $I_{\mathrm{LB}}(u, v)$. The right-

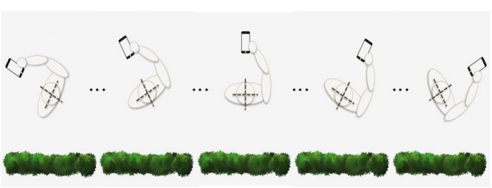

FIGURE 5 User scenario with a video selfie. Video is captured at the same location while the user is spinning around

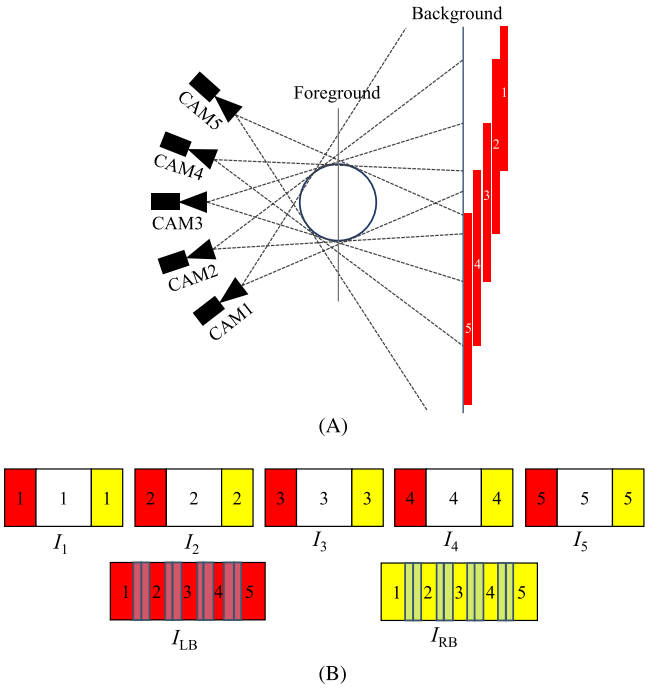

FIGURE 6 Construction of the background image from the frames of a video: (A) frames of images acquired in a video and (B) panoramic background images from the left and right sides of the images

side background image $I_{RB}(u, v)$ is constructed using the same method.

Figure 6 illustrates the construction of the left and right background images. For example, five video frames are acquired as shown in Figure 6A. The images on the left and right sides of the subject are depicted in the red and yellow regions in Figure 6B, respectively. The left-side background image $I_{LB}(u, v)$ is stitched from the displaced red regions, and the right-side background image $I_{RB}(u, v)$ is stitched from the displaced yellow regions.

4.2 | View synthesis

A virtual panoramic selfie is created from a selfie video in which the location of a subject is chosen post hoc. We use three images—the constructed left and right background images ($I_{LB}(u, v)$ and $I_{RB}(u, v)$, respectively) and the image that contains the subject at a particular location $I(u, v, c)$. Displacement (d_u^L, d_v^L) between the left background and center images is obtained by

$$\max_{d_u,d_v} \phi(I(u, v, c), I_{LB}(u+d_u, v+d_v)) \quad (17)$$

and displacement (d_u^R, d_v^R) between the right background and center images are obtained by

$$\max_{d_u,d_v} \phi(I(u, v, c), I_{RB}(u+d_u, v+d_v)). \quad (18)$$

Subsequently, the panoramic selfie is constructed using the center image $I(u, v, c)$ and the displaced images $I_{LB}(u+\hat{d}_u^L, v+\hat{d}_v^L)$ and $I_{RB}(u+\hat{d}_u^R, v+\hat{d}_v^R)$. For the overlapped regions, linear blending similar to (16) is used.

