IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 29, 2019, accepted November 10, 2019, date of publication November 15, 2019,
date of current version November 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953705

Minimum Error Seam-Based Efficient Panorama Video Stitching Method Robust to Parallax

JEONHO KANG, JUNSIK KIM, INHONG LEE, AND KYUHEON KIM
Department of Electronic Engineering, Kyung Hee University, Seoul 17104, South Korea

Corresponding author: Kyuheon Kim (kyuheonkim@khu.ac.kr)

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea Government (MSIT) under Grant 2017-0-00224, Development of Generation, Distribution and Consumption Technologies of Dynamic Media Based on UHD Broadcasting Contents.

ABSTRACT From stitching video sequences point of view, an existing image stitching technique may have limitations on object movement or loss in the stitched video sequence due to the different parallax of the individual video sequences. Also, an existing image stitching technique have higher computational complexity and have matching accuracy limitations in obtaining descriptors for many factors. Therefore, this paper proposes an edge-based weight minimum error seam method in order to overcome the limitations caused by parallax and video stitching and trigonometric ratio-based image match algorithm in order to reduce computational complexity and insensitivity to rotation, illumination, image size, or resolution.

INDEX TERMS Stitching method, robust to parallax, calculation of similar space, minimal error seam.

I. INTRODUCTION

Users have become accustomed to high quality content in high-definition (HD) and demand more immersive content with higher resolution and wider viewing angles, including ultra-high-definition (UHD), more excessive UHD, and panoramic views [1]. Higher resolution and wider viewing angle content require the use of fish-eye or other large field of view (FoV) lenses, stitching several sub-images by collecting data with a rotating line camera or reflecting the captured image through rotating, spherical, conical, hyperbolic, or parabolic mirrors [2]–[4]. Thus, it is required to stitch images to generate higher resolution and wider viewing angle content due to the fact that a single conventional camera cannot produce this content [5], [6]. In general, the stitching process extracts and matches feature points out of the two images to be stitched, sets an overlapping region in those two images on the basis of the match results, and performs a blending process of synthesizing the overlapping regions [7]–[9].

Scale-Invariant Feature Transform (SIFT) [10] and Speeded Up Robust Features (SURF) [11], which are both widely used in the stitching process, show very high accuracy when stitching images at the same focal distance. However, due to the difference between cameras and parallax, an error occurs depending on the distance between each object and the camera lens, as shown in Fig. 1 (a) [12]. For example, if there is an object between the cameras and the background, Viewpoints A and B display it differently in terms of the object position and the object scene to be captured, as shown in Fig. 1 (b).

When feature points from the background in an image are extracted and matched without considering the error due to the parallax explained in Fig. 1, the same object could be located at different positions in the images to be stitched, which could cause the objects in a stitched image to disappear [13]. Alternately, when the feature points are extracted and matched with respect to the objects, parallax error causes a mismatch of the background in the two images to be stitched [13]. In terms of stitching video sequences, an existing image stitching technique would have the limitation in object movement or disappearance in a stitched video sequence due to different parallax of individual video sequences [14]. Additionally, conventional stitching algorithms have higher computational complexity due to an integral computation [13] and have matching accuracy limitations in obtaining descriptors for rotation, illumination, image size, or resolution [13].

Therefore, this paper proposes a weighted minimum error seam method in order to overcome the limitations caused by parallax and video stitching as previously explained. Additionally, a trigonometric ratio-based image

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Zhao.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

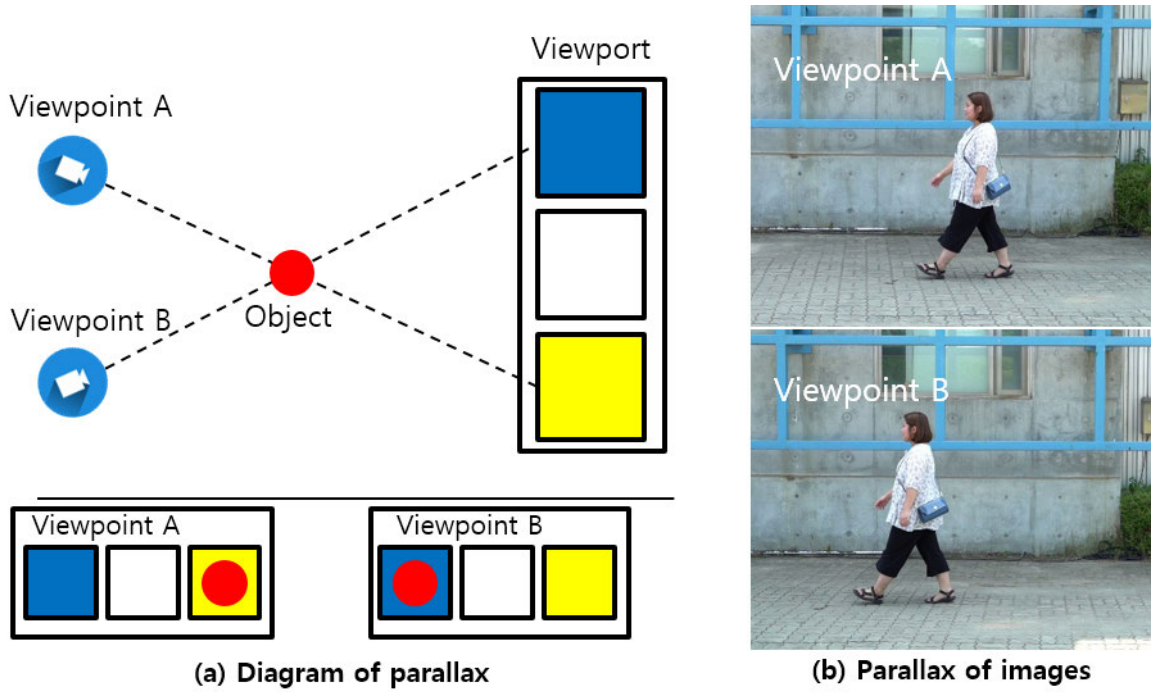


FIGURE 1. Example of parallax.

match algorithm is suggested for reducing computational complexity and insensitivity to rotation, illumination, image size, or resolution. Section II describes the feature point extraction method and the generation of a minimal error seam in the stitching region. Section III explains the proposed stitching systems, which include the technique that matches images based on the trigonometric ratio and the blending technique based on MES. In section IV, we analyze the experimental results of each proposed technique and conclude in section V.

