

Received December 16, 2020, accepted December 27, 2020, date of publication January 1, 2021, date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048759

# Quantitative Assessment Method of Image Stitching Performance Based on Estimation of Planar Parallax

KYUNGHWA JUNG<sup>1</sup> AND JAESUNG HONG<sup>1</sup>, (Member, IEEE)

Department of Robotics Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 42988, Republic of Korea

Corresponding author: Jaesung Hong (jhong@dgist.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (Ministry of Science and ICT) under Grant 2020R1A2C2100012.

**ABSTRACT** While parallax-tolerant image stitching is a relatively mature field, the performances of image stitching methods have been assessed subjectively and qualitatively. These methods primarily provide the stitched image itself to demonstrate the performance, rather than quantitative data. Although several objective assessment methods have been proposed for quantifying the quality of stitched images, only the stitched output images have been analyzed, without considering the parallax level in each input image. We propose a method for quantifying the parallax level of the input images and clustering them accordingly. This facilitates a quantitative assessment of the various stitching methods for each parallax level. The parallax levels of the images are grouped based on the magnitude and variation in the planar parallax, as estimated with the proposed metric using matching errors and patch similarity. The existing image stitching methods are compared experimentally in terms of the residual misalignment errors, based on 73 pairs of different levels of parallax images originally classified in this study. Among the existing methods, the elastic local alignment method exhibits the least error. The shape-preserving half-projective method produces a larger misalignment error, but creates a natural panorama with less geometric distortion. We introduce a quantitative assessment method for considering the parallax of input images in image stitching methods. It can aid in specifying their performances, and in finding an appropriate method depending on the parallax level of the input images.

**INDEX TERMS** Image alignment, image stitching, planar parallax, plane + parallax, quantitative assessment.

## I. INTRODUCTION

Image stitching has been extensively studied, and commercial software is available for various camera systems. One of the major challenges in image stitching is in correcting the parallax error [1], [2]. A parallax error occurs when the non-parallax point of a camera is moved while capturing a three-dimensional (3D) scene, resulting in image misalignments, such as a ghost artifact. Therefore, many parallax-tolerant image stitching methods have been developed, aiming to reduce image misalignments [3]–[5].

However, the performance of the parallax-tolerant image stitching methods is generally assessed subjectively and qualitatively by presenting an input parallax image and its resulting panoramic image [6]–[9]. An observer subjectively evaluates the level of parallax in the input image, and the

quality of the resulting panoramic image. This subjective and qualitative assessment has limitations in regards to specifying the performance of the image stitching methods, and individual evaluations can differ for an identical method. These limitations make it difficult to select an appropriate image stitching method for a given input image.

Several studies have attempted to quantify the quality of output panoramic images for an objective evaluation and comparison of the performance [5], [10]–[19]. However, they only focused on the resulting panoramic image, without considering whether the parallax of the input image was objectively large or small. When the parallax level of the input image differs, evaluating the performance with only the resulting image may result in inaccurate conclusions. In other words, it is difficult to accurately determine the extent to which the parallax in input images has been addressed in the panoramic image results. Until now, the parallax levels of input images have been expressed as small, moderate, large,

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar<sup>1</sup>.

**TABLE 1. Inconsistent expressions of parallax level for an identical image in previous studies.**

Image	Railtracks [3]	Vila [9]	Temple [21]
Parallax level expression	Small, moderate [7] Moderate [2] Large [41]	Small [42] Small, moderate [7] Moderate [2] Large [43]	Moderate [2] Large [41] Challenging [42]

and very large, based on subjective individual assessments. Therefore, some studies have assigned different levels of parallax to the same image (Table 1).

In this study, we propose a quantitative assessment method that provides an objective analysis for determining the parallax of an input image. The parallax in the input images was quantified with the proposed metric based on matching errors and patch similarity, and the images were divided into groups based on the quantified parallax. The alignment performances of the image stitching methods were quantitatively defined, particularly with respect to differences in the parallax level. In addition, we found that the open image dataset [3], [4], [7], [8], [20], [21] used in previous image stitching studies was biased toward small-parallax images; very few open images had moderate or large parallaxes. We created a supplementary image dataset with moderate and large parallaxes, and provided a full dataset with evenly distributed images at each parallax level. The full image dataset with the grouped parallax levels is publicly available. Our quantitative assessment can aid in searching for or developing an appropriate stitching method, depending on the parallax of the input images. The proposed assessment scheme can be used to compare various methods, including novel methods in the future. The proposed method can also be applied in various computer vision studies involving the parallax levels of images.

## II. RELATED WORKS

### A. PLANAR PARALLAX IN COMPUTER VISION

In a physics dictionary [22], parallax is defined as “the apparent displacement of a distant object with respect to a more distant background when viewed from two different positions.” In the field of computer vision, a plane + parallax method has been developed and used for 3D scene analysis [23], [24]. The plane + parallax method divides the 2D image motion between imaged 3D scene points caused by camera motion into the motion of a reference plane and a residual motion, called the “planar parallax.” The image motion of the reference plane is described by homography. The homography contains a rotation of the camera and a calibration parameter, whereas the planar parallax is related to the translation of the camera and depth of the 3D scene point. Therefore, the plane + parallax method has been employed to simplify geometric interpretations of scene structures, as aligning two images with the motion of the reference plane eliminates the effects from camera rotation, calibration, and zooming [25].

The plane + parallax division has been adopted to address various challenges in computer vision. In [26], a synthesized image was created between scene images using a

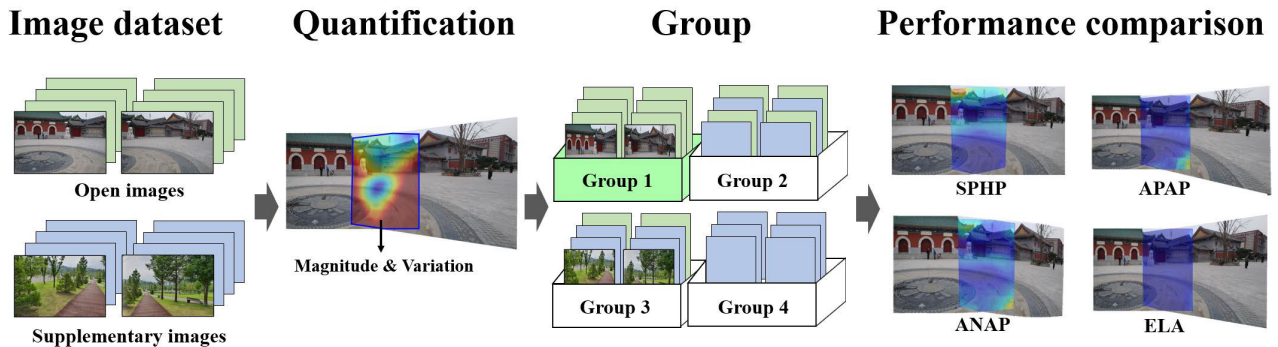
parallax constraint. In addition, planar parallax has been used to distinguish an independently moving object from the motion of a scene owing to camera motion [25], or to restore 3D scene structures relative to a reference plane [23], [27]. In this study, the planar parallax was used to quantify and group the parallax levels of the input images used for image stitching.

### B. PARALLAX-TOLERANT IMAGE STITCHING

A global homography transformation can be applied to stitch images of planar scenes or camera rotations around a non-parallax point, because no parallax exists between the images. However, if the non-parallax point of the camera is moved while capturing 3D scenes, a parallax will occur between the input images; this will hinder image alignment based on using only global homography.

