Warping Residual Based Image Stitching for Large Parallax

Kyu-Yul Lee and Jae-Young Sim
School of Electrical and Computer Engineering, UNIST, Ulsan, Korea
ever1135@unist.ac.kr and jysim@unist.ac.kr

Abstract

Image stitching techniques align two images captured at different viewing positions onto a single wider image. When the captured 3D scene is not planar and the camera baseline is large, two images exhibit parallax where the relative positions of scene structures are quite different from each view. The existing image stitching methods often fail to work on the images with large parallax. In this paper, we propose an image stitching algorithm robust to large parallax based on the novel concept of warping residuals. We first estimate multiple homographies and find their inlier feature matches between two images. Then we evaluate warping residual for each feature match with respect to the multiple homographies. To alleviate the parallax artifacts, we partition input images into superpixels and warp each superpixel adaptively according to an optimal homography which is computed by minimizing the error of feature matches weighted by the warping residuals. Experimental results demonstrate that the proposed algorithm provides accurate stitching results for images with large parallax, and outperforms the existing methods qualitatively and quantitatively.

1. Introduction

Image stitching is an important technique for diverse computer vision applications which aligns multiple images captured from different viewing positions onto a common coordinate domain to generate an image with wider field of view. Recently, many commercial products using image stitching techniques have been released such as 360° panorama camera[1] and surround-view monitoring systems [2]. Also, image stitching software products were provided to synthesize multiple images, e.g., Adobe Photoshop Photomerge™ and Autostitch [2].

Most of the conventional image stitching methods follow similar procedures [19]. Feature points are first detected from a pair of input images, and their correspondence matches are found between the images. Then parametric image warping models are estimated by using the detected feature matches, which warp a target image onto a reference image domain. Finally, we composite an output stitched image by determining the pixel values in the overlapped areas between the warped target image and the reference image.

One of the most crucial and challenging steps of image stitching is image warping. Homography is a simple and traditional image warping model which describes the parametric planar transformation based on the planar scene assumption [9]. However, when the captured scene is not planar including foreground objects at different scene depths and the camera baseline is large, we observe the parallax phenomenon where the relative positions of the objects are different from two images. In such cases, the stitching results using planar transformation models such as homography often exhibit parallax artifacts in the vicinity of the object boundaries.

To alleviate the parallax artifacts of image stitching, adaptive warping algorithms have been proposed which partition an image into regular grid cells or pixels and warp the partitions by different models [7, 10, 11, 15, 22, 24]. Energy minimization frameworks were applied to optimize the adaptive warps to prevent the distortion in the warped images [11, 15, 24]. Local alignment techniques were proposed which align only a specific image region while hiding the artifacts in other misaligned regions based on seam-cutting methods [8, 14, 23]. However, for images with large parallax, a group of neighboring pixels in one image may not have the corresponding pixels adjacent each other in another image, which causes severe parallax artifacts in resulting stitched images obtained by the existing smooth warping based methods [7, 11, 15, 22, 24]. A video stitching method has been proposed which addresses the large parallax problem based on the epipolar geometry [10], however it cannot be directly applied to image stitching due to the lack of temporal motion information of video sequences.

In this paper, we propose a warping residual based stitching algorithm for images with large parallax. We first partition input images into superpixels and warp the superpixels adaptively, since the parallax phenomenon usually occurs in the vicinity of object boundaries. We detect feature points

[1] https://www.panono.com/
[2] https://www.bmwblog.com/2019/04/18/video-bmw/

from two images and find their correspondence matches which are then used to estimate multiple homographies and their associated inlier matches. We find an optimal homography for each superpixel using the feature matches, where the contribution of each feature point is adaptively computed according to the warping residuals. The warping residuals for a given superpixel alleviate the parallax artifacts when warping the superpixel by emphasizing the feature points located on the regions with similar scene depths. Furthermore, we refine the initially estimated homography at each superpixel using that of the neighboring superpixels for reliable warping. Experimental results show that the proposed algorithm accurately aligns the images with large parallax and outperforms the conventional image stitching methods qualitatively and quantitatively.

The rest of this paper is organized as follows. Section 2 introduces the related work of image stitching. Section 3 proposes a novel concept of warping residuals. Section 4 describes the image warping algorithm. Section 5 presents the experimental results. Section 6 concludes the paper.

2. Related Work

Adaptive Warping Methods Gao *et al*. proposed a dual homography method that blends the homography estimated for the distant plane with the homography estimated for the ground plane adaptively according to the positions of feature points [7]. Lin *et al*. computed a spatially-varying affine transformation where an initially estimated global transformation is refined to the optimal one by minimizing a cost function [15]. Zaragoza *et al*. divided an input image into regular grid cells and estimated an optimal homography for each cell by moving direct linear transformation (MDLT) [22] which assigns more weights to the feature points spatially closer to the target cell when computing the alignment error. Zhang *et al*. employed a scale preserving term and a line preserving term to minimize the distortion in the warped images [24]. Li *et al*. approximated the projection bias caused by the homography of the matched points by using an analytical warping function based on the thin-plate spline [18] with radial basis functions [11]. Lee and Sim proposed a video stitching algorithm for large parallax based on the epipolar geometry [10]. Note that [7] does not handle the scenes with more than two planar structures. The other methods [11, 15, 22, 24] can warp the background composed of multiple planar regions, however, they usually assume continuous scene depths with small parallax and often fail to align foreground objects with large parallax having abrupt depth changes from the background. Also, [10] requires temporal motion information of foreground objects which is not available for image stitching. In contrary, the proposed method can warp both of the background and multiple foreground objects at different scene depths between two images with large parallax.