II. RELATED WORKS

In general, the image stitching process consists of a couple of functions, such as keypoint extraction, generation of a feature descriptor, matching, and blending.

The first step in an image stitching process is keypoint extraction. A good keypoint for matching images has the following two conditions. First, it should be easily identifiable, even if the shape, size, or position of the object changes. Second, it should be a keypoint that can be easily found in an image, even if the viewpoint and lighting change. The most widely used keypoint to satisfy these conditions in an image is a corner point. This is because a corner point has rather different pixel values when compared with neighboring pixel values [15].

Generation of a feature descriptor as the second step in an image stitching process is used in SIFT [10], SURF [11], Binary Robust Independent Elementary Features (BRIEF) [16], Oriented FAST and rotated BRIEF (ORB) [17], Fast Retina Keypoint (FREAK) [18], etc. A feature descriptor describes a local image characteristic, such as a gradient, histogram, etc., which is calculated from a corresponding keypoint and its neighbor pixels. The keypoint and feature descriptor have many requirements; for example, it should have invariance in translation, rotation, scale, affine transformation, noise, and blur effects for image matching.

The third step in the image stitching process is matching. After we have the information of keypoints and feature descriptors for all images, we can use this useful information to do image matching. The matching is in general a method of comparing descriptors, but also a method of comparing using distance information of feature points. The method of using the triangle similarity is one of matching methods of using distance information of feature points, where the ratio among each side lengths a triangle is used. This is called as Side-Side-Side (SSS) similarity [19].

As the final step in image processing, a blending function is applied in order to have a seamless stitched image. There are two popular ways of blending images. One is called alpha blending, which takes the weighted average of the stitching region in two images [19]. This alpha blending works extremely well when image pixels in an overlapped region are well aligned to each other and have only a difference in the overall intensity shift. Another popular approach is Gaussian pyramid [20]. This method essentially merges the images at different frequency bands. The lower the frequency band, the more it blurs the boundary. Gaussian pyramid blurs the boundary while preserving the pixels away from the boundary. As mentioned above, the generation of feature descriptors takes an important role in stitching accuracy and

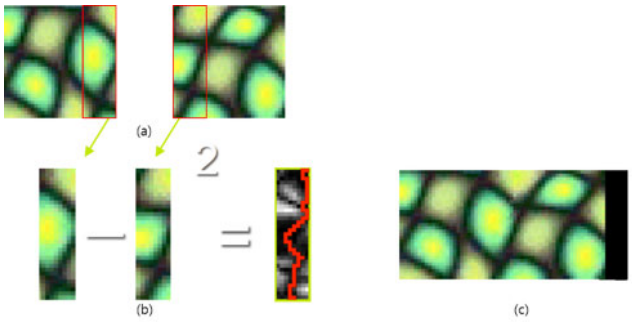


FIGURE 2. Example of minimal error seam extracting.

a huge computational burden due to keeping the invariance of many factors, which causes the stitching process to be slower. Therefore, we need a method to maintain accuracy while reducing the stitching process time with less computational effort.

Additionally, an object can exist between the background and the cameras, as shown in Fig. 1, which causes parallax to occur. Thus, there is a problem that a same object appears twice or disappears in a stitched image when the above blending technique is applied to perform a matching function based on the background. Therefore, in this paper, we use minimal error seam (MES) to find a boundary in the stitching region in order to avoid this object appearing twice or disappearing.

Fig. 2 shows an example of extracting MES. When the two images in Fig. 2 (a) are to be stitched and the red boxes in each image are considered as regions to be stitched, the MES method obtains the error by squaring the difference between the grayscale values in the overlapped region shown and produces a boundary line called MES by connecting pixels with the smallest error in Fig. 2 (b). On the basis of this boundary line, two images are stitched, as shown in Fig. 2 (c). Thus, the MES method produces a stitched image with remarkably low distortion in the stitching region [21]. However, the MES method still produces a boundary line inside an object, which causes distortion in a stitched image. This is because MES considers the difference error in terms of only pixels rather than of the object shape. Additionally, the MES method produces boundary lines for every frame when it is applied to a video stitching process. Since boundary lines are generated on the basis of two relevant frames from the two video sequences to be stitched, they can be abruptly changed from the current frame to the next one, which causes an object to disappear or movement in a stitched video sequence. Therefore, this paper proposes a stitching algorithm with an weighted minimum error seam method in order to overcome the computational burden in feature descriptor generation and the limitation of a traditional MES as described above.

III. EFFICIENT PANORAMA VIDEO STITCHING METHOD ROBUST TO PARALLAX

In this section, a minimum error seam-based efficient panorama video stitching method robust to parallax is proposed in this paper. This method consists of two functions, ''generation of location information'' and ''video blending''. The function of generating location information is to conduct a matching process by triangle keypoints classification based on triangle similarities, which will be explained in section III-A. Additionally, video blending has been achieved by the weighted minimum error seam method, which will be described in section III-B.

A. TRIANGLE KEYPOINT MATCHING SYSTEM BY USING TRIANGLE SIMILARITIES

The generation of location information in this paper is realized by the triangle keypoint matching system shown in Fig. 3. The triangle keypoint matching system consists of the following processes: keypoint extraction, keypoint grouping for a triangle, and comparing though triangle similarity.