Many image stitching methods have been developed to address parallax challenges by using local adaptive transformations. Gao *et al.* [21] used dual homography warp (DHW) to stitch images with both distant and near planes. Lin *et al.* [9] and Liu *et al.* [28] applied a smoothly varying local affine (SVA) and homography transformations to general scenes. However, these methods were deficient for parallax images with wide baselines.

The as-projective-as-possible (APAP) method [3] exhibits a high-precision local alignment in an overlap area. However, an unnatural projective distortion can appear in the non-overlap area for large-parallax images. The shape-preserving half-projective (SPHP) method [4] yields a natural panorama with less distortion in non-overlap regions, based on a combination of homography and similarity warps. Although the SPHP results appear to be considerably natural, the panoramas occasionally fail in parallax images with large viewpoint changes. Lin *et al.* developed the adaptive as-natural-as-possible (ANAP) warp [20] for natural panoramas, such as those from the SPHP method. The ANAP warp uses linearized homography and a point set with the smallest rotation angle for an optimal similarity transformation. However, unnatural rotation and scaling can still occur while stitching multiple images. Li *et al.* [5] used robust elastic local alignment (ELA) model based on a thin plate spline with a Bayesian model to remove outliers among feature matches. The robust elastic warping simultaneously achieved an accurate alignment and efficient processing. Overall, the existing methods have exhibited good performance in parallax conditions. However, their parallax-tolerant performances have not been specified and compared according to the parallax level(s).



**FIGURE 1.** Evaluation process using the grouped parallax images for performance verification of the image stitching.

### C. QUALITY ASSESSMENT OF STITCHED IMAGES

There has not been sufficient study regarding how to assess the quality of stitched images resulting from different stitching methods, while several image stitching methods have been developed and commercialized. An objective assessment is required to validate and compare the performances of the various stitching methods.

The quality-assessment methods for stitched images can be classified into subjective and objective approaches. A subjective approach depends on the human visual system [6]–[9]. Many studies have compared the resultant panoramic images to demonstrate the usefulness of their algorithms relative to other algorithms. Although demonstrating the performances of stitching algorithms based on the resultant panoramic images is highly intuitive, the assessment can vary from person to person.

To provide an objective evaluation of panoramic images, a matching error metric has been used to measure the misalignment in the overlap area [10]–[12], [29]. The matching error metric determines a corresponding point between two input images, and averages the Euclidean distance of the corresponding points. The matching error metric can directly calculate the error of the misaligned corresponding points, but only indicates the misalignments in the regions in which the corresponding points exist, i.e., not the entire overlap area. A patch similarity metric has also been used to quantify misalignments, even for areas where points are not extracted. This metric is an indirect measurement of the misalignment error, e.g., by using structural similarity (SSIM) [5], [13], [14], normalized cross-correlation [10], [11], or the peak signal-to-noise ratio [15]. The concept of the patch similarity metric is that the larger the misalignment, the less the similarity between the patches in the overlap area. However, the disadvantage of this metric is that the similarity value may be high despite the large parallax when the overlap region includes a large featureless area.

Recently, objective approaches based on mathematical models have been introduced [16]–[19]. Ling *et al.* [18] developed a convolutional sparse coding-based metric, and trained kernels to assess stitching specific distortions and their compound effects. Madhusudana and Soundararajan *et al.* [19] devised a stitched image quality

evaluator (SIQE) for capturing the ghosting and structural changes of edges, based on a Gaussian mixture model and steerable pyramid decompositions. The SIQE provided quality scores that could be validated using the consistency of human assessments. However, the reliability of these mathematical models has not yet been verified.

In this study, a metric combining the matching error metric and patch similarity metric was used to measure the planar parallax and panorama error, misalignment after the homography transformation, and misalignment after the final panorama transformation. We applied an optical flow to the proposed metric to calculate the misalignment vectors for full pixels in the overlap area, and used an SSIM-based patch similarity to filter the outliers in the misalignment vectors.

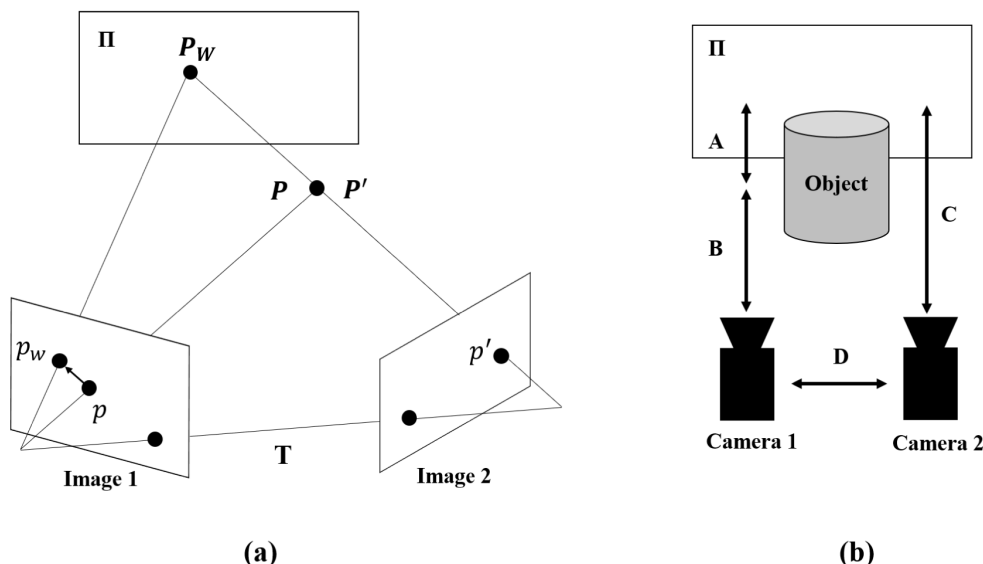
### III. METHODS

The scheme for the proposed assessment considering the parallax level of the input image is depicted in Figure 1. A planar parallax was used to define the parallax between the images. A metric was developed to measure the planar parallax of an open image dataset frequently cited in previous studies. From the results of the parallax quantification using the proposed metric, we observed that most of the open images had small parallaxes, and that a few open images had moderate or large parallaxes. We included a supplementary image dataset with moderate or large parallaxes to compare the various stitching methods. As a result, the parallax images of the full dataset were evenly distributed with each parallax level. The full dataset images were grouped into similar parallax images using K-means clustering, based on the parallax magnitude and variation. Finally, the alignment performance in the panoramic results for the grouped parallax images was verified and compared using the proposed metric.

#### A. DEFINITION OF PLANAR PARALLAX

The parallax between images is represented by the planar parallax of the plane + parallax method. The 2D image motion caused by camera motion can be decomposed into the motions of a reference plane and the residual planar parallax [25], [30]–[32].

The geometry of the planar parallax between the two images is depicted in Figure 2-a. The Cartesian coordinates of a scene point with respect to the two cameras are represented



**FIGURE 2.** Planar parallax. (a) Geometry of planar parallax, (b) four components to affect planar parallax level (A: the perpendicular distance from the object point to the reference plane, B: the depth of the object point with respect to the first camera, C: the perpendicular distance from the second camera to the reference plane, D: translational distance of the camera.)

by  $P = (X, Y, Z)^T$  and  $P' = (X', Y', Z')^T$ , and the homogeneous image coordinates of the scene point  $P$  are  $p = (x, y)$  and  $p' = (x', y')$ , respectively. The rotation and translation between the two cameras are expressed as a  $3 \times 3$  matrix ( $R$ ) and  $3 \times 1$  vector ( $T$ ), respectively.  $K_1$  and  $K_2$  represent  $3 \times 3$  matrices with the internal calibration parameters of cameras 1 and 2, respectively.  $\Pi$  denotes an arbitrary reference plane, and  $H$  denotes the homography that aligns the two images based on the reference plane  $\Pi$ .