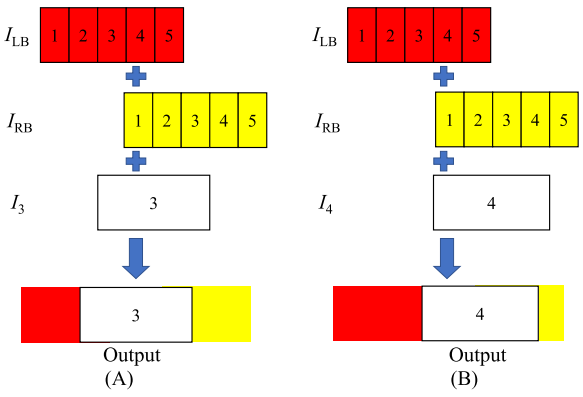

FIGURE 7 The panoramic selfie is a synthesis in which the subject can be placed anywhere in the background post hoc. (A) when using image #3 as the center image and (B) when using image #4 as the center image

Figure 7 illustrates the synthesis of the panoramic selfie. The frame numbers follow the numbering in Figure 6. In Figure 7A, the third image, $I(u, v, 3)$, is

FIGURE 8 Examples of the three selfies photographed under the user scenario. Images of the (A) first, (B) second, and (C) third shots are shown

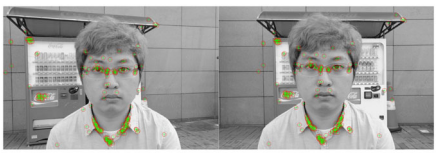

FIGURE 9 Matching points found in two of the three images: (A) I_1 and (B) I_2

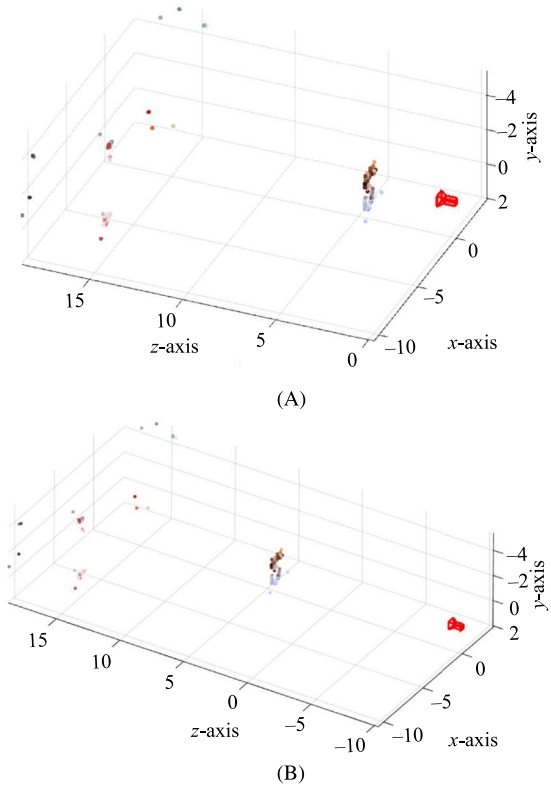

FIGURE 10 Result of scene analysis. (A) In 3D space, the camera positions and matching points from the subject and background are shown. (B) Moving the camera to a different location to render a view with a longer focal length lens

chosen as the center image; in Figure 7B, the fourth image, $I(u, v, 4)$ is chosen as the center image. After the selfie video is captured, the user can select the center image. Following that, the virtual selfie with the user's image at the selected location is synthesized.

5 | EXPERIMENTS AND DISCUSSIONS

5.1 | Portrait from multiple selfies

5.1.1 | Scene analysis

The examples of the three selfies are shown in Figure 8. Each step of the proposed algorithm is shown with the three photographs depicted in Figure 8. Figure 9 shows the matching points observed using the SIFT features in images I_1 and I_2. The matching points are used to obtain the fundamental matrix **F**, essential matrix **E**, and camera matrixes $\mathbf{P}_1$ and $\mathbf{P}_2$.