In the keypoint extraction, this paper uses corner information as keypoints since corner information is easily identifiable, even if the shape, size, or position of the object changes, as explained in section II. The corner information is obtained by a Harris corner detector [22], which can extract keypoints at high speed and easily control the number of extracted keypoints by a threshold. Fig. 4 (a) shows the keypoints extracted by a Harris corner detector.

As explained in section II, a stitching process uses extracted keypoints and their generated descriptors, which causes a huge amount of computation. Thus, this paper proposes a method of grouping keypoints into triangles and of comparing them in terms of triangle similarities for identifying an area to be stitched. The keypoint grouping process in Fig. 3 is to cluster extracted keypoints for forming triangles. That is, an extracted keypoint is clustered with its two closest keypoints into a triangle. Thus, all extracted keypoints are included in at least one triangle of keypoints. Fig. 4 (b) shows an example of generated triangle keypoints.

As the final step for the matching system proposed in this paper, two images need to be compared in terms of similarities of triangles produced by the keypoints grouping step in order to find the positions in the two images for stitching. The comparison of the two images uses the SSS similarity as being described in section II. When the SSS similarity between two relevant side ratios is less than a threshold, two triangle keypoints are determined to be matched ones as being shown n Fig. 5.

The SSS similarity is sequentially applied to all the triangles obtained by matched triangle keypoints in the two images. On the basis of the SSS similarity result, the area to be stitched in the two images is defined with a rectangle of the smallest size containing all matched triangle keypoints. The information of the stitching region is stored as a coordinate of the top-left and bottom-right of the rectangle in an XML format for a consequent blending process, as shown in Fig. 6.

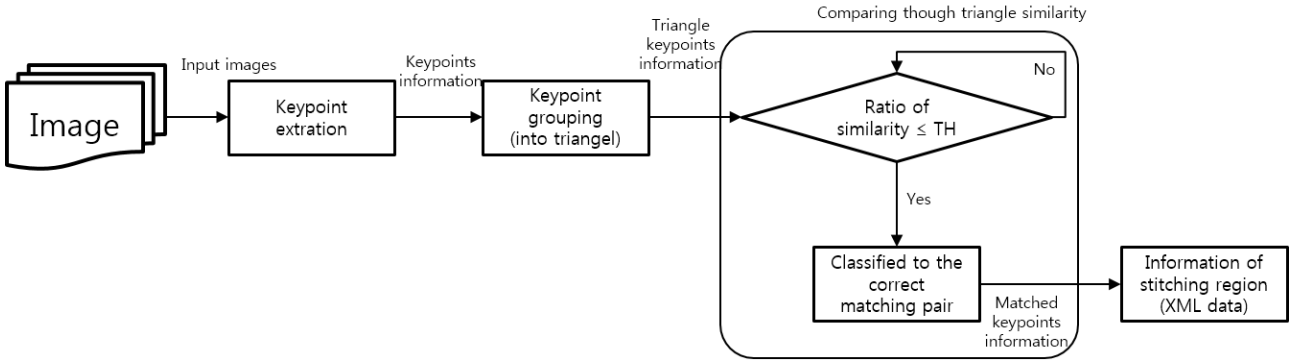
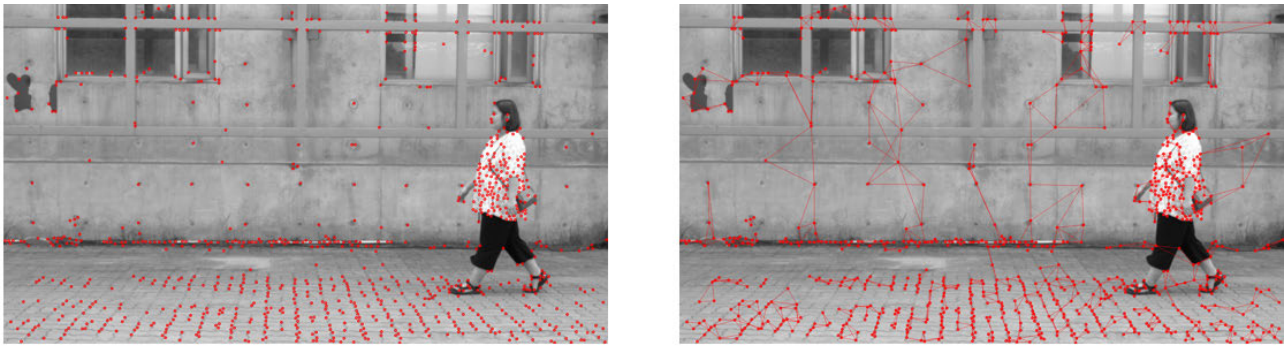


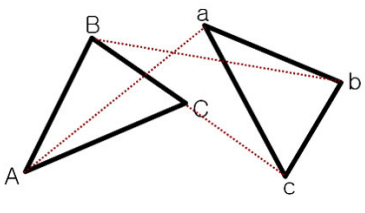
FIGURE 3. Triangle keypoint matching system.



(a) Keypoints extracted by a Harris corner detector

(b) Example of a triangle keypoints generated image

FIGURE 4. Results of extracted and triangle keypoints.



$$\left|\frac{\overline{ab}}{\overline{AB}} - \frac{\overline{bc}}{\overline{BC}}\right| \leq TH$$
$$\left|\frac{\overline{bc}}{\overline{BC}} - \frac{\overline{ca}}{\overline{CA}}\right| \leq TH$$
$$\left|\frac{\overline{ca}}{\overline{CA}} - \frac{\overline{ab}}{\overline{AB}}\right| \leq TH$$

FIGURE 5. Triangle similarity formula.

