Regional Linear Warping for Image Stitching with Dominant Edge Extraction

Jisung Yoo[1], Sung Soo Hwang[1], Seong Dae Kim[1], Myung Seok Ki[2] and Jihun Cha[2]

[1] Department of Electrical Engineering,
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, Republic of Korea
[e-mail: yjs1018@kaist.ac.kr]
[2] Broadcasting & Telecommunications Convergence Media Research Department,
Electronics and Telecommunications Research Institute (ETRI),
Daejeon, Republic of Korea
[e-mail: serdong@etri.re.kr]
*Corresponding author: Jisung Yoo

Received November 7, 2012; revised January 7, 2013; accepted January 25, 2013; published October 29, 2013

Abstract

Image stitching techniques produce an image with a wide field-of-view by aligning multiple images with a narrow field-of-view. While conventional algorithms successfully stitch images with a small parallax, structure misalignment may occur when input images contain a large parallax. This paper presents an image stitching algorithm that aligns images with a large parallax by regional linear warping. To this end, input images are first approximated as multiple planar surfaces, and different linear warping is applied to each planar surface. For approximating input images as multiple planar surfaces, the concept of dominant edges is introduced. Dominant edges are defined as conspicuous edges of lines in input images, and extracted dominant edges identify the boundaries of each planar surface. Dominant edge extraction is conducted by detecting distinct changes of local characteristics around strong edge pixels. Experimental results show that the proposed algorithm successfully stitches images with a large parallax without structure misalignment.

Keywords: Image stitching, dominant edge extraction, local histogram of intensity

This work was supported by the IT R&D This work was supported by the IT R&D program of MKE/ETRI (13ZR1110, HCI-based UHD Panorama Technology Development)

http://dx.doi.org/10.3837/tiis.2013.10.008

1. Introduction

Generating images with a wide field-of-view (FoV) is one of the key issues in computer vision society. In order to generate these images, various approaches have been developed, and these approaches can be categorized by the omnidirectional camera-based technique and the image stitching-based technique. Because an omnidirectional camera captures a wide FoV image on a single fixed image sensor, degradation of resolution of the captured image is unavoidable. On the other hand, the image stitching-based technique provides wide FoV images without degrading their resolution, and hence, image stitching methods are widely used.

Image stitching algorithms generally consist of global alignment and post-processing. Global alignment brings input images into the same coordinate, i.e., on a specific projection surface, generally using a single homography transformation. In the result of the global alignment, misalignments of input images frequently occur, as the single homography transformation is a linear warping without consideration of parallax effects. These undesirable artifacts are removed in the post processing stage to produce a fine stitching result.

While conventional image stitching algorithms work well for images with a small parallax, research on stitching images with a large parallax is necessary. Previous post-processing methods mostly focused on removing misalignments caused by a small parallax [1][2][3][4][5]. Obtaining small parallax images, however, is difficult: to obtain small parallax images, either input images must be captured with a very small translation of the camera center or all the content of the input images should be far away from the camera so that they can be approximated as a single planar surface. When images are captured with a large parallax, it is hard to alleviate structure misalignments between those images using conventional algorithms, as shown in **Fig. 1**.

Several depth-based nonlinear warping algorithms have been proposed to overcome the problems caused by linear warping of large parallax images [6][7]. These algorithms estimate depth information of the input images, and conduct image stitching considering the estimated depth values. They are kinds of nonlinear warping because each input image is warped pixel-wisely according to the depth value of each pixel. Hence, they require semi-dense depth estimation, which is computationally complex, and input images should contain plentiful textures and a large overlap between adjacent images.

The multi-homography warping method could be the solution for stitching images with a large parallax. Many outdoor scenes captured for image stitching could be approximated as multiple planar surfaces. The pixels included on the same planar surface could be warped by a single homography without sturucture misalignments, and hence, the multi-homography warping can align the input images a large parallax without sturucture misalignments. In [8], a fruitful opportunity was found that the multi-homography warping can be effective to stitch images with a large parallax. This algorithm, however, has a highly restrictive scene-type constraint such that the upper region of an image should consist of a scene on a distant plane while the lower region of an image should consist of a scene on a ground plane, i.e., these two planes should be connected in 3D space.

In this paper, we propose a regional linear warping-based image stitching algorithm that is robust to a large parallax. Our algorithm first approximates input images as multiple planar surfaces and then applies different linear warping to each planar surface. Unlike the depth-based nonlinear warping methods, our algorithm does not require either input images

Fig. 1. Example of structure misalignments caused from parallax effects

to have a large overlap or semi-dense depth map estimation. Unlike the previous regional linear warping methods, the proposed method is more free from the scene-type constraint.

Image stitching using multiple homographies requires modeling of the contents in input images into several planar surfaces. In our algorithm, the scene modeling procedure is performed by segmenting an image into several regions. The region segmentation is conducted by extracting *dominant edges*: dominant edges are defined as conspicuous edges or lines in the input images. Since regions, which belong to different planar surfaces in 3D space, in the input images tend to generate strong edges at their boundaries, region segmentation can be conducted by extracting the dominant edges. The extraction of dominant edges is performed by considering edge information and similarity of the local histogram around the edge pixels. Each region divided by dominant edges is then warped respectively using different homographies. As most of the misalignments detected by the human vision system are discontinuities of edges or lines [3][9], our algorithm produces a fine stitching result by preserving the dominant edges.