$u$  denotes the measurable 2D image displacement vector between the corresponding points  $p$  and  $p'$  in each image. The plane + parallax method defines  $u = u_{\Pi} + \mu$ , where  $u_{\Pi}$  is the motion of the reference plane (the homography owing to  $\Pi$ ), and  $\mu$  is the motion of the residual planar parallax. If  $p_w$  is a projected point of a scene point  $P_w$  on the reference plane  $\Pi$ , its corresponding point  $p'$  will be aligned with  $p_w$  when image 2 is warped by the homography. However, because  $p$  is a projected point of a non-planar point  $P$ , its corresponding point after warping is given by  $p_w$ . Subsequently, the displacement vector between  $p_w$  and  $p$  in image 1 is the planar parallax motion  $\mu$ . Therefore, the residual image motion between image 1 and the warped image of image 2 is the planar-parallax displacement. According to the literature [24], [25], [27], [30], [33], the planar parallax is affected by the four components A, B, C, and D (Figure 2-b). Components A and D are proportional to the planar parallax, and B and C are inversely proportional.

**B. PLANAR PARALLAX ESTIMATION**

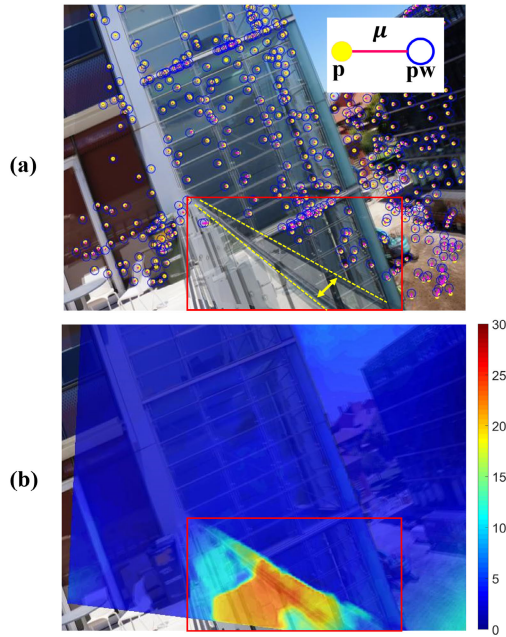
In the estimation of the planar parallax motion between the two input images, an image alignment was performed to detect the planar motion (Algorithm 1). A homography compatible with the planar motion was estimated using the corresponding extracted points, and was matched using a scale-invariant feature transform [34] from the two

**Algorithm 1** Planar parallax estimation

```

1: Input: Matched points  $p, p'$  and images  $I1, I2$ 
2: Output: Magnitude and variation of planar parallax
3: Procedure
4: for  $j = 1, \dots, 30$ 
5:   for  $i = 1, \dots, 500$ 
6:     Random 4 pairs  $(p, p')$  and solve for  $h$ 
7:      $\varepsilon_i = |h_i(p) - p'|$ 
8:     if  $\varepsilon_i < \gamma$  then
9:       Save  $h_i$  and  $S_i$ 
10:    end if
11:   end for
12:   Select initial  $h_j \leftarrow h_i$  with max  $S_i$ 
13:    $\text{argmin}_{K_{1j}, K_{2j}, R_j} |K_{2j}R_jK_{1j}^{-1}p - p'|$  with LM
14:   Optimal  $h_j = K_{2j}R_jK_{1j}^{-1}$ 
15:    $\varepsilon_j = |h_j(p) - p'|$ 
16: end for
17: Select optimal  $H \leftarrow h_j$  with min  $\varepsilon_j$ 
18:  $I2' \leftarrow H(I2)$ 
19: r map  $\leftarrow$  apply pwc-net to  $I1$  and  $I2'$ 
20:  $5 \times 5$  Patch extraction of the r map
21:   if SSIM (patch)  $> 0$  then
22:      $r'$  map  $\leftarrow$  r map
23:   end if
24: Median and quartile deviation  $\leftarrow r'$  map
    
```

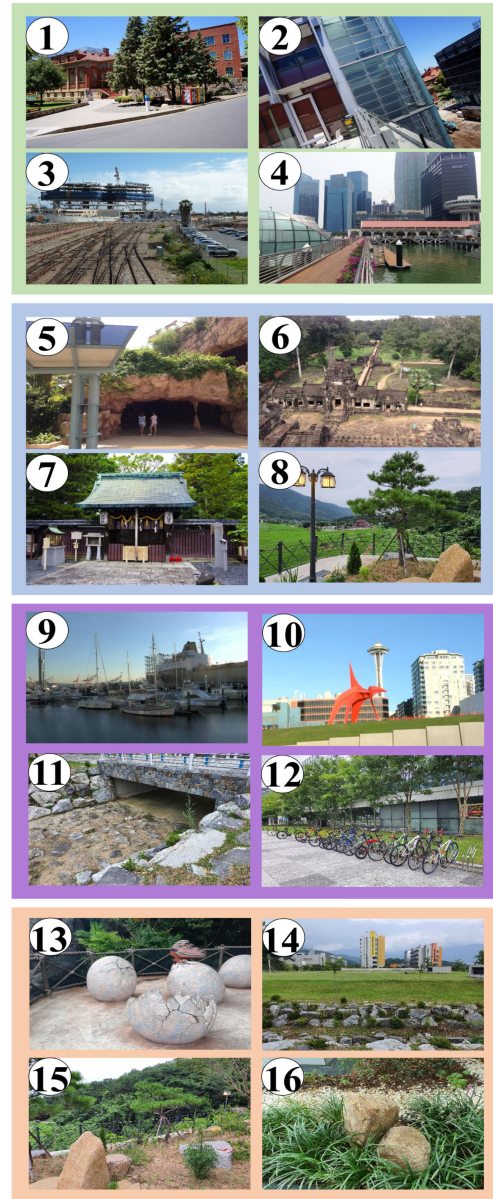
input images. We used a global homography optimization method. The homography was initialized using the direct least transformation (DLT) [35], and was iteratively optimized by applying the AutoStitch method [36]. The initial homography ( $h$ ) with the relationship  $p' = hp$  was estimated by the DLT, using four randomly selected pairs from the corresponding points. This process was repeated 500 times, and



**FIGURE 3.** Comparison of the planar parallax estimation. (a) In the results from the matching error metric, feature points were not extracted in the area of the rail (red box) where a parallax occurred (yellow lines); (b) the proposed metric can measure the displacement of planar parallax in the area of the rail. The color represents the magnitude of the planar parallax.

the homography with the most inlier-corresponding points ( $S$ ) within a threshold ( $\gamma$ ) was determined as the initial homography. Using the initial homography and inlier correspondences, the calibration parameters ( $K_1$  and  $K_2$ ) and camera rotation ( $R$ ) of the homography components were optimized using the Levenberg–Marquardt (LM) method [37], which is frequently used for non-linear least square problems because LM method can reliably find a solution even if an initial value is far from the solution. We repeated this optimization process 30 times, and the optimal homography with the least error ( $\epsilon$ ) was used to align the images.