B. WEIGHTED MINIMUM ERROR SEAM METHOD FOR VIDEO BLENDING

The video blending function in this paper is realized by the weighted minimum error seam (weighted MES) system shown in Fig. 7. The weighted MES system receives video sequences and location information, which is obtained from the triangle keypoint matching system explained in section III-A. The weighted MES system consists of a stitching region generation module, HSV-based image difference module, parallax adjustment module, seam consistency module, and MES-based image blending module, as shown in Fig. 7. The stitching region generation module as the first step in the weighted MES system generates multiple video sequences of the regions to be stitched on the basis of the location information. Then, these video sequences of the region will be fed into three other modules, as shown in Fig. 7. The HSV-based image difference module provides error matrixes between the two images to be stitched in terms of individual HSV domains rather than in terms of the grayscale domain, as explained in section II. The parallax adjust module yields edge weight matrixes, which prevent objects from disappearing due to a faulty seam caused by parallax. The seam consistency module generates horizontal variance weight (HVW) matrixes, which are used to inhibit the sudden movement of MES between consecutive frames. This sudden movement causes objects to abruptly move in successive frames. Finally, on the basis of three matrixes, the error matrix, edge weight matrix, gand HVW matrix, the MES-based image blending module generates MES, which is applied to the stitching region into a stitched video sequence. The details of the HSV-based image difference module, parallax adjustment module, and seam consistency module are described in section III-B-1), III-B-2), and III-B-3), respectively.

1) HSV-BASED IMAGE DIFFERENCE MODULE

As described in section II, a conventional MES-based stitching system uses a boundary line for stitching two images, where the boundary line is extracted based on the minimum

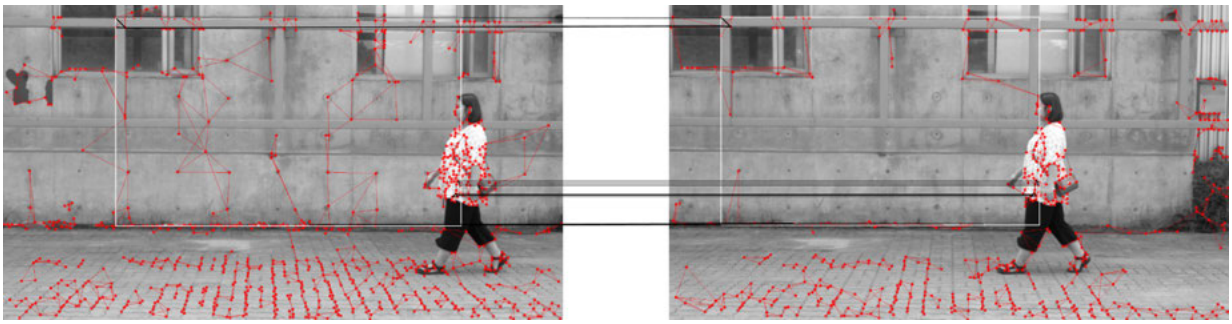


FIGURE 6. Example of images with a similar space calculation completed.

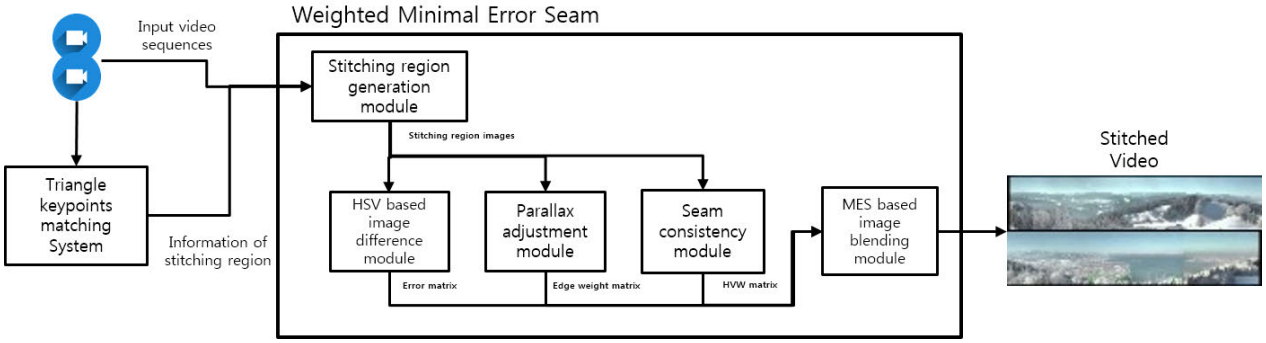


FIGURE 7. Weighted minimal error seam system.

error value using the mean square error, as expressed in (1):

$$e[i,j] = (X_1[i,j] - X_2[i,j])^2 \tag{1}$$

where i and j are the pixel positions on the vertical and horizontal axes, respectively, and $X_1[i,j]$ and $X_2[i,j]$ represent grayscale values at (i,j) in the two images to be stitched. The MES algorithm creates a boundary line based on the $e[i,j]$ values of each entry.

However, since the conventional MES is performed on the grayscale, unnatural boundary lines can be generated when images have similar brightness. Therefore, this paper proposes the method to apply the MES algorithm in terms of HSV domains, which help MES to reflect image differences by not only brightness but also both color and saturation. The MES in terms of the HSV domain proposed in this paper is expressed as follows:

$$H_e[i,j] = (H_1[i,j] - H_2[i,j])^2 \tag{2}$$
$$S_e[i,j] = (S_1[i,j] - S_2[i,j])^2 \tag{3}$$
$$V_e[i,j] = (V_1[i,j] - V_2[i,j])^2 \tag{4}$$

where the differences at (i,j) in terms of color, saturation, and brightness between the two images to be stitched are indicated. The difference between the two images to be stitched is represented as a single error value for the MES-based image blending as follows:

$$E[i,j] = H_e[i,j] + S_e[i,j] + V_e[i,j] \tag{5}$$

Thus, E[i, j] can be used to generate a boundary line for MES-based image blending with less sensitivity to brightness only.

2) PARALLAX ADJUSTMENT MODULE

This section describes the parallax adjustment module to prevent objects from disappearing due to a faulty seam caused by parallax. As explained in section I, the parallax causes the same object to be located at different positions in an image to be stitched. Fig. 8 shows an example of parallax, that is, the same object in the two images to be stitched appeared at different positions in a stitched image.