This paper is organized as follows: The proposed method is described in Section II. Experimental results in Section III show that our algorithm successfully alleviates structure misalignments from parallax effects. Lastly, we conclude in Section IV.

2. Proposed Method

This section describes the proposed regional linear warping method for image stitching. Two images of the same planar surface are related by a homography by definition of plane-induced homography [10]. However, when input images consist of points on multiple planar surfaces, they are not related by a single homography, and this causes critical structure misalignments when images are stitched by a homography. In our work, we approximated an image as multiple planar surfaces for the purpose of extracting the regions that could be stitched by a homography To remove structure misalignments, different homographies are applied to align two images of each planar surface.

Fig. 2 illustrates the overall structure of the proposed method. The proposed method consists of three parts: i) feature point clustering; ii) dominant edge extraction; and iii) regional linear warping of input images. In feature point clustering, feature points in the input images are first extracted and the extracted feature points are clustered according to the iterative RANSAC algorithm. In the second stage, seed points are extracted for each cluster of feature points, and the extracted seed points are then propagated to extract dominant edges.

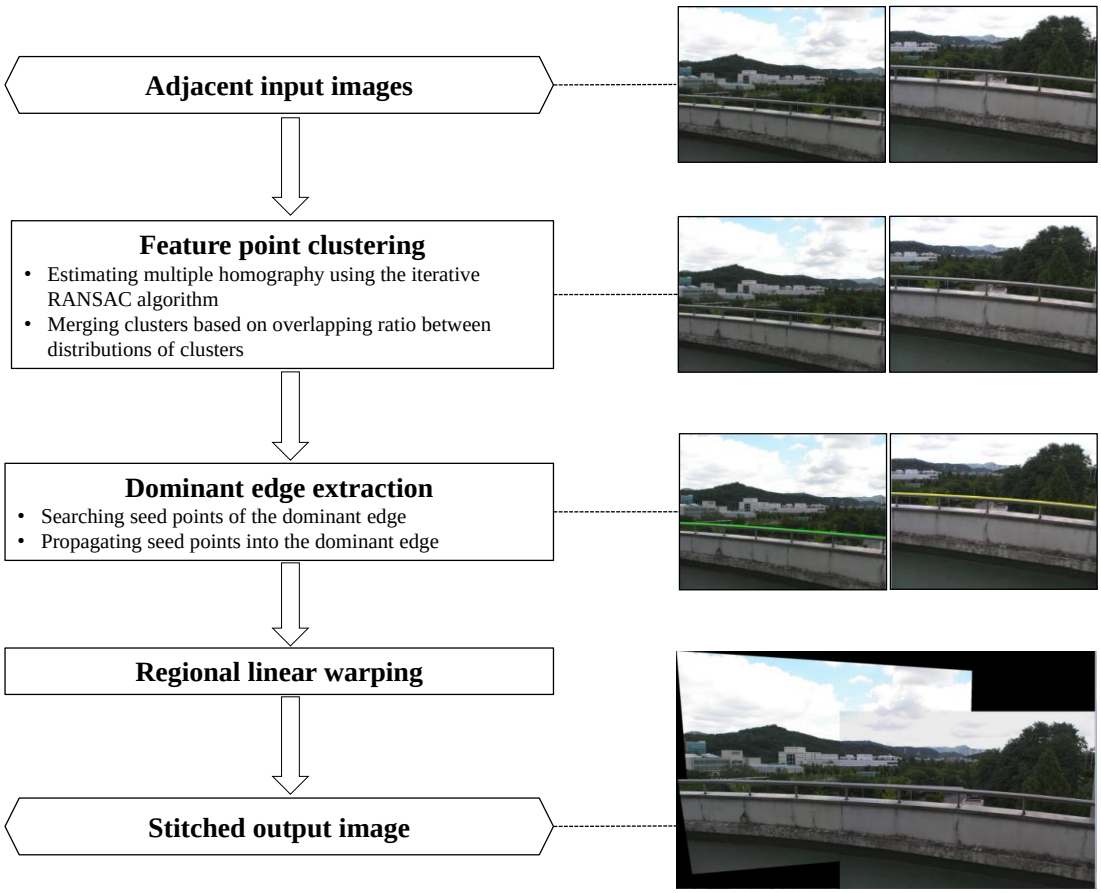


Fig. 2. Overall structure of proposed method

For extracting seed points and dominant edges, edge information and intensity distributions of the regions around the edge pixels are utilized. In regional linear warping, finally, a different homography warping is applied to each region divided by the dominant edges. We demonstrates the image stitching process for two adjacent images, i.e., an image pair, since the stitching process for multiple images is repetitions of the case of one image pair.

2.1 Feature Point Clustering

For regional linear warping using multiple homographies, the scene in input images is modeled as several planar surfaces. In our work, the scene modeling procedure is performed by segmenting an image into several regions. As the first step of the scene modeling, feature point clustering is performed to find the seed clusters of each region.