The locations of the matching points in the 3D space are obtained using the $\mathbf{P}_1$ and $\mathbf{P}_2$ matrixes. The locations are displayed in Figure 10A. The camera location for image $I2$ is shown in red, as is the origin of the 3D space. Figure 10A shows how the matching points from the subject and background are separated in 3D space. Our aim is to move the red camera to another location and use a telephoto lens to synthesize a view for a portrait, as depicted in Figure 10B.

5.1.2 | Matting

Figure 11 shows the result of the optical flow estimated between I_1 and I_2 using Middlebury color coding and the vector field. The estimated distances for the foreground subject are less than those for the background. For the matting algorithm, the distance is thresholded to form a scribble and color sampling.

The result of the matting is shown in Figure 12. The input image I_2 is shown in Figure 12. Scribble α_s built from the distance shown in Figure 11 is presented in Figure 12B. The foreground and background are extracted using the alpha matte, as shown in Figure 12C,D, respectively.

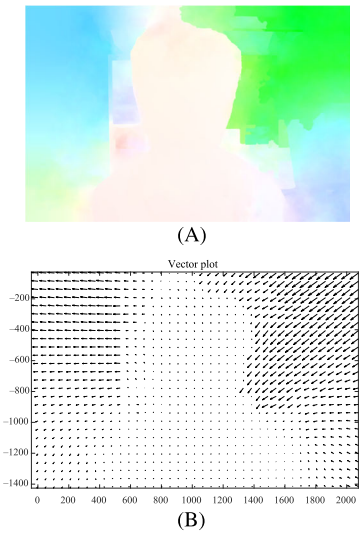

FIGURE 11 Optical flow results, shown using Middlebury (A) color coding and (B) vector field

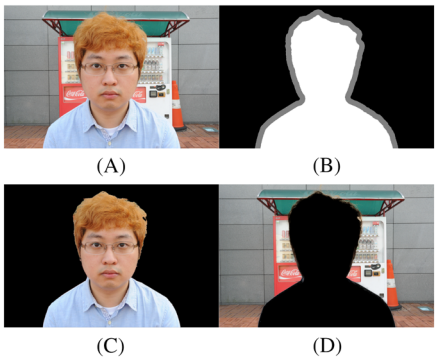

FIGURE 12 Result of matting: (A) image I_2, (B) trimap, (C) foreground, and (D) background

FIGURE 13 The panorama background was created by combining the backgrounds of the three images

FIGURE 14 Example of pincushion distortion simulation. (A) The foreground image was captured using a short focal length lens with barrel distortion. (B) Pincushion distortion was added to the foreground image to simulate the effect of a telephoto lens

5.1.3 | View synthesis

Figure 13 is the background constructed from the background of images I_1, I_2, and I_3. The panoramic background stitched up the background region behind the subject, black region in Figure 12D. The unfilled region, such as the lower parts of the vending machine, will be obscured by the foreground subject in the final view.

Figure 14 shows the simulation of pincushion distortion by a telephoto lens. The foreground image in Figure 14A is obtained from the three selfies captured with a short focal length lens. The illustration depicts barrel distortion. The image in Figure 14B results from the added pincushion distortion.

The final views of cameras with telephoto lenses of the 85-mm, 105-mm, 200-mm, and 300-mm focal lengths from farther distances are shown in Figure 15. The final results show that our proposed method renders portraits with the correct depth of field and field of view when using telephoto lenses. As the focal length of the simulated lens increases, so does the out-of-focus blur of the background. As the focal length of the simulated lens increases, so does the background inside the field of view. The background is shown in the simulated images of the lenses with a longer focal length, which is behind the subject in the original image from the 28-mm lens (the first images from the left in the figure). Because our proposed method renders a view from the rotating selfies, the regions in one of the original images that are occluded by the subject can be rendered accurately. Region filling techniques [26–29] to hide newly exposed occluded areas were not necessary. Through simulated lenses, our proposed method provides an accurate view of the scene. A view from any focal length lens can be synthesized using the proposed method. The figure includes, for example, the views of the 85- and 300-mm focal length lenses, which we do not currently own.