As illustrated in Fig. 2, a conventional MES needs to define a stitching region out of the two images to be stitched. As an example, Fig. 9 (a) and (b) shows the stitching region out of the two images to be stitched. A conventional MES may generate a boundary line, which is displayed as a blue and a red line in Fig. 9 (a) and (b), respectively, for the two input images. On the basis of these boundary lines, a stitched image is produced, as shown in Fig. 9 (c). This image is composed of two image regions: one is on the left of the boundary line and the other is on the right of the boundary line. Since the red circle part of the original image as being shown in Fig. 9 (a) is outside the boundary, the object is lost in the red circle area in the stitched image as being shown in Fig. 9 (c).

In order to prevent object disappearance in a conventional MES, an outer line value of an object produced by parallax

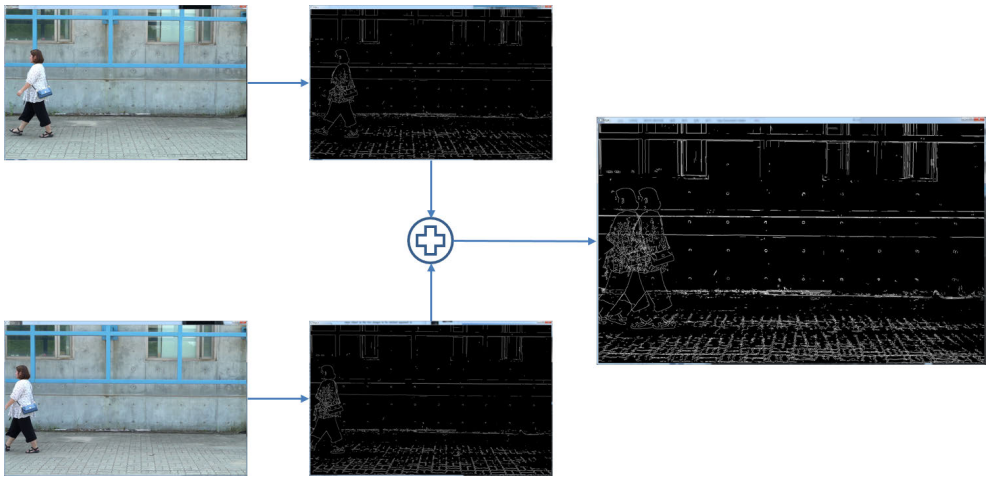


FIGURE 8. Example of parallax in terms of extracted edges.

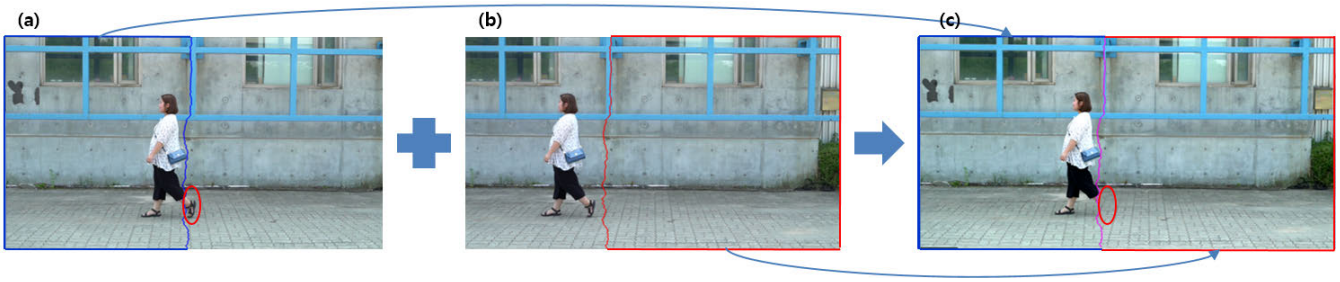


FIGURE 9. Example of objects disappearance due to a faulty seam.

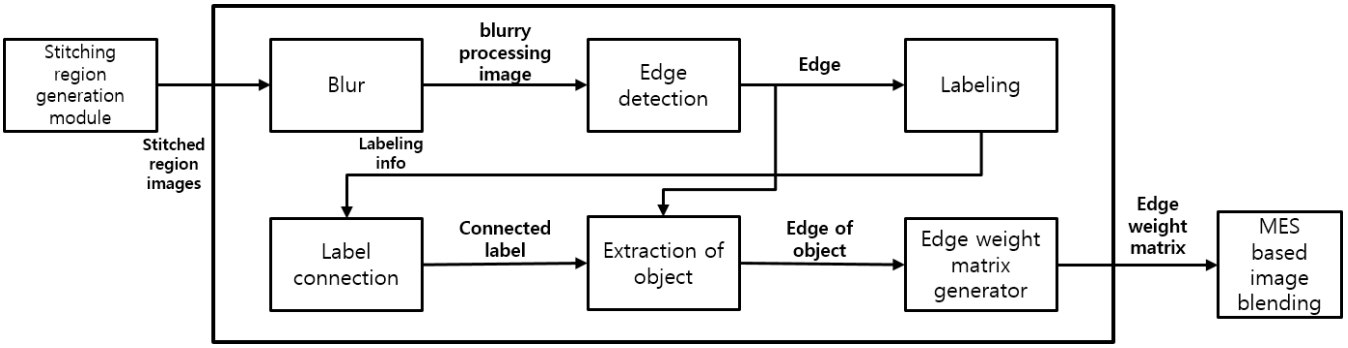


FIGURE 10. Parallax adjustment module.

should be considered when an error matrix in the MES is generated. The outer line value of the object can be yielded by an edge detection method. However, the outer lines of both an object and a background are generated. Thus, it is necessary to extract an outer line of only an object from all the detected edges in a stitching region.

In order to prevent a faulty seam caused by parallax, this paper therefore suggests the parallax adjustment module as shown in Fig. 10.