In feature point clustering, feature points in two adjacent images are extracted and matched by SIFT [11]. The matched feature points are then divided into several feature clusters by the RANSAC algorithm. The RANSAC algorithm calculates one optimal homography and selects the inliers for the homography among the remaining feature points. Hence, multiple iterations of the RANSAC algorithm compute multiple homographies and cluster feature points by assigning them into each homography as the inliers. At the first iteration, the remaining feature points are equal to the whole set of feature points extracted from two adjacent images. These feature points are classified into the inliers and the outliers

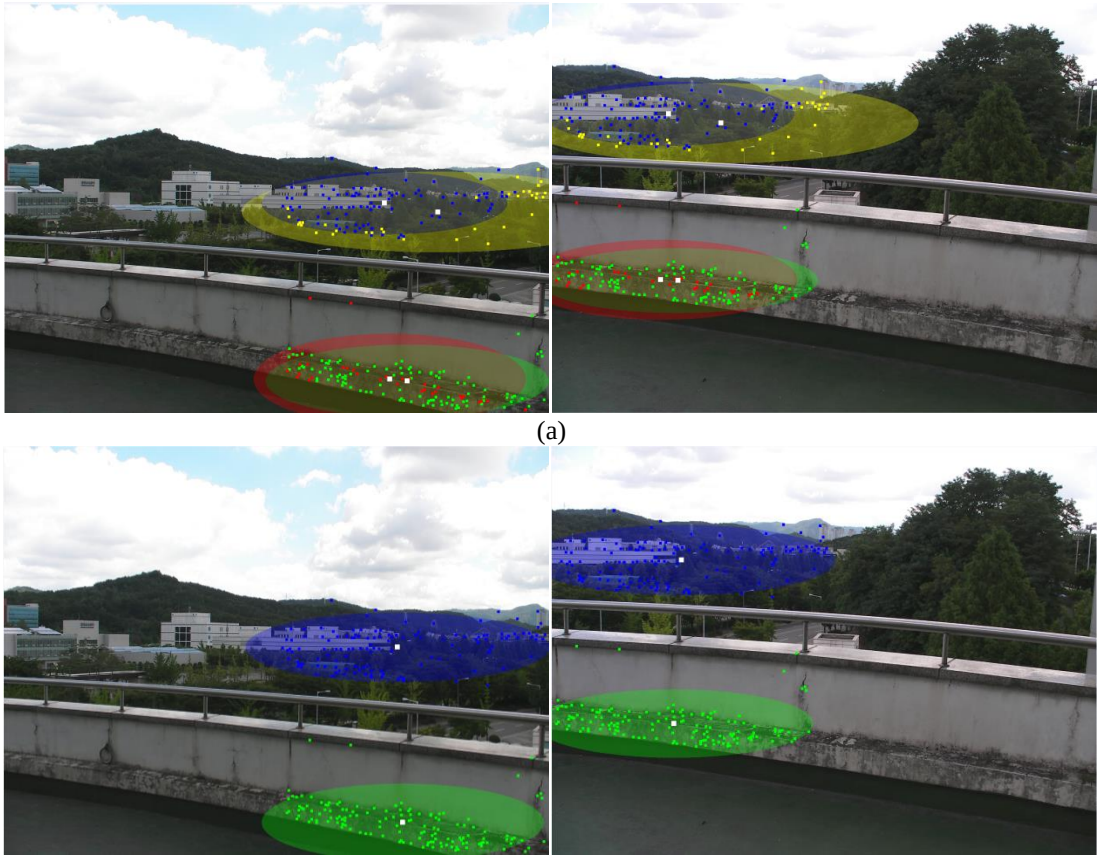


Fig. 3. (a) Modeling of feature point clusters as ellipsoidal distributions and (b) results of merging over-divided feature point clusters

for the first homography computed from the first RANSAC. These outliers are considered as the remaining feature points at the second iteration and they are divided into the inliers and the outliers for the second homography. This process is repeated until the number of remaining features is smaller than the number of a specific portion of all the feature points (in our work, this number is 20% of all the feature points or 20 feature points).

Clustered feature points from the iterative RANSAC algorithm go through an additional merging process to prevent creating an immoderate number of feature point clusters (or planar surface models). A feature point cluster is merged if its spatial distribution is highly overlapped with one of the other clusters. For the analysis of spatial distributions of feature point clusters, each feature point cluster is modeled as a 2D ellipsoidal distribution as shown in **Fig. 3 (a)**. Each ellipse model, E, is the set of pixels that satisfy the following inequality as

$$E = \{(x, y) \mid \frac{(x-\bar{x})^2}{(2\sigma_x)^2} + \frac{(y-\bar{y})^2}{(2\sigma_y)^2} \leq 1\} \tag{1}$$

where (x, y) is a pixel position on the image, $(\bar{x}, \bar{y})$ is the mean position of feature points belonging to each cluster, σ_x and σ_y are the standard deviation for x and y positions of feature points belonging to each cluster respectively. In order to find over-divided clusters we calculate overlapping ratios, ρ, between different ellipse models as

$$\rho(S_1, S_2, E_1, E_2) = \left(\frac{N(S_1, E_1 \cap E_2)}{N(S_1)} \right) \times \left(\frac{N(S_2, E_1 \cap E_2)}{N(S_2)} \right) \tag{2}$$

where S_i is the set of feature points in each cluster, E_i is the ellipse model of each cluster, $N(S_i)$ is the number of feature points that belongs to S_i, and $N(S_i, E_i \cap E_j)$ is the number of feature points that belongs to S_i in the overlapping region between E_i and E_j. If the overlapping ratio between two adjacent ellipse models exceeds the threshold, we regard two feature point clusters corresponding to these ellipse models as the same cluster, i.e., feature points in these two clusters belong to the same planar surface. If different two clusters are merged into the same cluster, the optimal homography are then newly computed for this new cluster. The results of the iterative RANSAC algorithm and additional merging process are shown in **Fig. 3**.