Table 1 shows the distance between the subject and virtual camera for the images in Figure 15. The distance for the 28-mm focal length lens is the actual distance between the subject and camera. When using the longer focal length lenses, the virtual camera is moved back and away from the subject. The distances become more than 6 m for the longest 300-mm focal length lens. This is the actual distance that photographers would place between them and a subject when taking a head and shoulder portrait of the subject. It is impossible for users to extend the distance between them and the camera when taking a selfie even with a selfie stick. By taking multiple selfies

TABLE 1 Distance between subject and virtual camera

Focal length	Distance
28 mm	0.60 m
85 mm	1.84 m
105 mm	2.24 m
200 mm	4.28 m
300 mm	6.74 m

FIGURE 15 Examples of virtual view synthesis of the telephoto lenses, for the pairs of images. Top: images photographed using the specified lenses, bottom: images synthesized from the three rotating selfies

and synthesizing a virtual view, we were able to render a portrait with a long focal length lens from a camera at a far-away distance.

Figure 16 shows the comparison to digital zoom [10]. A photograph is taken at a greater distance with a 28-mm focal length lens, and a portion of the image corresponding to the field of view of a 200-mm focal length lens is cropped and enlarged. The loss of resolution is apparent in digital zoom images. Furthermore, the image shows the depth of field of a 28-mm lens, and there is no out-of-focus blur in the background. In comparison, our proposed method renders the foreground subject with the same resolution as the original selfie, while the background has out-of-focus blur due to the shallow depth of field of a 200-mm lens.

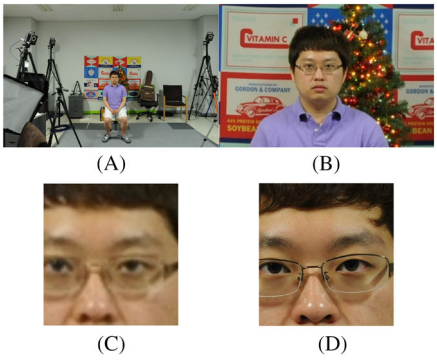

FIGURE 16 Comparison to digital zoom. (A) Image taken with a 28-mm lens in the same location as the image taken with a 200-mm lens; (B) digital zoom to simulate a 200-mm lens; (C) part of image obtained by digital zoom; and (D) part of image obtained by our proposed method

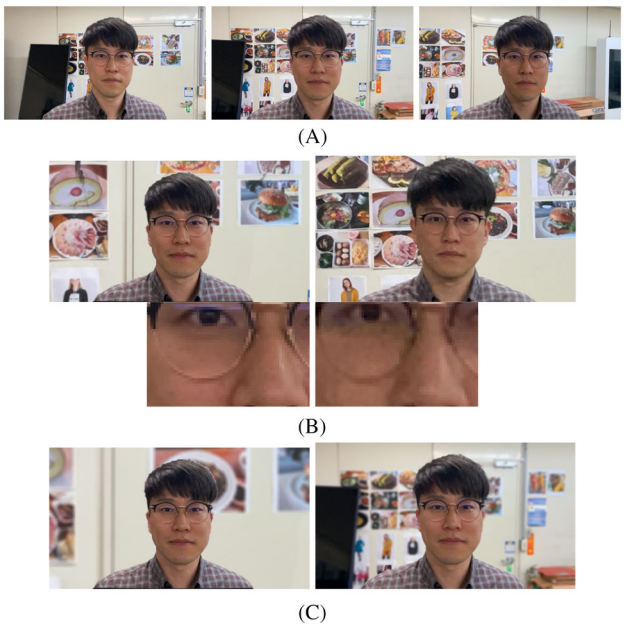

FIGURE 17 Comparison to currently available methods: (A) three selfie images taken with a 28-mm lens; (B) left: 85-mm f5.6 image synthesis from selfies; right: 85-mm image captured with Viewfinder Preview app; and (C) left: 300-mm f2.8 image synthesis from selfies; right: 300-mm image simulated by Pitu app