The parallax adjustment module presented in this paper consists of blur, edge detection, labeling, label connection, object extraction, and an edge weight matrix generator, as shown in Fig. 10. An edge detection method is applied to the stitching region images and produces a large amount of edge information from the objects and background together. In order to extract only the edge information of an object, this paper proposes the use of a blur sub-module for avoiding the sensitivity of the background, an edge detection sub-module to produce and add an edge image from blurred images, a labeling sub-module for binding continuous edge values as a rectangle, and a label connection sub-module for integrating adjacent labels by using dilation and erosion. Then,

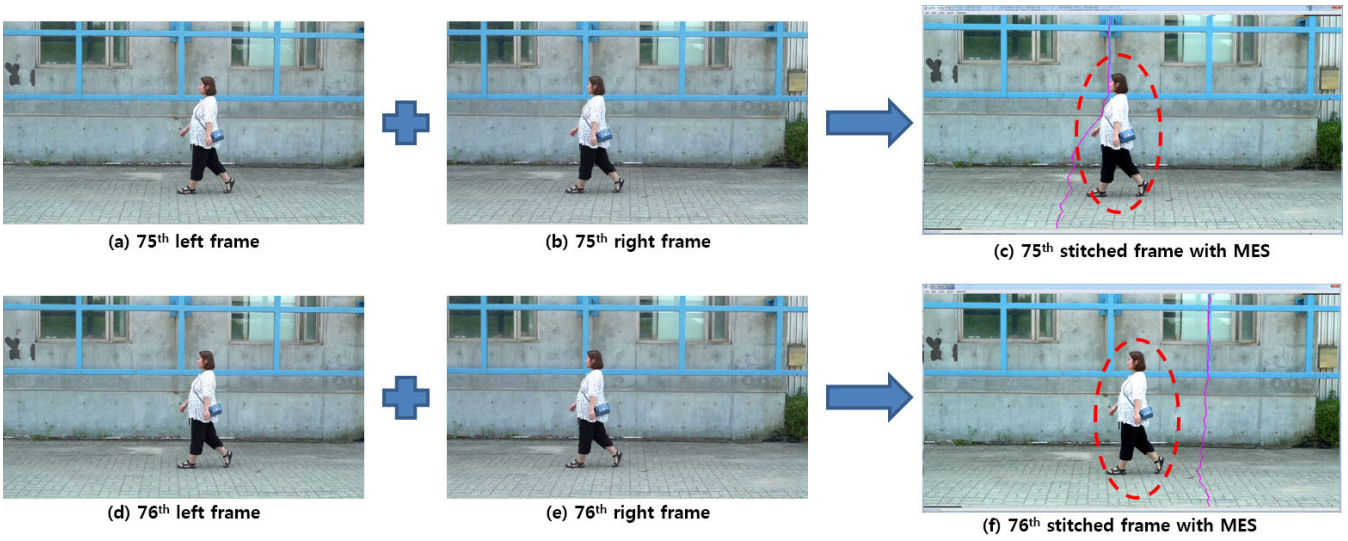


FIGURE 11. Object movement in successive frames.

edge information corresponding to the largest label rectangle can be used as object edge information. This object edge information will be used for compensating an error matrix of a conventional MES and is delivered by an edge weight matrix as follows:

$$Edge\ weight = (255 - edge\_value) \tag{6}$$

Therefore, the parallax adjustment module can prevent an object from disappearing in a stitched image, which is caused by a faulty seam in a conventional MES. Additionally, the parallax causes an object to abruptly move when MES is applied to a video sequence, which will be explained in the following section.

3) SEAM CONSISTENCY MODULE

This section describes the seam consistency module to overcome the disadvantage of unexpected object movement in successive frames. A conventional MES, as explained in section II, is generated on the basis of the two relevant frames to be stitched. When a conventional MES is applied to video sequences, an MES in a stitched frame can be abruptly changed in a subsequent one, which causes objects to unexpectedly move in successive stitched frames. For example, Fig. 11 (a) and (b) shows input frames captured from two cameras to be stitched, and the stitched frame is displayed in Fig. 11 (c) with a conventional MES. Additionally, subsequent frames to be stitched and the stitched frame are shown in Fig. 11 (d), (e), and (f), respectively. As presented in Fig. 11 (c) and (f), the object in the red circles appears to be abruptly moved to the right. This is because MES, shown as a purple line in Fig. 11 (c), is changed in a subsequent stitched frame, as exhibited with a purple line in Fig. 11 (f).

Thus, this paper proposes the seam consistency module in order to prevent this sudden movement of MES between consecutive stitched frames. The seam consistency module proposed in this paper produces the HVW matrix that considers MES in a previous frame when MES in a current frame is calculated. HVW suggested in this paper is defined as follows:

$$HVW[i, j] = (j - j_{error\_seam})^2 \tag{7}$$

where i and j are pixel positions on the vertical and horizontal axes, respectively, and j_{error_seam} is a horizontal position of an MES. Additionally, HVW in (7) is obtained for every vertical line i.

As an example, the error matrixes in a previous $t-1$ and current t frames are shown as Fig. 12 (a) and (b), where individual matrix indicates the overlapped region of two input images to be stitched, and the value of each matrix shows the graylevel difference between two input images. When these error matrixes are applied by the conventional MES explained in section II, the results are displayed as the red numbers in Fig. 12 (c) and (d) for a previous $t-1$ and a current t frames. As shown in Fig. 12 (c) and (d), the location of MESs are very different, even though two frames subsequently follow each other. In order to overcome this problem, this paper proposes the method of HVW-based MES, where a $t-1$ frame HVW obtained by (7) is added with an error matrix at t frame, as presented in Fig. 12 (e) and (f). The HVW-based MES at t frame is much closer to MES at $t-1$ frame, which can help the seam consistency module proposed in this paper to prevent objects from unexpectedly moving in successive frames.