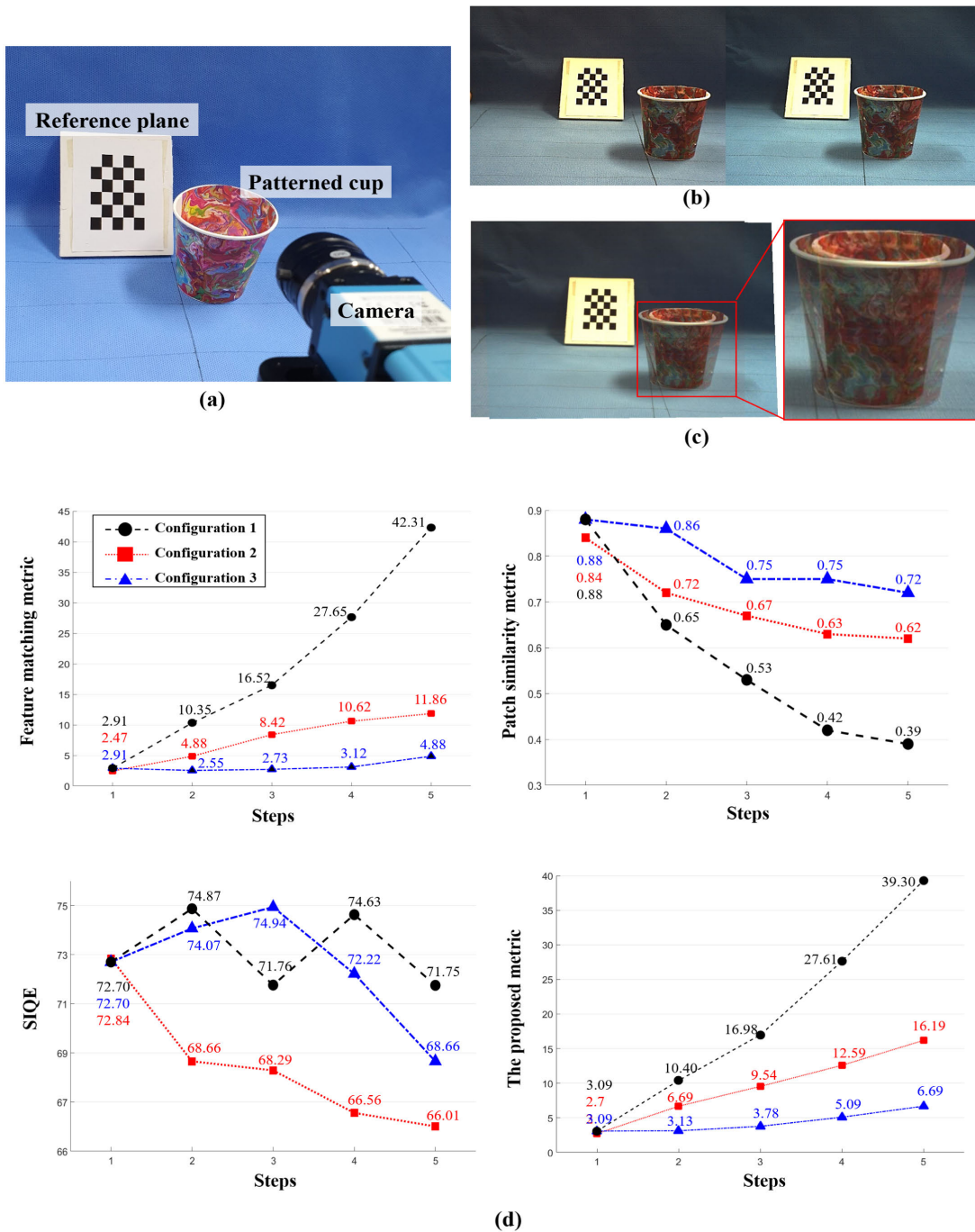
We estimated the planar parallax in the overlap area of the two aligned images using the proposed metric, which combined the matching error metric and patch similarity metric. Additionally, we improved the matching error metric by incorporating a “PWC-Net” optical flow [38] to extract the parallax vector in the overlap area where the feature points were not extracted (Figure 3). Maps were obtained for the x-directional parallax vector ( $x$ ) and y-directional parallax vector ( $y$ ), and a map of the planar parallax vector ( $r$ ) was obtained using  $x^2 + y^2$ . From the  $r$  map, we extracted the vector in the overlap area, and excluded the vectors whose final pixel of the vector displacement was outside the overlap area. In addition, if the SSIM similarity of the  $5 \times 5$  pixel corresponding patch (centered on the final pixel of the vector displacement) had a negative value, the vector was eliminated from the  $r$  map of the parallax vector. The filtered planar parallax vector ( $r'$ ) was expressed as a histogram with a left-skewed distribution. Therefore, the median and quartile deviation (QD) were used as the representative magnitude and variation in the planar parallax, respectively.



**FIGURE 4.** Image dataset. The magnitude and variation in the planar parallax in the image increase gradually from the first to fourth group. (1–3 are from [3]; 4–6 and 13 are from [7]; 7, 9, and 10 are from [8]; 8, 11, 12, and 14–16 are ours).

**C. COMPLETION OF THE IMAGE DATASET**

A total of 73 pairs of image datasets were quantified using the proposed metric. 44 pairs of images were obtained from the open image dataset provided in the SVA [9], DHW [21], APAP [3], SPHP [4], ANAP [20], and Seagull studies [7], along with the Zhang & Liu dataset [8]; such images have been used in many image stitching studies. The open image dataset included parallaxes owing to camera motion without moving objects in the static image. We included 29 pairs of images as supplementary images with moderate or large magnitudes, and variations that were insufficient for the open image dataset (Figure 4). Static indoor and outdoor environments without moving objects were captured using a Galaxy Note 10 camera.

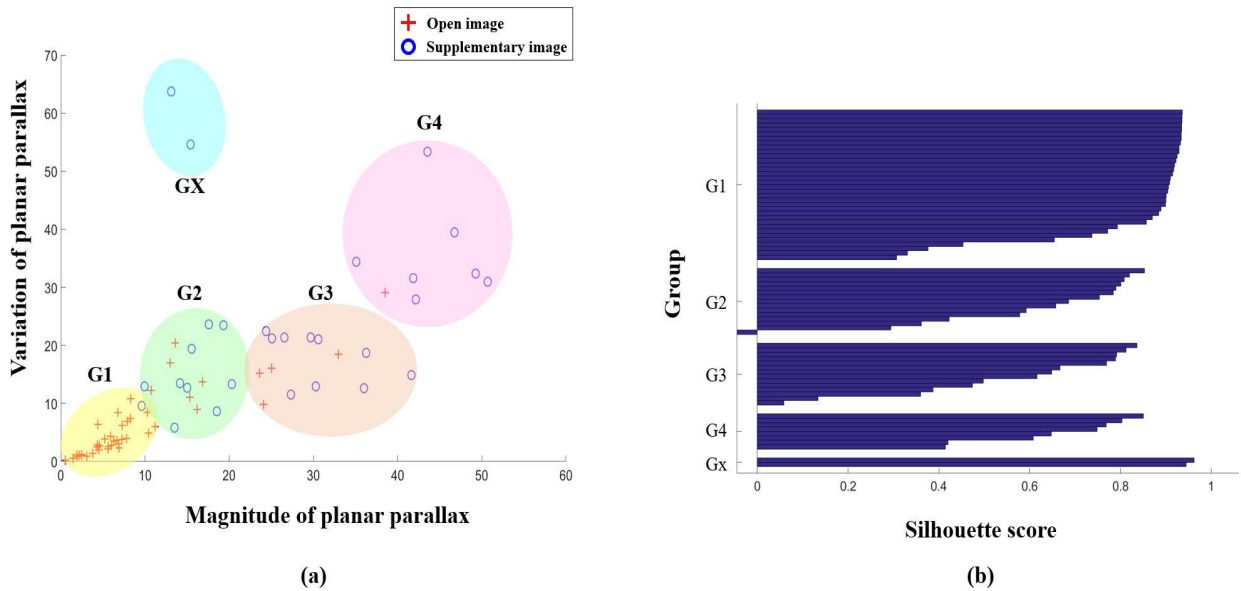


**FIGURE 5.** Reliability evaluation of the parallax estimation metric. (a) experimental setting, (b) parallax image pair, (c) residual planar parallax after warping using the homography, and (d) conventional and proposed metric results on planar parallax level of five steps in three configurations.