Figure 17 shows comparisons to currently available apps: Viewfinder Preview [14] and Pitu [15]. Figure 17A is the three selfie images taken while rotating using Viewfinder Preview app with a 28-mm lens. In Figure 17B, an 85-mm image synthesized from the three selfies is shown on the left, and an image captured using an 85-mm lens in the Viewfinder Preview app is shown on the right. While using the 85-mm lens, the Viewfinder Preview app requires the photographer to move away from the subject. As a result, the image on the right is a portrait taken by a separate photographer a long distance away from the subject, rather than a selfie. The zoomed images show that the proposed method provided more resolution than Viewfinder Preview software. In Figure 17C, a 300-mm image synthesized from the three selfies is shown on the left. The Viewfinder Preview app's lens focal length is limited to 180 mm, and a 300-mm image could not be captured. To simulate a 300-mm image, we used the Pitu app. Pitu app simply blurs the background of a 28-mm image, and the perspective between the camera, subject, and background differs from that of a 300-mm image. There is currently no app that allows a user to create a portrait of themselves from selfies taken with various focal length lenses.

Figure 18 shows an example of virtual composition. We increase the distance between the camera and the background while keeping the distance between the camera and the subject constant. The simulated images show that as the distance between the subject and the background increases, so does the out-of-focus blur of the background. The background included in the field of view increases as we virtually move the subject and camera away from the background. Because more background information is available from the rotating selfies, our proposed method added more background than the original image (the first images from the left in the figure). Through virtual composition, we propose a proper view of the scene.

5.2 | Panoramic selfie from a selfie video

5.2.1 | Scene analysis

Figure 19 displays images from a video captured while the user is rotating. The subject (user) is in the center of

FIGURE 18 Examples of virtual composition at different distances between the camera and the background and the images captured at different distances between the camera and background are shown

FIGURE 19 Examples of frames in a selfie video captured while a user is rotating; four frames are shown as an example

the images, and the left and right sides contain different parts of the scene behind the subject.

The panoramic background images are constructed by stitching the parts of the images on the left side of the subject and those on the right side of the subject. The two background images, $I_{\mathrm{LB}}(u,v)$ and $I_{\mathrm{RB}}(u,v)$, constructed using the video are shown in Figure 20. There are luminance mismatches in the background images due to exposure differences between video frames. Because of the use of three different center images, the stitching locations between the center and background images differ. The difference in lightness disappears as the center images move to the right sides, indicating different stitching locations. The difference in exposure can be compensated for by adjusting the color and luminance between the frames [30].

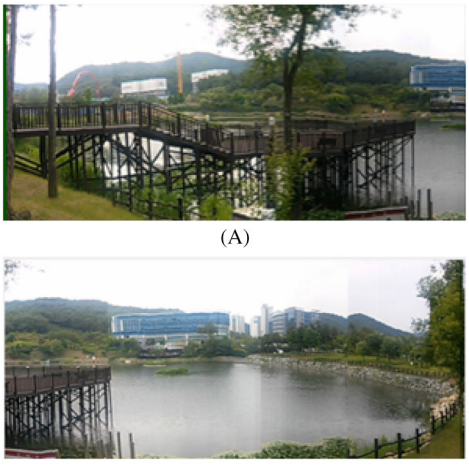

FIGURE 20 Left and right panoramic background images: (A) background constructed from parts of the images on the left side of a subject; (B) background constructed from parts of the images on the right side of a subject

FIGURE 21 Virtual panoramic selfie, where the subject can be placed in any locations, and three examples with three different subject locations are shown

5.2.2 | View synthesis

With any frame of the video as the center image $I(u,v,c)$, a virtual panoramic selfie can be synthesized. Because the frames in the video are captured from different parts of the scene, using a different frame as the center image moves the subject in the panoramic selfie to a different location. After the video has been captured, the location can be selected. Figure 21 shows examples of the synthesized panoramic selfies, where three different parts of the scene are selected as the subject locations. The proposed method assumes that a user is taking a panorama while pointing the camera at himself/herself. There is currently no app that assumes this scenario.