As shown in Fig. 7, three module outputs such as HSV based error matrix, the edge weight matrix and HVW matrix explained before are used for MES based image blending module in Weighted minimal error seam system as follows:

$$E[i, j] = H_e[i, j] + S_e[i, j] + V_e[i, j] + Edge\ weight + HVW[i, j] \tag{8}$$

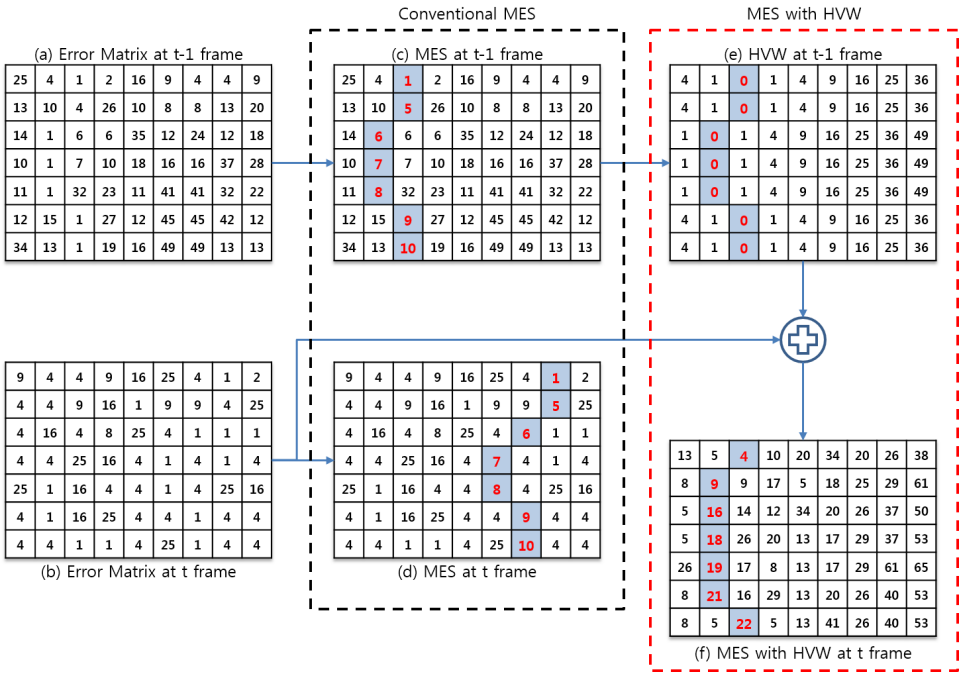


FIGURE 12. Method for generation of the horizontal variance weight.

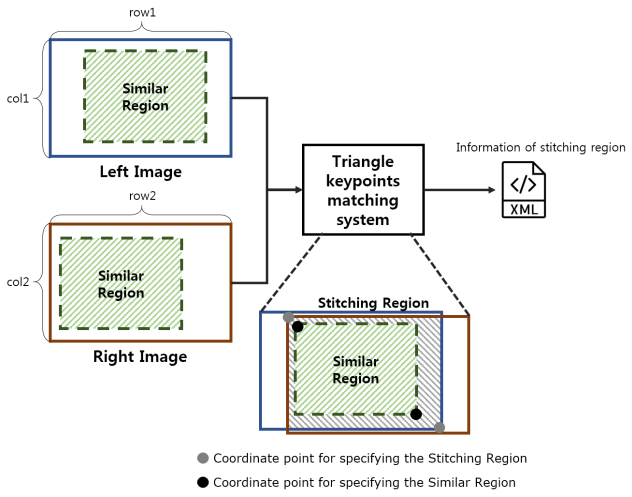


FIGURE 13. Generation of a stitching region by the triangle keypoint matching system.

```
▼<opencv_storage>
    <col1>1920</col1>
    <row1>1080</row1>
    <similar_region1>183 43 1796 1059</similar_region1>
    <stitching_region1>1749 2 1920 1080</stitching_region1>
    <col2>1920</col2>
    <row2>1080</row2>
    <similar_region2>4 41 1617 1056</similar_region2>
    <stitching_region2>0 0 1741 1078</stitching_region2>
</opencv_storage>
```

FIGURE 14. XML description of similar and stitching regions.

Section III-A proposed the triangle keypoint matching system for generation of location information, which is to conduct a matching process by triangle keypoints classification based on triangle similarities. This system reduces complexity and processing time.

And then, section III-B suggested the weighted MES system for video blending, which consisted of three modules to enhance the conventional MES: the HSV-based image difference module, the parallax adjustment module, and the seam consistency module. The HSV-based image difference module applies the MES algorithm in terms of HSV domains and helps MES to reflect image differences by not only brightness but also both color and saturation. The parallax adjustment module uses the extracted edges of objects, which prevents object disappearance due to a faulty seam caused by parallax. Finally, the seam consistency module overcomes the disadvantage of object movements in successive frames by using the HVW matrix.

IV. EXPERIMENTAL RESULTS

In this section, the performance of the MES-based efficient panorama video stitching method robust to parallax, which is composed of two functions, ''generation of location information'' and ''video blending'', as explained in the section III, is demonstrated. Before demonstrating the experimental results of the proposed method verification and performance evaluation, the experimental environment will be described. The experiments are conducted on a PC platform, which includes Microsoft Windows 7 as the OS and Visual Studio 2013 and OpenCV 3.1.0 for development. For the experiments, two