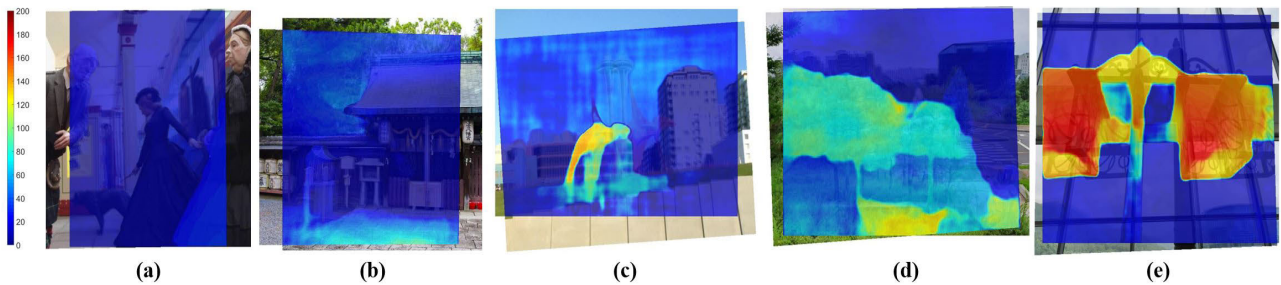
**D. GROUPING OF PARALLAX IMAGES**

Similar images in the 73 pairs of images were grouped based on the vector magnitude and on variation in the planar parallax in the overlap area. In most images, the small-sized parallax vector constituted a large portion of the r map, and the parallax vector in the r map formed a left-skewed distribution. Thus, the median value was used as a representative magnitude for the planar parallax vector in the input image. The variation in parallax was represented as a quartile deviation,

and indicated whether the magnitude of the parallax was uniform or non-uniform. The magnitude and variation values were normalized using feature scaling between 0 and 1, and the groups were determined using K-means clustering. The cluster results were evaluated using the Dunn index [39] and silhouette score [40]. The Dunn index is the ratio of the minimum distance between clusters to the maximum distance within clusters. Therefore, a higher index indicates better clustering, with data compactness and cluster separation.



**FIGURE 6.** Groups of parallax images. (a) Five groups by the magnitude and variation of the parallax, (b) silhouette score of the parallax images.



**FIGURE 7.** Planar parallax level of a representative image in each group using a color map. (a) G1, (b) G2, (c) G3, (d) G4, and (e) GX.

The silhouette score shows the suitability of the cluster for each dataset. The silhouette score ranges from  $-1$  to  $1$ , and a score closer to  $1$  indicates better clustering.

**IV. EXPERIMENTS AND RESULTS**

**A. RELIABILITY OF PARALLAX ESTIMATION METRIC**

The reliability of the proposed parallax estimation metric was verified based on correlations between known planar parallax levels and the outputs from the metric. Various parallax images with five levels were created by changing the four components (A, B, C, D) affecting the planar parallax. The experimental setup is shown in Figure 5. Using a patterned cup, chessboard, and camera, five levels of parallax images were produced in three configurations.

The first configuration involved changing the position of the cup in five steps, while the positions of the chessboard and camera were fixed. This indicated that the parallax level increased as component A increased and component B decreased from step 1 to step 5. The second configuration changed the distance between the cameras in five steps; this increased component D, resulting in an increase in the parallax in the images. In the third configuration, the positions of the chessboard and cup were fixed, and the position of the camera was moved along the depth direction

in five steps. The parallax level increased as components B and C decreased from step 1 to step 5. Theoretically, in the three configurations, the parallax would increase as the step increased. Thus, we confirmed the results from the metric based on increasing the parallax in five steps in each configuration.

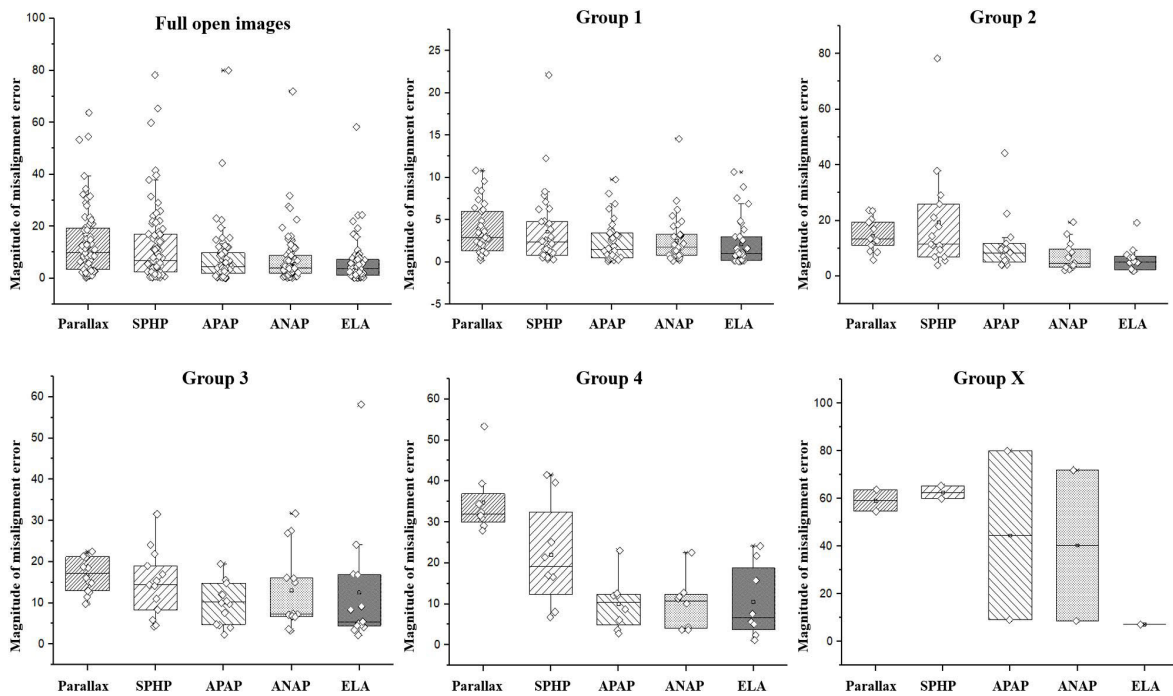
The parallax images with five levels in each configuration were aligned using the optimal homography estimated using chessboard points on the reference plane, and the planar parallax in the overlap was quantified. To exclude the parallax caused by the background other than the patterned cup, we manually extracted a boundary of the cup.

The parallax estimation performance of the proposed metric was compared with those of the conventional feature matching metric, the patch similarity metric, and the mathematical model SIQE. As the outputs of the feature matching metric and proposed metric are a distance of the parallax vector, the outputs are expected to increase as the parallax step increases. In contrast, the outputs of the patch similarity metric and SIQE are expected to decrease as the parallax step increases.

However, we found in the results that the feature matching metric, patch similarity metric, and SIQE did not reliably estimate the parallax. Although the feature matching metric

**TABLE 2.** Residual misalignment errors of existing image stitching methods for the parallax groups. The magnitude/ variation of the misalignment errors are described, and parallax improvement rate is shown in parentheses (in %). SPHP: Shape-preserving half-projective; APAP: As-projective-as-possible; ANAP: As-natural-as-possible; ELA: Elastic local alignment.