2.2. Dominant Edge Extraction

In Section 1, dominant edges are defined as conspicuous edges or lines in input images. We also suggest that those edges tend to exist on boundaries between regions that belong to different planar surfaces in the 3D space. Since these regions have different regional characteristics, strong edge components and distinct change of local characteristics usually exist at the dominant edges. Hence, dominant edges can be extracted by detecting strong edge components and distinct change of local characteristics around the edges of the components. The dominant edge extraction consists of two parts: 1) searching seed points of dominant edges in the overlapping FoV between adjacent images; and 2) propagating the dominant edges from each seed point to the adjacent images.

To detect distinct changes of local characteristics, we compare the characteristics of local regions by their intensity distributions. Intensity distribution is a good descriptor to compare characteristics of local regions, as it is usually insensitive to small translations or rotations: if a descriptor has the spatial information of the region, it may be too sensitive to small translations or rotations of the texture in the region. Similarity measurement of intensity distributions between local regions is conducted using their local histogram of intensity (LHI). To measure the similarity of histograms, each local histogram is first normalized such that the sum of the number of pixels in each histogram eqauls one. The similarity between local histograms of intensity (SLHI) is then measured by the histogram intersection ratio, s_{inter}, and the proportion of the pixels that have the same intensity value:

$$s_{inter}(H_1, H_2) = \sum_{n=1}^{N} \min(H_1(n), H_2(n)) \tag{3}$$

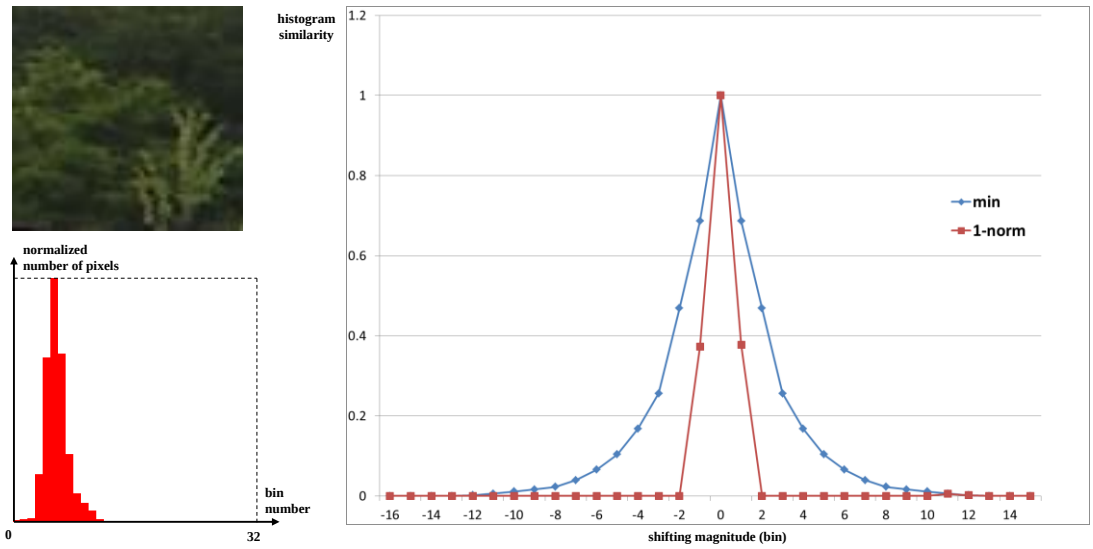


Fig. 4. Example of comparison between two histogram similarity measures, the histogram intersection ratio and the 1-norm distance-based similarity. The left top image is a local region and the left bottom image represents the intensity histogram of the local region. The right image is the graph that shows that the histogram intersection ratio is more robust to histogram shift than the 1-norm distance-based similarity.

where H_i is an intensity histogram of region i, N is the total number of the histogram bins, $H_i(n)$ is the normalized number of pixels of which the intensity belongs to the n th bin of H_i. As **Fig. 4** illustrates, the histogram intersection ratio is more robust to the histogram shift, which usually occurs from local illumination changes, than the similarity based on the widely used 1-norm distance calculated by

$$s_{1-norm}(H_1, H_2) = \left| 1 - \sum_{n=1}^{N} |H_1(n) - H_2(n)| \right| \tag{4}$$

where H_i is the intensity histogram of region i, N is the total number of the histogram bins, and $H_i(n)$ is the normalized number of pixels of which intensity belongs to the n th bin of H_i.

Using the histogram intersection ratio, two types of SLHI (intra- and inter-SLHI) are defined and they are used in detecting reliable seed points and boundary propagation respectively. Intra-SLHI represents the similarity between two sub-regions of a local region while inter-SLHI represents the similarity between two local regions of two different edge pixels. Edge pixels on the dominant edges have low intra-SLHI, as distinct changes of intensity distributions occurs in the dominant edges, and edge pixels on the dominant edges have high inter-SLHI. In order to compute two types of SLHI for an edge pixel, we set a local region around the edge pixel and divide this region into two sub-regions (upper regions and lower region) by the parallel line of the edge direction shown as **Fig. 5 (a)**. These two types of SLHI are then computed as