Shape-Preserving Warps Whereas the overlapping regions between two images are well aligned by using valid feature matches, the non-overlapping regions usually exhibit severe perspective distortions. Chang *et al*. applied the homography transformation to the overlapping regions and applied the similarity transformation to the non-overlapping regions, respectively [3]. Lin *et al*. proposed a homography linearization method that smoothly extrapolates the warps for the overlapping regions to the non-overlapping regions [13]. Chen *et al*. estimated proper scale and rotation for each image and designed an objective function for warping estimation based on a global similarity-prior [4]. Li *et al*. proposed a quasi-homography warp to solve the line bending problem between the homography transformation and the similarity transformation of [3] by linearly scaling the horizontal component of the homography [12]. Note that the shape-preserving warps are usually designed to mitigate the perspective distortion in non-overlapping regions between two images, whereas the main purpose of this paper is to align common visual contents in the overlapped regions between two images with large parallax.

Seam-Based Methods Gao *et al*. defined a seam-cutting loss for the homography that measures the discontinuity between the warped target image and the reference image [8]. They estimated multiple homographies using RANSAC [6] and selected an optimal homography having the minimum seam-cutting loss. Zhang *et al*. estimated the local homography which aligns a certain image region only, and applied the contents preserving warping (CPW) [16] to further refine the alignment [23]. The misalignment artifacts are hidden by the seam-cutting method. Lin *et al*. improved the stitching performance gradually by using the iterative warp and seam estimation [14]. The seam-based methods usually align certain local image regions only to provide visually pleasing results of image stitching, which may not be geometrically accurate over an entire image area.

3. Warping Residuals for Large Parallax

We review the mathematical framework of MDLT [22], which is one of the adaptive image warping models. Then we introduce a novel concept of warping residuals which assign high weights to the feature points located at similar scene depths to a given superpixel when computing the alignment error of the warped superpixel.

3.1. Moving Direct Linear Transformation

Let $\mathbf{X}$ be a real-world point on a plane π in 3D space, and let $\mathbf{x} = [x_1, x_2, 1]^T$ and $\mathbf{y} = [y_1, y_2, 1]^T$ are the pixels of $\mathbf{X}$ projected onto two images $\mathcal{I}$ and $\mathcal{J}$, respectively. The relationship between the two pixels is described by a 3×3

homography matrix $\mathbf{H}$ induced by the plane π.

$$\mathbf{y} \sim \mathbf{H}\mathbf{x} \tag{1}$$

where $\sim$ indicates the equality up to scale. Since the two locations of $\mathbf{H}\mathbf{x}$ and $\mathbf{y}$ are same in $\mathcal{J}$, we have [22]

$$\begin{aligned} \mathbf{0}_{3\times 1} &= \mathbf{y} \times \mathbf{H}\mathbf{x} \\ &= \begin{bmatrix} \mathbf{0}_{1\times 3} & -\mathbf{x}^T & y_2\mathbf{x}^T \\ \mathbf{x}^T & \mathbf{0}_{1\times 3} & -y_1\mathbf{x}^T \\ -y_2\mathbf{x}^T & y_1\mathbf{x}^T & \mathbf{0}_{1\times 3} \end{bmatrix} \mathbf{h} \end{aligned} \tag{2}$$

where $\mathbf{h}$ is obtained by vectorizing the rows of $\mathbf{H}$. When $\mathbf{x}$ and $\mathbf{y}$ are not projected from a same scene point on π and the camera baseline between $\mathcal{I}$ and $\mathcal{J}$ is not sufficiently small, the equation (2) does not hold and the norm $\|\mathbf{y} \times \mathbf{H}\mathbf{x}\|$ can be regarded as the alignment error of $\mathcal{I}$ and $\mathcal{J}$.

Direct linear transformation (DLT) [25] estimates the optimal homography $\hat{\mathbf{h}}$ from K matching pairs of pixels between $\mathcal{I}$ and $\mathcal{J}$, which minimizes the algebraic error as

$$\hat{\mathbf{h}} = \text{argmin}_{\mathbf{h}} \sum_{k=1}^{K} \|\mathbf{g}_k\mathbf{h}\|^2, \qquad \text{s.t.} \|\mathbf{h}\| = 1, \tag{3}$$

$$= \text{argmin}_{\mathbf{h}} \|\mathbf{G}\mathbf{h}\|^2, \qquad \text{s.t.} \|\mathbf{h}\| = 1, \tag{4}$$

where $\mathbf{g}_k$ is the first-two rows of the RHS matrix in (2) for the k-th matching pair of pixels and $\mathbf{G} \in \mathbb{R}^{2K\times 9}$ is obtained by stacking $\mathbf{g}_k$'s for all the K matching pairs. The optimal homography matrix $\hat{\mathbf{H}}$ is obtained from the solution $\hat{\mathbf{h}}$, the least significant right singular vector of $\mathbf{G}$.

Note that this global homography may yield significant parallax artifacts when aligning the images with non-planar 3D scenes and large camera baselines. To alleviate this artifacts, Zaragoza *et al*. [22] modified DLT to the moving DLT (MDLT) which partitions the images into regular grid cells and estimates the homography for each cell adaptively. The optimal homography $\hat{\mathbf{H}}_i$ (equivalently its vectorized form $\hat{\mathbf{h}}_i$) for the i-th cell is estimated by employing a weight matrix $\mathbf{W}_i$ to (4) as

$$\hat{\mathbf{h}}_i = \text{argmin}_{\mathbf{h}} \|\mathbf{W}_i\mathbf{G}\mathbf{h}\|^2, \qquad \text{s.t.} \|\mathbf{h}\| = 1, \tag{5}$$

where $\mathbf{W}_i \in \mathbb{R}^{2K\times 2K}$ is a diagonal weight matrix for the i-th cell given by

$$\mathbf{W}_i = \text{diag}([w_{i,1}\ w_{i,1}\ w_{i,2}\ w_{i,2}\ \cdots\ w_{i,K}\ w_{i,K}]). \tag{6}$$

The weight $w_{i,k}$ is defined using the spatial distance between the center pixel $\mathbf{c}_i$ of the i-th cell and the k-th feature point $\mathbf{x}_k$ given by

$$w_{i,k} = \max(\exp(-\|\mathbf{c}_i - \tilde{\mathbf{x}}_k\|^2/\sigma^2), \gamma). \tag{7}$$

where $\tilde{\mathbf{x}}$ means the 2D pixel coordinates of the homogeneous coordinates $\mathbf{x}$.