6 | CONCLUSION

In this study, a scene is constructed from multiple selfies taken at the same location while the user rotates. Following that, virtual selfies are obtained by moving the camera closer to the subject and changing the focal length of the lens. Furthermore, the change in perspective and lens characteristics is simulated. We were able to obtain portraits with the same composition as the portraits with the selected lenses at the new locations in our experiments. We were also able to obtain a virtual panorama, in which the subject's image is placed in a preferred location. Our proposed algorithms can provide new use cases, wherein the composition of selfies is not limited by the focal length of a camera or the distance from a camera. We are currently investigating how, with the help of computations, using multiple images of a scene can open up new and exciting use cases.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

ORCID

Yongsik Lee https://orcid.org/0000-0003-2461-2266
Jinhyuk Jang https://orcid.org/0000-0002-5082-7108
Seungjoon Yang https://orcid.org/0000-0001-9109-1582

REFERENCES

1. A. Watt, *3D computer graphics*, CUMINCAD, 1993.
2. R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2003.
3. M. Tanimoto, *FTV (free-viewpoint television)*, APSIPA Trans. Signal Inform. Process. **1** (2012), E4.
4. J. Starck and A. Hilton, *Model-based multiple view reconstruction of people*, (IEEE International Conference on Computer Vision, Nice, France), 2003, pp. 915–922.
5. H.-Y. Lin and Y.-H. Xiao, *Free-viewpoint image synthesis based on non-uniformly resampled 3D representation*, (IEEE International Conference on Image Processing, HongKong, China), 2010, pp. 2745–2748.
6. J. Starck and A. Hilton, *Virtual view synthesis of people from multiple view video sequences*, Graph. Models **67** (2005), no. 6, 600–620.
7. Z. Liu, P. An, S. Liu, and Z. Zhang, *Arbitrary view generation based on DIBR*, (International Symposium on Intelligent Signal Processing and Communication Systems, Xiamen, China), 2007, pp. 168–171.
8. C.-M. Cheng, S.-J. Lin, S.-H. Lai, and J.-C. Yang, *Improved novel view synthesis from depth image with large baseline*, (19th International Conference on Pattern Recognition, Tampa, FL, USA), 2008, pp. 1–4.
9. X. Chen, H. Liang, H. Xu, S. Ren, H. Cai, and Y. Wang, *Virtual view synthesis based on asymmetric bidirectional DIBR for 3D video and free viewpoint video*, Appl. Sci. **10** (2020), no. 5, 1562.
10. S. Watanabe, *Digital camera with electronic zooming function*, 2003. US Patent 2003/0007082 A1.
11. T. Chen, *NeLF: Neural light-transport field for portrait view synthesis and relighting*, arXive preprint, 2021. https://doi.org/10.48550/arXiv.2107.12351
12. J. Freer, K. M. Yi, W. Jiang, J. Choi, and H. J. Chang, *Novel-view synthesis of human tourist photos*, (Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA), 2022, pp. 3069–3076.
13. T. Leimkühler and G. Drettakis, *FreeStyleGAN: Free-view editable portrait rendering with the camera manifold-supplemental materials*, arXive preprint, 2021. https://doi.org/10.48550/arXiv.2109.09378
14. A. Fowler, *Viewfinder Preview*. https://apps.apple.com/kr/app/viewfinder-preview/id1216484605
15. Ltd Tencent Technology Co., *Pitu—Best selfie and PS Soft*. https://apps.apple.com/us/app/pitu-best-selfie-and-ps-soft/id724295527
16. D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, Int. J. Comput. Vision **60** (2004), no. 2, 91–110.
17. M. A. Fischler and R. C. Bolles, *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography*, Commun. ACM **24** (1981), no. 6, 381–395.
18. A. Levin, D. Lischinski, and Y. Weiss, *A closed-form solution to natural image matting*, IEEE Trans. Pattern Anal. Machine Intell. **30** (2007), no. 2, 228–242.
19. J. Wang and M. F. Cohen, *Optimized color sampling for robust matting*, (IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA), 2007, pp. 1–8.
20. D. Sun, S. Roth, and M. J. Black, *Secrets of optical flow estimation and their principles*, (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA), 2010, pp. 2432–2439.
21. J. Demers, *Depth of field: A survey of techniques*, GPU Gems **1** (2004), no. 375, U390.
22. J. Wu, C. Zheng, X. Hu, Y. Wang, and L. Zhang, *Realistic rendering of bokeh effect based on optical aberrations*, Visual Comput. **26** (2010), no. 6–8, 555–563.
23. P. Perona, *A new perspective on portraiture*, J. Vision **7** (2007), no. 9, 992–992.
24. O. Fried, E. Shechtman, D. B. Goldman, and A. Finkelstein, *Perspective-aware manipulation of portrait photos*, ACM Trans. Graph. **35** (2016), no. 4, 1–10.