$$\text{intra-SLHI}(R) = s_{inter}(H_{R_U}, H_{R_L}) \tag{5}$$

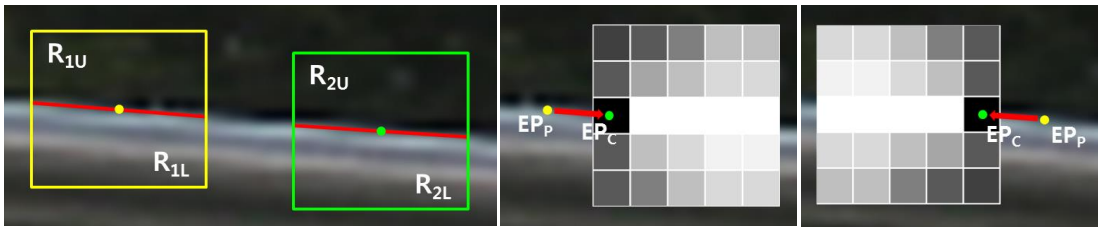
and

$$\text{inter-SLHI}(R_1, R_2) = s_{inter}(H_{R_{1U}}, H_{R_{2U}}) + s_{inter}(H_{R_{1L}}, H_{R_{2L}}) \tag{6}$$

where R_i is local regions represented in **Fig. 5 (a)** as color boxes and R_{iU} and R_{iL} represent the upper and lower regions of R_i respectively.

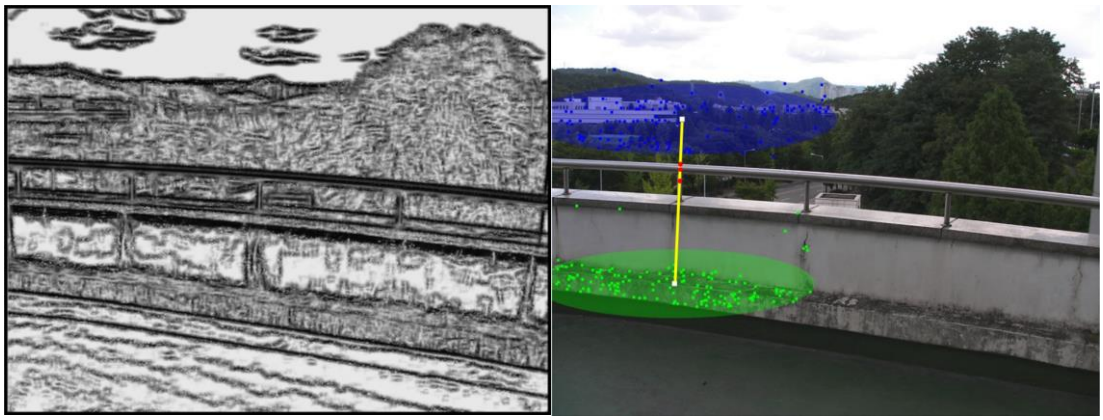
As the first step of the dominant edge extraction, we determine seed points of the dominant edge. In section 2.1, we classified feature points extracted from overlapping FoV into several clusters. The feature points in each cluster are on different planar surfaces. As mentioned before, the dominant edge is likely to exist between different planar surfaces. Hence, we search a region between two different feature point clusters for seed points of the dominant edge. In consideration of searching speed and accuracy, we make a segment that connects two mean pixels of difference clusters and find the seed points on the segment only. Among pixels on this segment, we select pixels as seed points when i) edge magnitude of the pixels is large and ii) the pixels have low intra-SLHI. Edges are extracted using the Sobel edge operator and intra-SLHI is computed using (5). **Fig. 6 (a)** shows the computed intra-SLHI map of one of the input images and **Fig. 6 (b)** shows the selected seed points.

The selected seed points are then propagated into a single dominant edge. Two different types of tracing mask are used for propagation of the seed points, as shown in **Figs. 5 (b)** and **(c)**. Considering that the difference between tangent gradients at adjacent edge pixels in the same boundary is relatively small, we compute weights according to the angle between the current and previous tracing directions and these weights are applied to each element in the tracing mask. We trace the next edge pixel that has the highest inter-SLHI with the seed point and the previously traced edge pixel using the pre-defined weighted mask.



(a) (b) (c)

Fig. 5. (a) Example of local regions for computing SLHIs and two types of weighted masks for edge propagation of (b) right direction and (c) left direction.



(a) (b)

Fig. 6. (a) The intra-SLHI map of one of the input images. The seed points for the dominant edge are selected among the pixels on the line between mean positions of feature clusters based on the intra-SLHI. This line is represented as yellow line in (b) and the selected seed points are represented as the red spots in (b).

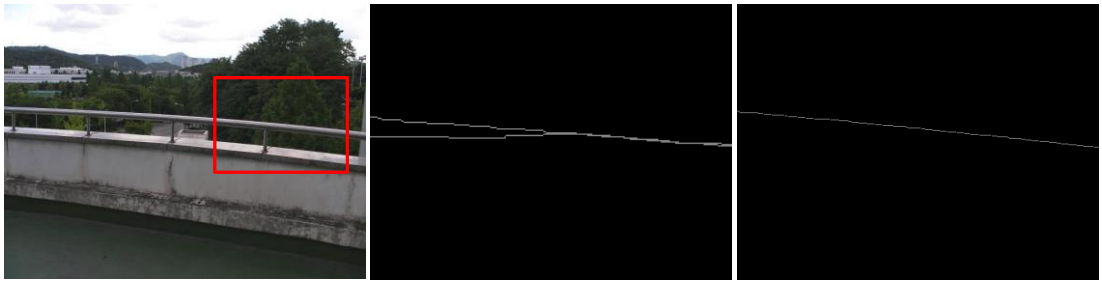


Fig. 7. Crossed dominant edge removal

Fig. 8. The final results of dominant edge extraction for input images.