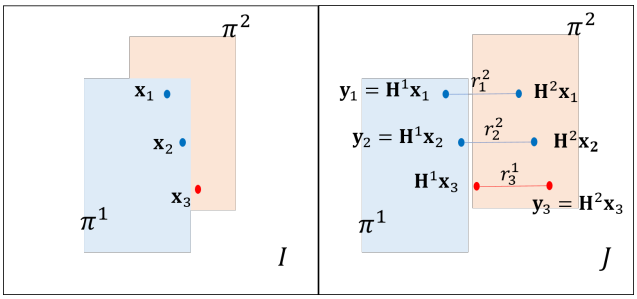


Figure 1. Visualization of warping residuals associated with the three feature matches of $\{\mathbf{x}_k \leftrightarrow \mathbf{y}_k\}_{k=1}^3$ with respect to the two planes of π^1 and π^2.

3.2. Warping Residual Based Transformation

The spatial distance based weight defined in (7) alleviates the artifacts of image warping methods based on the global homography, however it still suffers from severe misalignment artifacts especially on the challenging images with large parallax. As illustrated in Figure 1, let us consider three feature points where the two points $\mathbf{x}_1$ and $\mathbf{x}_2$ lie on the plane π^1 while the point $\mathbf{x}_3$ lies on the other plane π^2, respectively. We assume the two planes π^1 and π^2 have different scene depths from each other causing severe parallax, and therefore the relative locations of π^1 and π^2 in $\mathcal{J}$ are largely different from that in $\mathcal{I}$. Consider the three matching pairs $\{\mathbf{x}_k \leftrightarrow \mathbf{y}_k\}_{k=1}^3$ between $\mathcal{I}$ and $\mathcal{J}$. While the relative locations between $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$ in $\mathcal{I}$ are same to that of the matched pixels $\tilde{\mathbf{y}}_1$ and $\tilde{\mathbf{y}}_2$ in $\mathcal{J}$, the relative locations between $\tilde{\mathbf{x}}_2$ and $\tilde{\mathbf{x}}_3$ in $\mathcal{I}$ are quite different from that of the matched pixels $\tilde{\mathbf{y}}_2$ and $\tilde{\mathbf{y}}_3$ in $\mathcal{J}$, as shown in Figure 1. The existing weighting scheme in (7) assigns higher weights to spatially closer feature points regardless of different scene depths, and hence yields parallax artifacts in the vicinity of object boundaries.

To overcome this drawback, we first extract multiple possible homography matrices of $\mathbf{H}^m$'s between $\mathcal{I}$ and $\mathcal{J}$, and introduce a warping residual r_k^m for the k-th matched point $\mathbf{x}_k$ with respect to the m-th homography $\mathbf{H}^m$ as

$$r_k^m = \|\tilde{\mathbf{y}}_k - \tilde{\mathbf{y}}_k^m\|. \tag{8}$$

where $\mathbf{y}_k^m = \mathbf{H}^m\mathbf{x}_k$. In Figure 1, we consider two homographies $\mathbf{H}^1$ and $\mathbf{H}^2$ associated with π^1 and π^2, and warp the three points $\{\mathbf{x}_k\}_{k=1}^3$ using the two homographies, respectively. Since $\mathbf{x}_1$ and $\mathbf{x}_2$ lie on π^1, the warping residuals of r_1^1 and r_2^1 induced by $\mathbf{H}^1$ are small but r_1^2 and r_2^2 induced by $\mathbf{H}^2$ becomes large. In contrary, $\mathbf{x}_3$ lies on π^2 and thus results in a small warping residual of r_3^2 but a large residual of r_3^1, respectively. For a given k-th feature point $\mathbf{x}_k$, we define the warping residual vector $\mathbf{r}_k$ using M homographies $\{\mathbf{H}^m\}_{m=1}^M$ given by

$$\mathbf{r}_k = [r_k^1\ r_k^2\ \cdots\ r_k^M]^T. \tag{9}$$

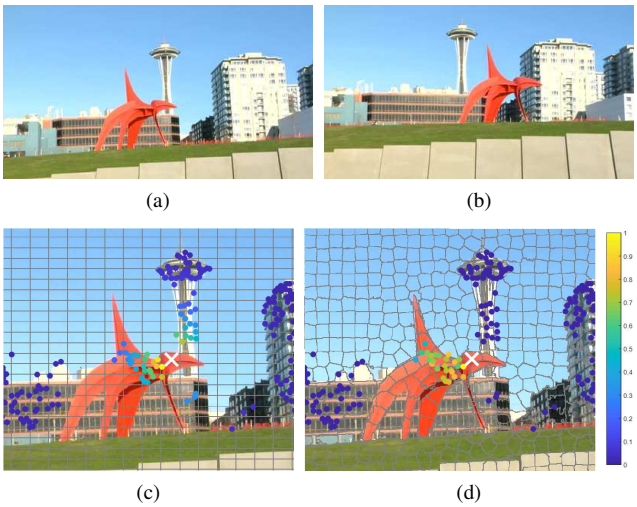


Figure 2. Effect of warping residuals. (a) A target image $\mathcal{I}$ and (b) a reference image $\mathcal{J}$. (c) The partitioned regular grid cells and the feature points with spatial distance based weights [22] used to estimate the homography for a given cell marked by the white cross. (d) The partitioned superpixels and the feature points with the proposed warping residual based weights used to estimate the homography for a given superpixel marked by the white cross. We draw large grid cells and superpixels for visualization purpose.