Methods	Parallax groups					Average of groups	Success (%)
	G1	G2	G3	G4	GX		
Planar parallax	5.8/2.9	15.3/13.3	28.5/17.3	42.9/32.0	14.3/59.1	21.3/24.9	-
SPHP	2.5/2.3 (57.0/20.3)	10.8/11.5 (30.0/13.5)	19.4/14.4 (32.0/16.6)	23.7/19.2 (44.7/39.8)	17.5/62.6 (-22.9/-5.8)	14.8/22.0 (28.1/16.9)	98.57
APAP	2.3/1.5 (60.8/50.2)	7.6/8.4 (50.5/37.1)	10.0/10.2 (65.0/40.8)	10.9/10.3 (74.6/67.7)	14.0/44.6 (2.1/24.7)	8.9/15.0 (50.6/44.1)	<b>100</b>
ANAP	2.8/1.7 (52.0/42.0)	6.7/4.5 (56.5/66.1)	12.3/7.3 (57.0/57.6)	14.1/10.8 (67.2/66.3)	13.5/40.4 (5.3/31.7)	9.9/12.9 (47.6/52.7)	98.57
ELA	<b>0.9/1.0</b> <b>(85.1/65.8)</b>	<b>3.0/5.0</b> <b>(80.7/62.5)</b>	<b>6.8/5.4</b> <b>(76.2/68.6)</b>	<b>7.0/6.7</b> <b>(83.6/79.1)</b>	<b>9.4/7.2</b> <b>(33.9/87.8)</b>	<b>5.4/5.1</b> <b>(71.9/72.8)</b>	92.86



**FIGURE 8.** Box plots of the magnitude of the residual misalignment error of the existing image stitching methods with respect to the parallax groups.

showed better performance than the patch similarity metric and SIQE, the metric results at step 1 (with the smallest parallax level) were greater than those at steps 2 and 3. In the results of the patch similarity metric, identical outputs were obtained at different parallax levels (steps 3 and 4) in configuration 3. In the SIQE results, the SIQE scores were not related to the changes in parallax levels in configurations 1 and 3. In contrast, the output of the proposed metric constantly increased when the parallax increased from steps 1 to 5, for all

three configurations (Figure 5-d). In the first configuration, the planar parallax exhibited the greatest increase at each step. The rate of increase varied depending on the arrangements of the reference plane, object, and camera.

**B. GROUPING OF PARALLAX IMAGES**

The 73 pairs of parallax images (comprising the open and supplementary images) were divided into five groups according to the magnitude and variation of the parallax (Figure 6-a).



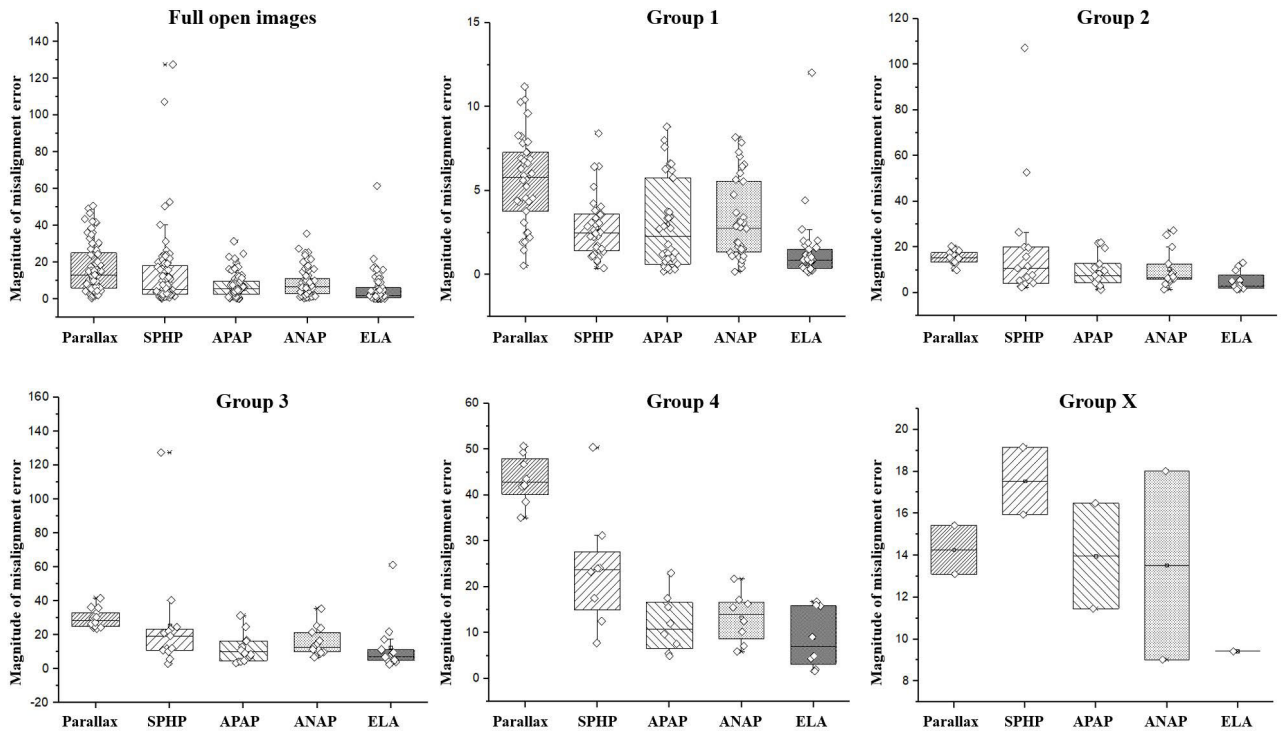


FIGURE 9. Box plots of the variation in the residual misalignment error of the existing image stitching methods with respect to the parallax groups.

The five groups were termed G1, G2, G3, G4, and GX. Most of the open image dataset belonged to G1, with a magnitude ranging from 0 to 10, and a variance ranging from 0 to 10. The supplementary images belonged to other groups, with magnitudes ranging from 10 to 50 and variances ranging from 5 to 65. Most of the images exhibited a tendency of the variation increasing with the parallax magnitude. Two exceptional images in GX had a small magnitude, but a large variation. This occurred very rarely, e.g., when distant points were aligned by homography and a large displacement occurred at the near points in the image with a large depth difference. The Dunn index of the five parallax groups was 0.1147 and the average silhouette score was 0.7172, including a minus silhouette score (Figure 6-(b)).

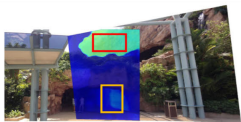
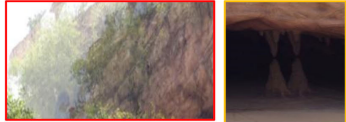
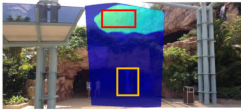
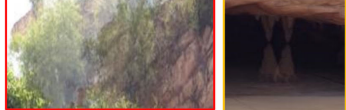
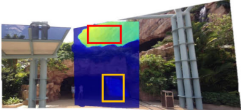

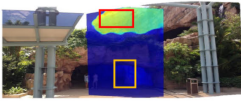

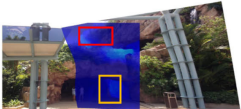

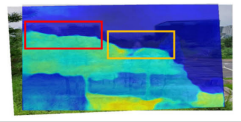
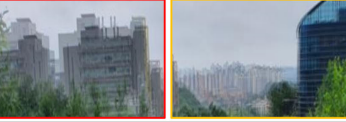
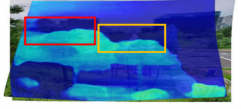

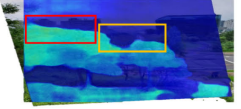

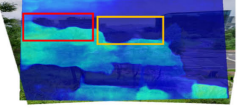

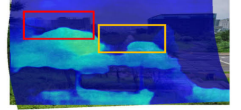

Figure 7 shows the parallax levels for a representative image from each group using a color map. Blue and red represent small and large planar parallax levels, respectively. The planar parallax magnitudes and variations of each grouped image are provided in Table 2. The median magnitude of the planar parallax for each group increased in the order of G1, G2, G3, and G4. The median parallax magnitude of group GX was similar to that of G2. In the variation, the average QD value of each group increased in the order of groups G1, G2, G3, G4, and GX. A two-way ANOVA was used to statistically compare the parallax level in each group; the parallax magnitude and variation were considered as two independent variables. The level of significance was set at  $P < 0.01$ . Significant differences were found between the groups ( $P < 0.01$ ) in both the parallax magnitude and variation.