There should be only one dominant edge between two adjacent regions. In the propagetion process, however, two types of ambiguities could appear. First, multiple seed points may exist between two adjacent regions. When two different dominant edges are crossed in the middle of tracing, the one, of which the total sum of inter-SLHIs is lower than

that of the other, is removed as shown in **Fig. 7**. Second, there are more than two dominant edges after the end of the tracing. In this case, we select one dominant edge only from detected candidates based on the similarity between the intensity histograms of the regions on both sides of each candidate. This similarity of a dominant edge can be computed by summing intra-SLHIs of the pixels belonging to the dominant edge. The final dominant edge is selected as the dominant edge which has the highest sum of intra-SLHIs.

2.3. Regional Linear Warping

The last step of the proposed method is regional linear warping for the segmented regions of the input images. We select one of the input images as the reference image and warp other input images into the reference image coordinates using a different homography for each region. All homographies were already estimated in the feature point clustering step in section 2.1. The process and result of regional linear warping are as shown in **Fig. 9**.

(a) (b)

Fig. 9. (a) The process and (b) the result of regional linear warping.

3. Experimental Results

The proposed method is compared with three state-of-the-art software programs, Microsoft ICE [12], Photoshop CS5 [13], and Autostitch [14]. Image stitching algorithms in these software programs use only one homography to stitch two adjacent images while our algorithm uses a different homography for each region.

Figs. 10 (a), (b), (c), (d) show total stitched results of our algorithm, Microsoft ICE, Photoshop CS5 , and Autostitch, respectively. **Figs. 10 (f), (g), (h)** highlighted by green, pink, and cyan, respectively, show structure misalignments generated from three state-of-the-art software programs, while they are removed from our results shown as **Fig. 10 (e).** These structure misalignments are caused by stitching images with a large parallax using a single homography. Hence, the proposed method that uses regional linear warping using multiple homographies outperforms other stitching methods based on warping using a single homography when the input images could be approximated as multiple planar surfaces reasonably. **Fig. 11** and **12** shows additional results of our algorithm, Microsoft ICE, Photoshop CS5, Autostitch (in **Fig. 11**) and the state-of-the-art multi-homography warping method (in **Fig. 12**).