We see that the warping residual vectors implicitly describe the overall scene structures, since the feature points located on a same planar region tend to have similar warping residual vectors to each other, while the points located at different scene depths yield largely different warping residual vectors from each other.

We adopt the warping residual vectors for image stitching with large parallax. Specifically, we first partition input images into superpixels which are warped adaptively, since the parallax phenomenon usually occurs in the vicinity of the object boundaries. We estimate the optimal homography $\hat{\mathbf{h}}_i$ for the i-th superpixel $\mathcal{S}_i$ by solving the equation in (5) where the existing spatial distance based weight $w_{i,k}$ is replaced with the proposed warping residual based weight $w_{i,k}^{\text{prop}}$ given by

$$w_{i,k}^{\text{prop}} = \max(\exp(-\left\|\mathbf{R}_i - \mathbf{r}_k\right\|^2 / \sigma^2), \gamma), \tag{10}$$

where $\mathbf{R}_i$ is the warping residual vector of the i-th superpixel $\mathcal{S}_i$ which is defined by the warping residual vector of the nearest feature point $\mathbf{x}_k$ to the center of $\mathcal{S}_i$. We empiricaly set $\sigma = 4$ and $\gamma = 0.01$. Note that each superpixel is warped adaptively by imposing high weights to the feature points located on similar planar regions to the target superpixel.

Figure 2 shows the effect of the warping residual based weighting scheme, where two input images in Figures 2(a) and (b) exhibit large parallax, since the red object and the buildings have quite different relative positions from each view. Figure 2(c) visualizes the partitioned regular grid cells and the detected feature points obtained by APAP [22]. We compute the spatial distance based weights in (7) for the feature points with respect to a target grid cell marked by the white cross. We see that higher weights are assigned to the feature points spatially closer to the target grid cell. In particular, even though the target grid cell is located on the red object, not only the points located on the red object but also the points on the white tower are assigned high weights. On the other hand, Figure 2(d) shows the partitioned superpixels and the warping residual based weights computed in (10) for a given target superpixel marked by the white cross. We observe that the proposed algorithm assigns high weights to only the feature points located on the same red object to the target superpixel, while effectively suppressing the weights for the spatially close feature points located on the white tower at a different scene depth from the red object. Also, note that large parallax usually occurs in the vicinity of the object boundaries which are effectively captured by the boundaries of superpixels.

4. Image Warping

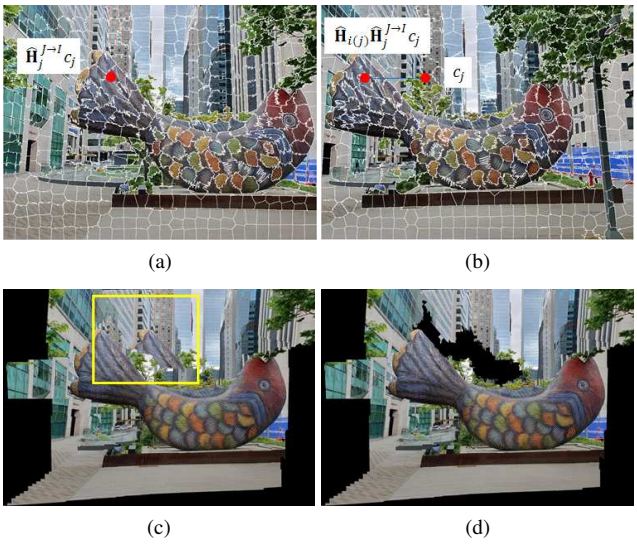
We first estimate multiple valid homographies between two images, which are used to compute the warping residuals. We partition input images into superpixels and estimate an optimal homography at each superpixel based on the warping residuals. We also refine the initially estimated homography to handle the occlusion for more reliable image alignment.

4.1. Multi-Homography Estimation

We find feature points by using SIFT [17] to obtain a set of initial matches $\mathbf{F}_{\text{init}}$ between two images $\mathcal{I}$ and $\mathcal{J}$. Then we use $\mathbf{F}_{\text{init}}$ to estimate M multiple homographies $\{\mathbf{H}^m\}_{m=1}^M$ and the associated inlier matches $\{\mathbf{F}_{\text{inlier}}^m\}_{m=1}^M$. We initialize the set $\mathbf{F}_{\text{cand}}$ of the candidate matches as $\mathbf{F}_{\text{init}}$, and estimate the first homography $\mathbf{H}^1$ with its inlier set $\mathbf{F}_{\text{inlier}}^1$ from $\mathbf{F}_{\text{cand}}$ using the outlier removal method of MULTI-GS with a threshold of $\eta = 0.01$ [5]. For stable homography estimation, we also use the normalized coordinates of feature points [25], and remove the matches of isolated feature points from $\mathbf{F}_{\text{inlier}}^1$ which have no neighboring points within 50 pixel distance. We subtract $\mathbf{F}_{\text{inlier}}^1$ from $\mathbf{F}_{\text{cand}}$, and continue to estimate the next homography $\mathbf{H}^2$ and its inlier set $\mathbf{F}_{\text{inlier}}^2$ from the updated $\mathbf{F}_{\text{cand}}$. This process is iteratively repeated up to 5 times until satisfying the stopping condition, $\|\mathbf{F}_{\text{inlier}}^m\| < 8$ or $\|\mathbf{F}_{\text{cand}}\| \,/\, \|\mathbf{F}_{\text{init}}\| < 0.02$. In addition, when only a single homography is estimated as valid during the iteration, we regard the input images exhibit small parallax and estimate $\mathbf{H}^1$ and $\mathbf{F}_{\text{inlier}}^1$ again using a more relaxed inlier threshold of $\eta = 0.1$. Finally, we combine the obtained sets $\{\mathbf{F}_{\text{inlier}}^m\}_{m=1}^M$ into the