C. PERFORMANCE ASSESSMENT ON GROUPED IMAGES

The residual misalignment error of the existing image stitching methods was quantitatively verified and compared for each grouped image (Table 2). The existing parallax-tolerant methods, i.e., SPHP, APAP, ANAP, and ELA, generated the resultant panoramic images for the 73 pairs of images. The misalignment errors in the overlap areas of the panoramic images were estimated using the proposed metric to assess the performance of each method.

We quantitatively confirmed the extent to which the stitching methods improved the magnitude and variation in the parallax on the images of each group (Figures 8 and 9). All methods exhibited the smallest magnitude and variation in the misalignment error in group G1, and the largest error in group G4. In most groups, ELA exhibited the best improvement in magnitude and variation, resulting in the lowest average error regarding residual misalignment. APAP and ANAP exhibited similar improvements in magnitude and variation; SPHP showed less improvement. A comparison of SPHP and ANAP (which were developed for naturality rather than accurate alignment) indicated that SPHP outperformed ANAP in group G1, but the results were reversed in other groups. The misalignment errors of the existing methods can be observed in a color map (Figure 10). Overall, the parallax improvement performance of the ELA algorithm resulted in the lowest misalignment error in the G2 and G4 images, respectively.

However, the panorama generation of the ELA algorithm occasionally failed, with severe image distortions. The probability of success for creating a panoramic image was

Methods	G2 image		
	Magnitude/ variation	Metric result	Misaligned image
Planar parallax	13.0/17.0		
SPHP	4.3/5.5		
APAP	2.5/11.7		
ANAP	5.8/11.6		
ELA	2.7/4.7		
Methods	G4 image		
	Magnitude/ variation	Metric result	Misaligned image
Planar parallax	41.9/31.6		
SPHP	24.1/25.1		
APAP	23.0/23.0		
ANAP	16.3/22.6		
ELA	16.0/24.3		

**FIGURE 10.** Resulting stitched images of the existing method using the input images of groups G2 and G4.

the lowest in ELA, and the highest in APAP. The average improvement from the four existing methods with respect to each parallax group was the least for group GX.

**V. DISCUSSION**

A quantitative and objective parallax analysis of input images is essential to specifying the performance of parallax-tolerant image stitching methods. Assessing only the resulting panoramic image may result in an inaccurate understanding.

In this study, a parallax estimation metric was proposed for quantifying the parallax level, and was used to group the images within a constant range. Using the grouped images, the parallax-tolerant performances of the existing methods were verified for various parallax levels.

In the reliability test for the parallax estimation metric, the output of the proposed metric showed a continuous increase responding to the increase in the parallax level. In contrast, the feature matching metric and patch similarity

metric (as well as the SIQE scores) did not show consistent trend with respect to the parallax level, particularly in configuration 3. When the increase rate of the parallax level was relatively small like the case of configuration 3, it appears difficult to employ the existing metrics.

Significant differences between the parallax image groups were found by the two-way ANOVA test. The five groups, constructed based on different parallax levels, can be used as a standard evaluation dataset for any image stitching method. All of the image stitching methods exhibited the least misalignment error for group G1, and the largest error for group G4. This result suggested that the different parallax levels of the input images could affect the performance of the resultant stitched image. Therefore, the parallax level of an input image should be evaluated when performance verification is required. In the experiments, the quantitative assessment using the proposed five groups showed the changeable performance of each stitching method depending on the parallax level. In most groups, ELA exhibited the least residual misalignment error. However, the probability of success in the image stitching was the lowest. As the number of inlier matching points decreased, the performance of the feature refinement decreased, affecting the warping model. ELA seemed to be sensitive to the outliers of matching points, as the alignment term was based on a few anchor points [5]. When comparing SPHP and ANAP, which were developed for natural panoramic images as well as accurate alignment, SPHP appeared better for G1, and ANAP appeared better for the other groups. For a natural panoramic image with an accurate alignment, either SPHP or ANAP could be selected, based on the parallax level(s) of the input images.

The five groups with similar parallax levels were created using K-means clustering, and the validity of the group assignment was evaluated using the Dunn index and silhouette score. An image of G2 had a low negative silhouette score. The average distance between images in G2 seemed to be larger than that in G1, owing to the different levels of compactness between the image data in each group. Therefore, if additional data is included in the G2 parallax group, the Dunn index and silhouette score will increase.

We also observed it is difficult to completely correct the misalignment in the entire overlap area by warping the image when the parallax magnitude is in the range of groups G2 to G4. A seam-based image composition is more suitable than a blending-based image composition that should align all the pixels of the overlap area by image warping.

This study had the following limitations. The planar parallax was affected by global homography. We used the homography that satisfied most of the feature points in the entire area with multiple iterations. If a reference homography can be calculated for the input image, the parallax quantification will be more reliable. An additional limitation is that error vectors may have arose in the proposed metric when the image had a large plain texture or wide baseline. Although the parallax vector calculated by PWC-Net was filtered through the SSIM image similarity, the error vector may have been included.

In addition, different optical flow methods may affect the parallax vector estimation. However, minor changes may have occurred in the representative parallax of the image owing to the error vector or the different methods, as the median value was selected as the representative parallax magnitude of the image. Finally, the parallax groups were created based on the magnitude and variation in the parallax. These criteria can be changed depending on the aim, and the changed criteria will result in different groups.

The proposed assessment method may aid in tuning the performance of an image stitching method according to the parallax level of the input images. This study may also contribute to other computer vision fields, including reconstruction from 2D to 3D data or analyzing 3D motion with parallax images.

## VI. CONCLUSION

Quantifying and clustering parallax input images is important for objectively evaluating image stitching methods, as the performance differs according to the parallax level of the input images. In this study, using a proposed parallax estimation metric, parallax images were assigned into five groups to create a standard testing dataset. In the dataset, the parallax was evenly distributed, by supplementing the insufficient parallax-level images. The proposed method can quantitatively assess the improvements of the parallax in a stitched panoramic image, and can aid in finding an appropriate image stitching method suitable for a parallax problem. This study suggests a new performance assessment method for image stitching, and indicates the importance of analyzing the input parallax image.