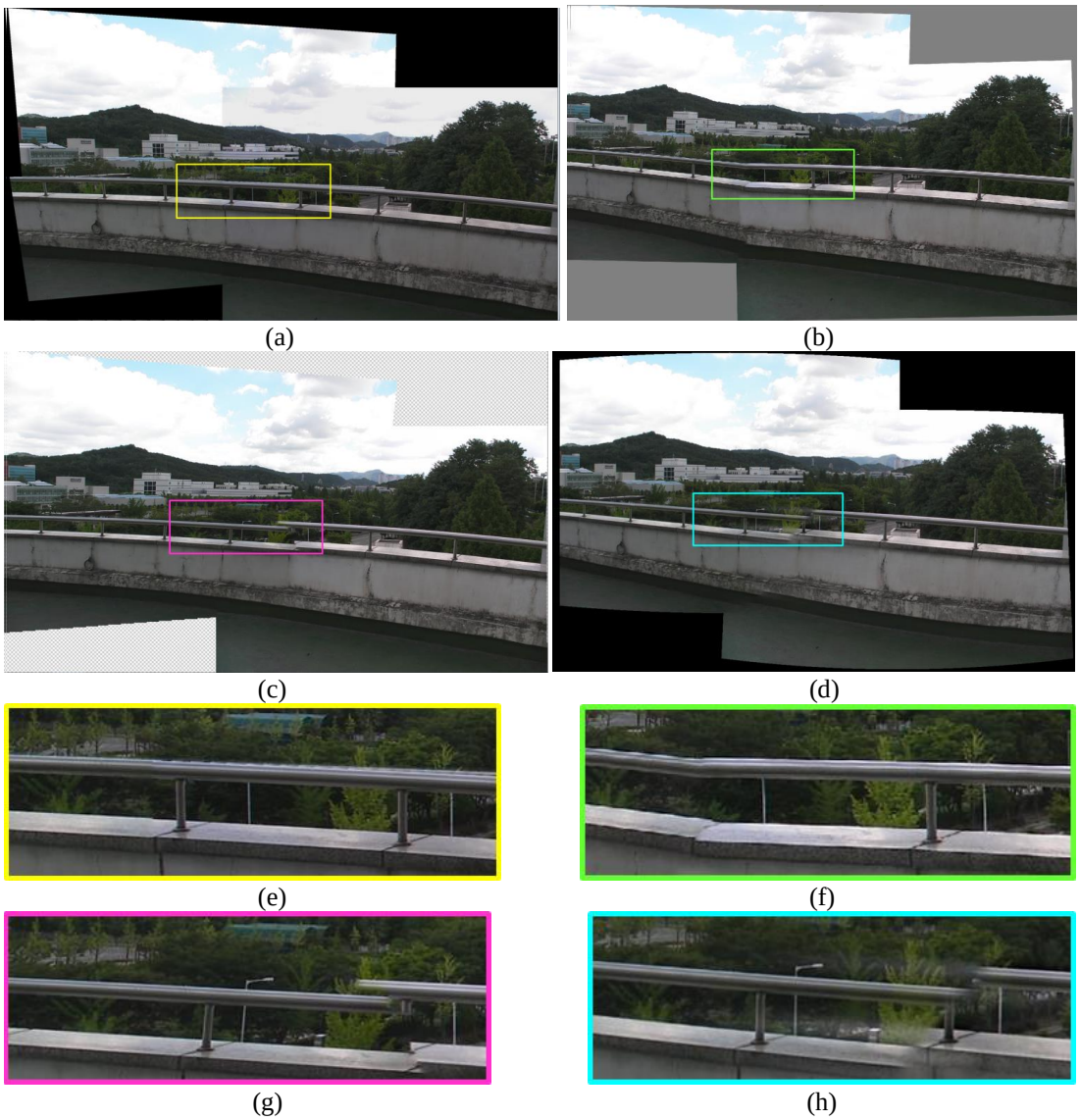
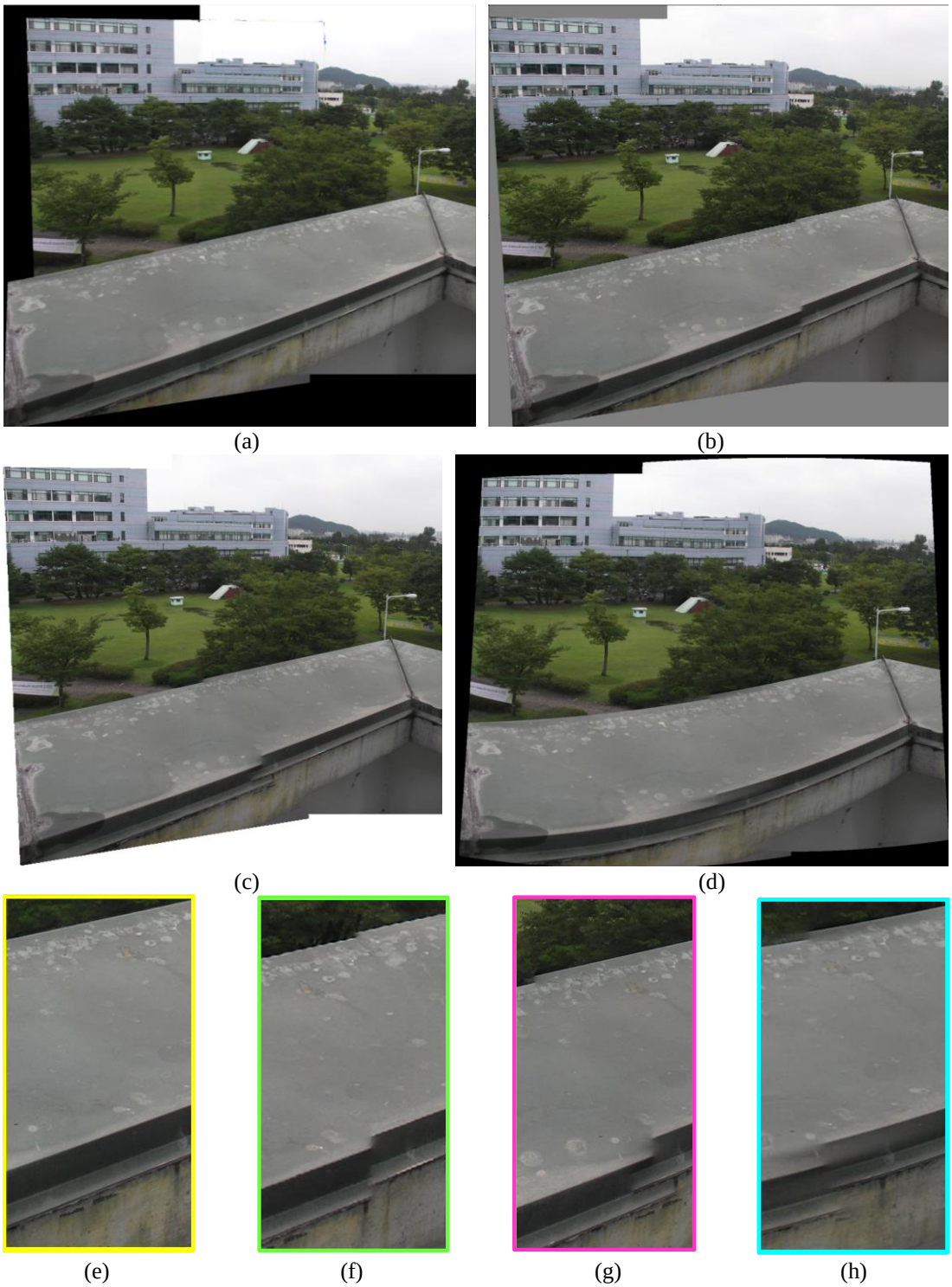


Fig. 10. Example of results of our algorithms and other state-of-the-art algorithms. **Figs. 10 (a), (b), (c), (d)** show the total stitched results of our algorithm, Microsoft ICE, Photoshop CS5 and Autostitch, respectively. **Figs. 10 (f), (g), (h)** show the structure misalignments generated from three state-of-the-art software programs while they are removed from our results shown as **Fig. 10 (e)**.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 11. Another example of the results of our algorithm and other state-of-the-art algorithms. **Figs. 11 (a), (b), (c), (d)** show the total stitched results of our algorithm, Microsoft ICE, Photoshop CS5 and Autostitch, respectively. The **Figs. 11 (f), (g), (h)** show structure misalignments generated from three state-of-the-art software programs while they are removed from our results shown as **Fig. 11 (e).**

Fig. 12. Examples of the results of the state-of-the-art multi-homography (dual-homography) warping method. The yellow and green circles show structure misalignments generated from this method while they are removed from our results shown as **Figs. 10 (a) and 11 (a).**

Fig. 13. Examples of the failure cases of the proposed method. The left and right image shows difficulty of tracing the dominant edge when intensity values of both sides of the dominant edge are similar.