(a) (b) (c) (d)

Figure 3. Image warping by multiple homographies. (a) Input images and (b∼d) the results of image stitching where the target image is warped by the estimated three homographies, respectively.



(a) (b) (c) (d)

Figure 4. (a) A target image and (b) a reference image with large parallax. The warped target images onto the reference image domain according to (c) the initially estimated homography and (d) the refined homography, respectively.

set of total inlier matches $\mathbf{F}_{\text{inlier}}$. Figure 3 shows a pair of input images with large parallax and their stitching results obtained by warping the target image onto the reference image according to the three estimated homographies, respectively. We see that each of the multiple homographies aligns a certain image region only, e.g., the first homography aligns the white building, the second homography aligns the ground plane, and the third homography aligns the blue object, respectively.

4.2. Optimal Warping Estimation

As explained in Section 3, we partition an input image $\mathcal{I}$ into superpixels using SLIC [1], and we warp each superpixel of $\mathcal{I}$ onto $\mathcal{J}$ according to the optimal homography computed by solving (5) based on the warping residual based weights in (10). However, such forward warping may cause holes in the warped image domain, and therefore, we perform the inverse warping from $\mathcal{J}$ to $\mathcal{I}$ instead. Specifically, we consider a large canvas on $\mathcal{J}$ domain which includes $\mathcal{J}$ and the warped image of $\mathcal{I}$. We partition the canvas into superpixels where the inside of $\mathcal{J}$ is partitioned by using SLIC [1] and the outside of $\mathcal{J}$ is uniformly partitioned into superpixels of 100×100 regular grids. For each j-th superpixel $S_j^{\mathcal{J}}$ in the canvas, we estimate the initial homography $\hat{\mathbf{H}}_j^{\mathcal{J}\to\mathcal{I}}$ based on the equation (5) similarly to the homography estimation for the forward warping. Then we take the pixels in $\mathcal{I}$ corresponding to $S_j^{\mathcal{J}}$ according to $\hat{\mathbf{H}}_j^{\mathcal{J}\to\mathcal{I}}$, and generate a hole-free warped image of $\mathcal{I}$ on the canvas domain.

Figures 4(a) and (b) show two images $\mathcal{I}$ and $\mathcal{J}$, respectively, and Figure 4(c) shows the warped image of $\mathcal{I}$. We see that most of the pixels are warped accurately, however, the tail of the fish statue appears twice due to the occlusion as depicted in the yellow box. To handle this occlusion artifact, we check whether the superpixels in the canvas domain, whose corresponding pixels lie within $\mathcal{I}$, are occluded in $\mathcal{I}$ or not. In practice, we first compute a warping loss for $S_j^{\mathcal{J}}$ given by

$$L(S_j^{\mathcal{J}}) = \frac{1}{|S_j^{\mathcal{J}}|} \sum_{\mathbf{q}\in S_j^{\mathcal{J}}} \left\| \mathcal{J}(\mathbf{q}) - \mathcal{I}(\hat{\mathbf{H}}_j^{\mathcal{J}\to\mathcal{I}}\mathbf{q}) \right\| \tag{11}$$

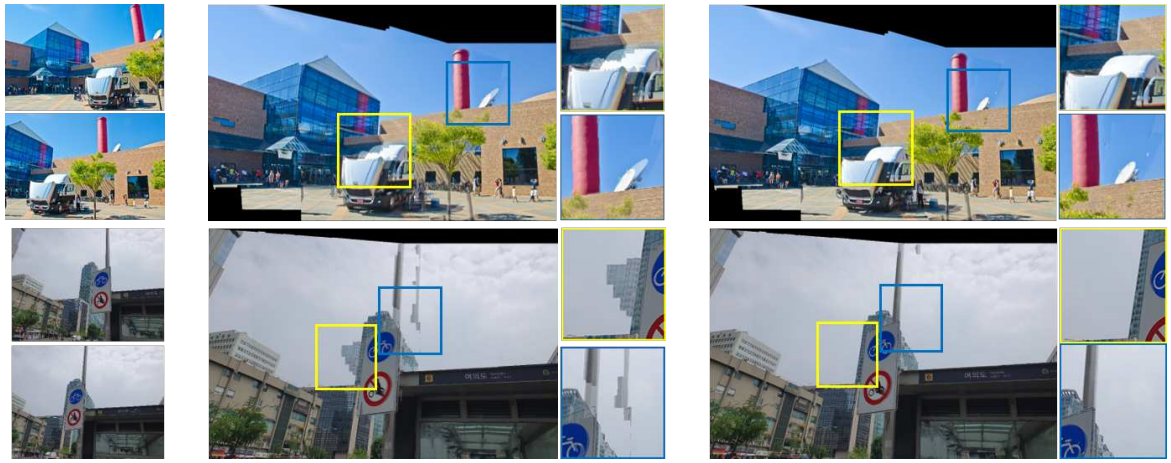
where large warping losses are usually evaluated at the superpixels occluded in $\mathcal{I}$. We also compute a bidirectional warping distance for $S_j^{\mathcal{J}}$ as