## REFERENCES

- [1] J. Zheng, Y. Wang, H. Wang, B. Li, and H.-M. Hu, "A novel projective-consistent plane based image stitching method," *IEEE Trans. Multimedia*, vol. 21, no. 10, pp. 2561–2575, Oct. 2019.
- [2] G. Zhang, Y. He, W. Chen, J. Jia, and H. Bao, "Multi-viewpoint panorama construction with wide-baseline images," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3099–3111, Jul. 2016.
- [3] J. Zaragoza, T.-J. Chin, Q.-H. Tran, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1285–1298, Jul. 2014.
- [4] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, "Shape-preserving half-projective warps for image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3254–3261.
- [5] J. Li, Z. Wang, S. Lai, Y. Zhai, and M. Zhang, "Parallax-tolerant image stitching based on robust elastic warping," *IEEE Trans. Multimedia*, vol. 20, no. 7, pp. 1672–1687, Jul. 2018.
- [6] Y.-S. Chen and Y.-Y. Chuang, "Natural image stitching with the global similarity prior," in *Proc. 14th Eur. Conf. Comput. Vis.*, Sep. 2016, pp. 186–201.
- [7] K. Lin, N. Jiang, L.-F. Cheong, M. Do, and J. Lu, "Seagull: Seam-guided local alignment for parallax-tolerant image stitching," in *Proc. 14th Eur. Conf. Comput. Vis.*, Sep. 2016, pp. 370–385.
- [8] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3262–3269.
- [9] W.-Y. Lin, S. Liu, Y. Matsushita, T.-T. Ng, and L.-F. Cheong, "Smoothly varying affine stitching," in *Proc. CVPR*, Jun. 2011, pp. 345–352.
- [10] T.-Z. Xiang, G.-S. Xia, X. Bai, and L. Zhang, "Image stitching by line-guided local warping with global similarity constraint," *Pattern Recognit.*, vol. 83, pp. 481–497, Nov. 2018.

- [11] G. M. Jacob and S. Das, "Large parallax image stitching using an edge-preserving Diffeomorphic warping process," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst. (ACIVS)*, Sep. 2018, pp. 521–533.
- [12] W. Yan, G. Yue, Y. Fang, H. Chen, C. Tang, and G. Jiang, "Perceptual objective quality assessment of stereoscopic stitched images," *Signal Process.*, vol. 172, Jul. 2020, Art. no. 107541.
- [13] V. Dissanayake, S. Herath, S. Rasnayaka, S. Seneviratne, R. Vidanaarachchi, and C. Gamage, "Quantitative and qualitative evaluation of performance and robustness of image stitching algorithms," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2015, pp. 1–6.
- [14] H. S. Qureshi, M. M. Khan, R. Hafiz, Y. Cho, and J. Cha, "Quantitative quality assessment of stitched panoramic images," *IET Image Process.*, vol. 6, no. 9, pp. 1348–1358, Dec. 2012.
- [15] M. S. Billah and H. Ahn, "Objective quality assessment method for stitched images," *J. Broadcast Eng.*, vol. 23, no. 2, pp. 227–234, Mar. 2018.
- [16] G. Cheung, L. Yang, Z. Tan, and Z. Huang, "A content-aware metric for stitched panoramic image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 2487–2494.
- [17] L. Yang, J. Liu, and C. Gao, "An error-activation-guided blind metric for stitched panoramic image quality assessment," in *Proc. Chinese Conf. Comput. Vis. (CCCV)*, Nov. 2017, pp. 256–268.
- [18] S. Ling, G. Cheung, and P. Le Callet, "No-reference quality assessment for stitched panoramic images using convolutional sparse coding and compound feature selection," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.
- [19] P. C. Madhusudana and R. Soundararajan, "Subjective and objective quality assessment of stitched images for virtual reality," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5620–5635, Nov. 2019.
- [20] C.-C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin, "Adaptive as-natural-as-possible image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1155–1163.
- [21] J. Gao, S. J. Kim, and M. S. Brown, "Constructing image panoramas using dual-homography warping," in *Proc. CVPR*, Jun. 2011, pp. 49–56.
- [22] R. Rennie and J. Law, *A Dictionary of Physics*, 8th ed. London, U.K.: Oxford Univ. Press, 2019.
- [23] M. Sawhney, "3D geometry from planar parallax," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1994, pp. 929–934.
- [24] M. Irani, P. Anandan, and M. Cohen, "Direct recovery of planar-parallax from multiple frames," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 11, pp. 1528–1534, Nov. 2002.
- [25] M. Irani and P. Anandan, "A unified approach to moving object detection in 2D and 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 6, pp. 577–589, Jun. 1998.
- [26] M. Irani, P. Anandan, and D. Weinshall, "From reference frames to reference planes: Multi-view parallax geometry and applications," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, May 1998, pp. 829–845.
- [27] R. Kumar, P. Anandan, and K. Hanna, "Direct recovery of shape from multiple views: A parallax based approach," in *Proc. 12th Int. Conf. Pattern Recognit.*, 1994, pp. 685–688.
- [28] J. Liu, B. Wang, W. Hu, P. Sun, J. Li, H. Duan, and J. Si, "Global and local panoramic views for gastroscopy: An assisted method of gastroscopic lesion surveillance," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 9, pp. 2296–2307, Sep. 2015.
- [29] K. Joo, N. Kim, T.-H. Oh, and I. S. Kweon, "Line meets as-projective-as-possible image stitching with moving DLT," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 1175–1179.
- [30] H. S. Sawhney, "Simplifying motion and structure analysis using planar parallax and image warping," in *Proc. 12th Int. Conf. Pattern Recognit.*, 1994, pp. 403–408.
- A. Criminisi, I. Reid, and A. Zisserman, "Duality, rigidity and planar parallax," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jun. 1998, pp. 846–861.
- [32] M. Irani and P. Anandan, "Parallax geometry of pairs of points for 3D scene analysis," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Apr. 1996, pp. 17–30.
- [33] J. Wulff, L. Sevilla-Lara, and M. J. Black, "Optical flow in mostly rigid scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4671–4680.
- [34] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1150–1157.
- [35] Z. Zhang, "Parameter estimation techniques: A tutorial with application to conic fitting," *Image Vis. Comput.*, vol. 15, no. 1, pp. 59–76, Jan. 1997.
- [36] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, Apr. 2007.
- [37] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Soc. for Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, Jun. 1963.
- [38] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "PWC-net: CNNs for optical flow using pyramid, warping, and cost volume," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8934–8943.
- [39] M. K. Pakhiza, S. Bandyopadhyay, and U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognit.*, vol. 37, no. 3, pp. 487–501, Mar. 2004.
- [40] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.
- [41] X. Pan and G. Wang, "Parallax-tolerant image stitching based on mesh optimization," in *Proc. IEEE 2nd Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Mar. 2017, pp. 414–420.
- [42] L. Wei, Z. Zhong, C. Lang, and Z. Yi, "A survey on image and video stitching," *Virtual Reality Intell. Hardw.*, vol. 1, no. 1, pp. 55–83, Aug. 2019.
- [43] S. Li, L. Yuan, J. Sun, and L. Quan, "Dual-feature warping-based motion model estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4283–4291.



**KYUNGHWA JUNG** received the B.E. degree from the Department of Biomedical Engineering, Yonsei University, Wonju, South Korea, in 2013, and the M.S. degree in robotics engineering from the Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, South Korea, in 2015, where she is currently pursuing the Ph.D. degree. Her research interests include image processing, computer vision, and surgical navigation.



**JAESUNG HONG** (Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from Kyungpook National University, Daegu, South Korea, in 1995 and 1997, respectively, and the Ph.D. degree in frontier science from The University of Tokyo, Tokyo, Japan, in 2004. In 2004, he worked as a Foreign Researcher of the Japan Society for the Promotion of Science with the Graduate School of Information Science and Technology, The University of Tokyo. From 2005 to 2010, he worked as a Faculty Member with Kyushu University, Fukuoka, Japan. Since 2010, he has been with the Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, where he is currently a Professor with the Department of Robotics Engineering. His research interests include surgical navigation, surgical robots, and medical augmented reality. He has been a Board Member of the Korean Society of Medical Robotics and the Asian Society of Computer Aided Surgery.

...