Fig. 13 shows examples of the failure cases of the proposed algorithm. The proposed dominant edge extraction method uses the intenisty histogram as the similarity measure. Hence, although the regions at both sides of the dominant edge have totally different color values, in case that the regions have similar intenisty distributions, our algorithm traces improper edge pixels.

4. Conclusion

We presented an image stitching method that aligns images with a large parallax. Input images were first approximated as multiple planar surfaces, and a different homography was applied for each planar surface. To approximate the input images as multiple planar surfaces, the concept of dominant edges was introduced, and dominant edges were extracted by considering the edge magnitude and the distinct change of intensity distributions. For measuring similarity of the intensity distributions, the concept of intra- and inter-SLHI,

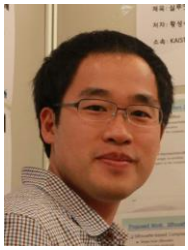
which measures similarity of the intensity histogram, was also introduced. We verified that the proposed method avoids structure misalignments that occur in conventional single-homography-based image stitching algorithms.

As structure misalignments can be alleviated by the proposed method, our work can help produce realistic panoramic images. The image capturing for panoramic images, which requires the captured images to have a small parallax, can also be improved, as the proposed method successfully aligns images with a large parallax. Lastly, as the proposed method segments input images into multiple planar surfaces that are estimated to have difference depth levels, our work can be extended to 2D-to-3D conversion of panoramic images.

References

[1] M. Brown et al., "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, 2007, pp. 59–73. Article (CrossRef Link)
[2] P. Burt et al., "A multiresolution spline with applications to image mosaics," *ACM Transactions on Graphics*, vol. 2, no. 4, 1983, pp. 217–236. Article (CrossRef Link)
[3] J. Jia and C.-K. Tang, "Image stitching using structure deformation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, 2008, pp. 617–631. Article (CrossRef Link)
[4] A. Levin, A. Zomet, S. Peleg, and Y. Weiss, "Seamless image stitching in the gradient domain," in *Proc. European Conference on Computer Vision*, 2006. 4. Article (CrossRef Link)
[5] M. Uyttendaele, et al., "Eliminating ghosting and exposure artifacts in image mosaics," in *Proc. of IEEE Computer Vision and Pattern Recognition*, 2001, pp. 509–516. Article (CrossRef Link)
[6] S.B. Kang, et al., Seamless Stitching using Multi-Perspective Plane Sweep, Technical Report MSR-TR-2004-48, Microsoft Research, 2004. Article (CrossRef Link)
[7] Q. Zhi and J.R. Cooperstock. "Toward Dynamic Image Mosaic Generation With Robustness to Parallax," *IEEE Transactions on Image Processing*, vol. 21, no. 1, Jan. 2012, pp. 366-378. Article (CrossRef Link)
[8] J. Gao, S.J. Kim, and M. S. Brown, "Constructing Image Panoramas using Dual-Homography Warping," in *Proc. of IEEE Computer Vision and Pattern Recognition*, 2011. Article (CrossRef Link)
[9] C.-K. Tang, "Eliminating structure and intensity misalignment in image stitching," in *Proc. of IEEE International Conference on Computer Vision*, 2005. Article (CrossRef Link)
[10] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd Ed., Cambridge University Press, 2003.
[11] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110. Article (CrossRef Link)
[12] http://cs.unc.edu/~ccwu/siftgpu/
[13] Microsoft Image Composite Editor, http://resear ch.microsoft.com/enus/um/redmond/groups/ivm /ICE/
[14] http://www.adobe.com/products/ photoshop.
[15] http://cvlab.epfl.ch/brown/autostitch/autostitch .html.

Jisung Yoo is currently working towards his PhD degree at the Korea Advanced Institute of Science and Technology (KAIST). His current research interests include feature extraction and matching, image registration and object detection.



Sung Soo Hwang is a PhD student at the Korea Advanced Institute of Science and Technology (KAIST). His current research interests include 3D object coding and image understanding.

Seong Dae Kim received his Dr.Ing in electrical engineering from ENSEEIHT, INPT, Toulouse, France, in 1983. Since 1984, he has been a professor in the Department of Electrical Engineering and Computer Science at the KAIST. His research interests are image processing, computer vision, pattern recognition, and image coding.

Myung Seok Ki is currently working at the Telecommunications Research Institute (ETRI) of Korea. His research interests are computer vision and interactive broadcasting system.

Jihun Cha is currently working at the Telecommunications Research Institute (ETRI) of Korea. He has served as the Team leader for Convergence Media Research Team of ETRI. His research interests include multimedia streaming, interactive broadcasting system, and feature extraction/tracking in motion pictures.