$$d(\mathbf{c}_j) = \|\mathbf{c}_j - \hat{\mathbf{c}}_j\| \tag{12}$$

where $\mathbf{c}_j$ is the center pixel of $S_j^{\mathcal{J}}$ and $\hat{\mathbf{c}}_j = \hat{\mathbf{H}}_i\hat{\mathbf{H}}_j^{\mathcal{J}\to\mathcal{I}}\mathbf{c}_j$ where $\hat{\mathbf{H}}_i$ denotes the homography of the forward warping at the i-th superpixel in $\mathcal{I}$ including the pixel of $\hat{\mathbf{H}}_j^{\mathcal{J}\to\mathcal{I}}\mathbf{c}_j$. As marked by the red points in Figure 4(b), $\mathbf{c}_j$ and its bidirectionally warped point have a large distance in the warped image since $\mathbf{c}_j$ is occluded in $\mathcal{I}$. We empirically detect $S_j^{\mathcal{J}}$ is occluded when $L(S_j^{\mathcal{J}}) > 20$ and $d(\mathbf{c}_j)$ is larger than 2%

(a) (b) (c) (d)

Figure 5. Effects of the warping residual based weighting and the occlusion handling in the proposed algorithm. (a) Input images with large parallax. The stitched images obtained by using (b) APAP [22], (c) an extended APAP by replacing the spatial distance based weight with the proposed warping residual based one, and (d) the proposed algorithm with occlusion handling and homography refinement.



(a) (b) (c)

Figure 6. Image stitching results of the proposed algorithm. (a) Input images with large parallax. The stitched images obtained by using (b) the initial homography and (c) the refined homography, respectively. "Truck (top)" and "Subway (bottom)" images.

of the diagonal length of $\mathcal{J}$. In addition, we perform the connected component analysis on the detected superpixels and select $M-1$ largest connected regions as occluded regions. Note that M is the number of estimated multiple homographies. We observe that, in most cases, occlusions mainly occur between the foreground objects and the background, and thus we simplify the number of occluded regions as $M-1$. Figure 4(d) illustrates the warped image of $\mathcal{I}$ on the canvas domain of $\mathcal{J}$, where the regions of occlusion are represented as holes.

Note that the warping residual vector of a superpixel is defined by that of the feature point nearest to the superpixel, which may cause the alignment error when they are not on the same object. Therefore, we obtain an increased number of feature points by adjusting the parameters when implementing SIFT. We also additionally refine the estimated homography $\hat{\mathbf{H}}_j^{\mathcal{J}\to\mathcal{I}}$ for $S_j^{\mathcal{J}}$ to the homography of the neighboring superpixel which has the minimum warping loss in (11) among 100 nearest superpixels to $S_j^{\mathcal{J}}$. We apply the refinement method to non-occluded superpixels only, since no valid pixel values are extracted from the occluded superpixels.