25. G. Vass and T. Perlaki, *Applying and removing lens distortion in post production*, (Proceedings of the 2nd Hungarian Conference on Computer Graphics and Geometry), 2003, pp. 9–16.
26. I. Daribo and B. Pesquet-Popescu, *Depth-aided image inpainting for novel view synthesis*, (IEEE International Workshop on Multimedia Signal Processing, Saint_Malo, France), 2010, pp. 167–170.
27. C. Guillemot and O. Le Meur, *Image inpainting: Overview and recent advances*, IEEE Signal Process. Mag. **31** (2013), no. 1, 127–144.
28. M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, *Image inpainting*, (Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques), 2000, pp. 417–424. https://doi.org/10.1145/344779.344972
29. A. Criminisi, P. Pérez, and K. Toyama, *Region filling and object removal by exemplar-based image inpainting*, IEEE Trans. Image Process. **13** (2004), no. 9, 1200–1212.
30. Y. Xiong and K. Pulli, *Fast panorama stitching for high-quality panoramic images on mobile phones*, IEEE Trans. Consumer Electron. **56** (2010), no. 2, 298–306.

AUTHOR BIOGRAPHIES

Yongsik Lee received the B.S. degree from Pusan National University, Pusan, Republic of Korea, in 2010, and Ph.D. degree in Electrical and Computer Engineering from the Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea, in 2018. He is currently with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea. His research interests are image processing and computer vision

Jinhyuk Jang received the B.S. and M.S. degrees from the School of Electrical and Computer Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea, in 2014 and 2016, respectively. He is currently with the Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea. His current research interests include image processing, blur estimation, human facial recognition, and human action recognition.

Seungjoon Yang (S'99-M'00) received the B.S. degree from Seoul National University, Seoul, Republic of Korea, in 1990, and the M.S. and Ph.D. degrees from the University of Wisconsin-Madison in 1993 and 2000, respectively, all in Electrical Engineering. From 2000 to 2008, he worked at Samsung Electronics Co., Ltd.'s Digital Media Research and Development Center. He is currently employed at Ulsan National Institute of Science and Technology's School of Electrical and Computer Engineering in Ulsan, Republic of Korea. Image processing, estimation theory, and optimization theory are among his research interests.

How to cite this article: Y. Lee, J. Jang, and S. Yang, *Virtual portraits from rotating selfies*, ETRI Journal **45** (2023), 291–303. https://doi.org/10.4218/etrij.2021-0429