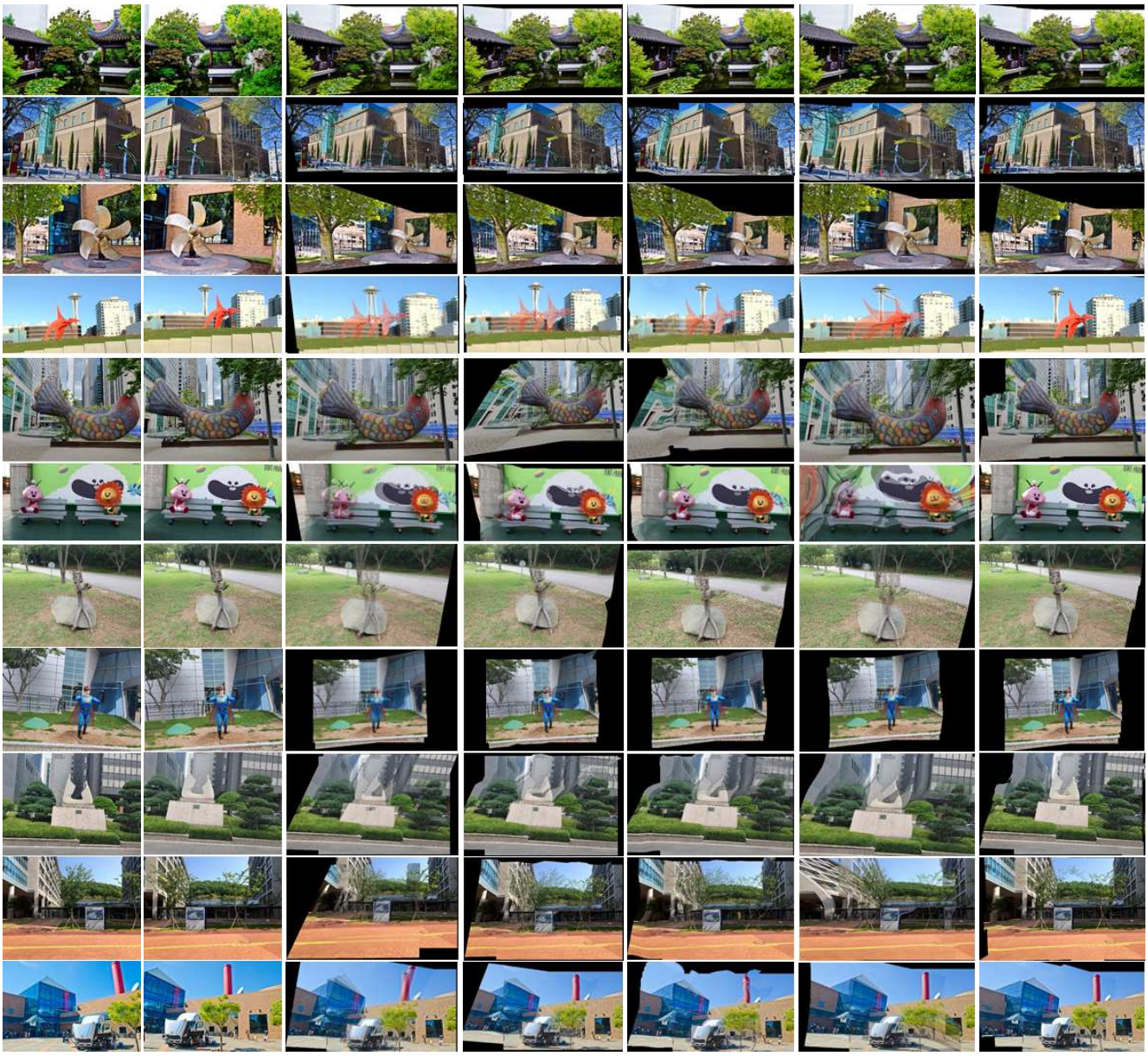
5. Experimental Results

We evaluate the performance of the proposed image stitching algorithm using 40 pairs of test images: 20 image pairs are from [23] and the other 20 pairs are newly captured. In order to compare the alignment performances effectively, each pair of images share sufficient overlapping areas and we apply the average blending scheme to combine the warped target image and the reference image.

5.1. Performance of Image Warping

Figure 5 shows ablation experimental results of the proposed algorithm. As shown in Figure 5(b), the existing method of APAP [22] yields severe parallax artifacts around the foreground object of the standing board. When we replace the spatial distance based weighting scheme of APAP with the proposed warping residual based one, most of the parallax artifactis are effectively alleviated as shown in Figure 5(c). In addition, the occlusion handling and the homography refinement further remove the artifacts as shown in Figure 5(d).

Figure 6 shows the results of image stitching obtained

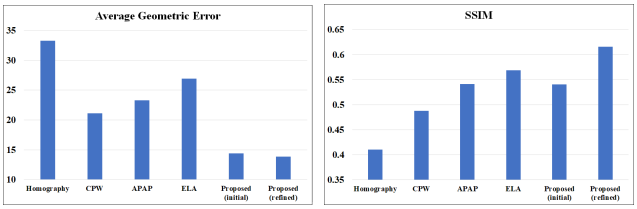


(a) $\mathcal{I}$ (b) $\mathcal{J}$ (c) Homography (d) CPW [16] (e) APAP [22] (f) ELA [11] (g) Proposed

Figure 7. Comparison of the image stitching results obtained by the proposed algorithm with that of the four existing methods: Homography, CPW [16], APAP [22], and ELA [11]. From top to bottom, "Garden," "Building," "Propeller," "Seattle," "Fish statue," "Comic characters," "Puppet," "Superman," "Company statue," "Street," and "Truck" images.

by the proposed algorithm on two challenging test sets with large parallax, which include multiple foreground objects with complex shapes at different scene depths. Figure 6(b) shows the stitching results obtained by warping images according to the initial homographies, where we see that most of the images with large parallax are well aligned without causing severe artifacts. However, some superpixels are assigned false warping residuals derived from their nearest feature points located at different scene depths, which often exhibits stitching artifacts such as ghosting in the vicinity of the object boundaries as highlighted in Figure 6(b). Figure 6(c) shows the final stitching results obtained by warping images according to the refined homographies, which accurately aligns the foreground objects at different scene depths, e.g., the truck and the satellite dish in "Truck" and the traffic sign in "Subway."

Figure 8. Examples of ground truth matches.



(a) (b)

Figure 9. Comparison of the quantitative performance of image alignment. (a) Average geometric error. (b) SSIM.

5.2. Comparison With Conventional Methods

We compare the performance of the proposed algorithm with that of the four existing algorithms: Homography, CPW [16], APAP [22], and ELA [11]. Homography is a global image alignment method that estimates a single dominant homography, and the other three methods are adaptive image alignment methods. CPW adaptively deforms the vertices of grid cells based on the energy minimization framework. APAP uses MDLT computed with the spatial distance based weighting scheme to estimate the homographies for the grid cells. ELA adaptively aligns the grid cells by approximating the warping error analytically. Note that we do not compare the seam-cutting based methods [8, 14, 23] since they try to hide the misalignment artifacts, and we also do not compare the shape-preserving warping based methods [3, 4, 12, 13] since they mainly focus on aligning non-overlapping areas. We implemented Homography and CPW, and we obtained the stitching results of APAP[3] and ELA[4] using the source codes provided by the authors. Note that we used MULTI-GS [5] to Homography and CPW for more reliable outlier removal.

Fig. 7 compares the image stitching results. "Garden" and "Building" images exhibit relatively small parallax, and thus most of the methods provide reliable stitching results. However, on the other test images with large parallax, the conventional algorithms usually fail to align both of the foreground objects and the background simultaneously since they are designed based on the assumption that the neighboring pixels in the target image should be neighboring pixels in the warped image domain as well. Specifically, the conventional methods yield severe ghosting artifacts in the vicinity of the object boundaries, e.g., the red object in "Seattle," the characters in "Comic characters," and the wooden object in "Puppet." On the contrary, while the foreground object in "Fish statue" is well-aligned by the conventional methods, the stitched background images yield severe artifacts. Also, both of the foreground objects and the background in "Company statue" and "Street" are misaligned by the conventional methods. On the other hand, the proposed algorithm adaptively estimates optimal homographies at local image regions with similar scene depths according to the warping residuals, and aligns both of the foreground objects and the background reliably while alleviating the parallax artifacts in stitched images successfully. For example, the proposed algorithm accurately aligns the foreground objects, e.g., the red object in 'Seattle," the characters in "Comic characters," and the truck in "Truck," and also aligns the background regions faithfully, e.g., the background buildings in "Fish statue" and "Street."

In addition, we evaluate the quantitative performance of image alignment using the ground truth feature matches for the 40 pairs of test images. Specifically, we distributed query pixels regularly on $\mathcal{I}$ and used the dense feature descriptor DAISY [20] to obtain the matched pixels in $\mathcal{J}$ which are then refined manually, as shown in Figure 8. We have about 90 ground truth matches on average per each image pair. We measure the difference between the warped query pixels and their ground truth matching pixels on $\mathcal{J}$ domain in terms of the average geometric error and SSIM [21] as shown in Figures 9(a) and (b), respectively. Homography, CPW, APAP, and ELA yield the average geometric errors of 33.3, 21.1, 23.3, and 26.9, and the average SSIM of 0.41, 0.49, 0.54, and 0.57, respectively. However, the proposed algorithm achieves the minimum error of 13.9 and the highest SSIM score of 0.62.

6. Conclusions

We proposed a warping residual based image stitching algorithm robust to the large parallax. We partitioned input images into superpixels and warped each superpixel according to an optimal homography which minimizes the warping error of the feature matches. We introduced the warping residual which adaptively assigns high weights to the feature points located at local image regions with similar depths to the target superpixel. Experimental results demonstrated that the proposed algorithm accurately aligns both of the foreground objects as well as the background on challenging test images with large parallax, and yields a significantly better performance compared with the conventional image stitching methods.

Acknowledgements

This work was supported by the NRF of Korea through the Ministry of Science and ICT (MSIT) under Grant 2017R1A2B4011970, and the IITP through MSIT under Grant 20170006670021001.

[3] http://cs.adelaide.edu.au/~tjchin/apap/

[4] https://ieeexplore.ieee.org/document/8119833/algorithms

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11):2274–2282, Nov. 2012. 5

[2] M. Brown and D. G Lowe. Automatic panoramic image stitching using invariant features. *Int'l J. Comput. Vis.*, 74(1):59–73, 2007. 1

[3] C.-H. Chang, Y. Sato, and Y.-Y. Chuang. Shape-preserving half-projective warps for image stitching. In *Proc. CVPR*, 2014. 2, 8

[4] Y.-S. Chen and Y.-Y. Chuang. Natural image stitching with the global similarity prior. In *Proc. ECCV*, 2016. 2, 8

[5] T.-J. Chin, J. Yu, and D. Suter. Accelerated hypothesis generation for multistructure data via preference analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(4):625–638, Apr. 2012. 4, 8

[6] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *ACM Comm.*, 24(6):381–395, 1981. 2

[7] J. Gao, S. J. Kim, and M. S Brown. Constructing image panoramas using dual-homography warping. In *Proc. CVPR*, 2011. 1, 2

[8] J. Gao, Y. Li, T.-J. Chin, and M. S Brown. Seam-driven image stitching. In *Proc. Eurographics*, 2013. 1, 2, 8

[9] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003. 1

[10] K.-Y. Lee and J.-Y. Sim. Stitching for multi-view videos with large parallax based on adaptive pixel warping. *IEEE Access*, 6:26904–26917, May 2018. 1, 2

[11] J. Li, Z. Wang, S. Lai, Y. Zhai, and M. Zhang. Parallax-tolerant image stitching based on robust elastic warping. *IEEE Trans. Multimedia*, 20(7):1672–1687, July 2018. 1, 2, 7, 8

[12] N. Li, Y. Xu, and C. Wang. Quasi-homography warps in image stitching. *IEEE Trans. Multimedia*, 20(6):1365–1375, June 2018. 2, 8

[13] C.-C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin. Adaptive as-natural-as-possible image stitching. In *Proc. CVPR*, 2015. 2, 8

[14] K. Lin, N. Jiang, L.-F. Cheong, M. Do, and J. Lu. Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In *Proc. ECCV*, 2016. 1, 2, 8

[15] W.-Y. Lin, S. Liu, Y. Matsushita, T.-T. Ng, and L.-F. Cheong. Smoothly varying affine stitching. In *Proc. CVPR*, 2011. 1, 2

[16] F. Liu, M. Gleicher, H. Jin, and A. Agarwala. Content-preserving warps for 3d video stabilization. *ACM Trans. Graph.*, 28(3):44, 2009. 2, 7, 8

[17] D. G Lowe. Distinctive image features from scale-invariant keypoints. *Int'l J. Comput. Vis.*, 60(2):91–110, 2004. 4

[18] R. Sprengel, K. Rohr, and H. S. Stiehl. Thin-plate spline approximation for image registration. In *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc. Bridging Disciplines Biomed.*, Oct. 1996. 2

[19] R. Szeliski et al. Image alignment and stitching: A tutorial. *Found. Trends Comput. Graph. Vis.*, 2(1):1–104, 2007. 1

[20] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(5):815–830, May 2010. 8

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, Apr. 2004. 8

[22] 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.*, 36(7):1285–1298, July 2014. 1, 2, 3, 4, 6, 7, 8

[23] F. Zhang and F. Liu. Parallax-tolerant image stitching. In *Proc. CVPR*, 2014. 1, 2, 6, 8

[24] G. Zhang, Y. He, W. Chen, J. Jia, and H. Bao. Multi-viewpoint panorama construction with wide-baseline images. *IEEE Trans. Image Process.*, 25(7):3099–3111, July 2016. 1, 2

[25] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Image Vis. Comput.*, 15(1):59–76, 1997. 3, 4