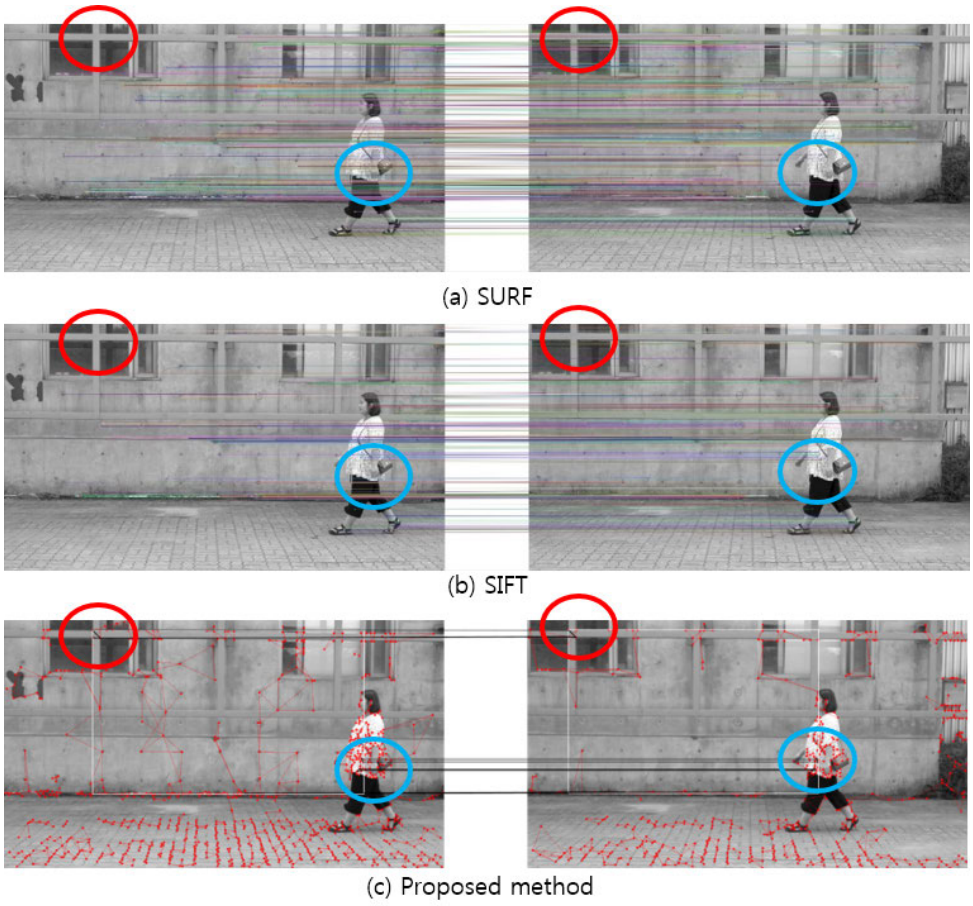


FIGURE 15. Results of keypoints matching.

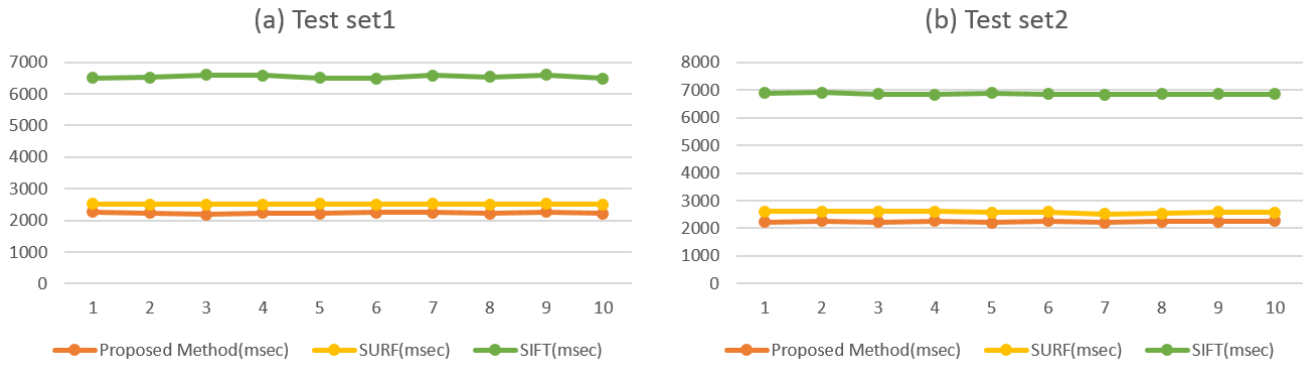


FIGURE 16. Comparison of processing time.

original video sets are used, which are obtained by using the video sequences from two different cameras. The spatial resolutions of the video sequences are 1920 × 1080. The result of generating location information, which is to conduct a matching process by triangle keypoints classification based on triangle similarities, will be explained in section IV-A. Additionally, the result of video blending, which has been achieved by the weighted MES method, will be described in section IV-B.

A. RESULT OF THE TRIANGLE KEYPOINT MATCHING SYSTEM USING TRIANGLE SIMILARITIES

A conventional stitching method extracts and describes features over whole regions of two images to be stitched, even

though a stitched region is a partial region of images, which increases the complexity and processing time. In order to reduce the complexity and processing time, this paper proposed the triangle keypoint matching system in section III-A, which obtains and describes the location information of a stitching region in two images by using the similarity ratio of triangles. As shown in Fig. 13, the triangle keypoint matching system takes the two images to be stitched as input and acquires the coordinate information of the similar region by using a similar triangle ratio, and then produces stitching region information. Since the stitching process is to be applied to only this stitching region rather than entire input images, it will reduce the computation in subsequent blending process. This stitching region information in this paper is described in an XML document. This is because a stitching region information can be handled independently of two input images to be stitched.

As being displayed in Fig. 13, the XML document defines the two tags such as <similar_region> for the region between two original frames calculated using the triangle keypoint matching system, and <stitching_region> for the region where stitching process is to be conducted in two input images.

As an example, the XML description displayed in Fig. 14 provides an input image size by <col> and <row> elements, and the location of similar and stitching regions by <similar_region> and <stitching_region> elements. In particular, these two regions indicated by <similar_region> and <stitching_region> elements are specified by the location of both the upper-left and lower-right coordinate points.

Then, the result of keypoints matching using the SURF, SIFT, and proposed method are shown in Fig. 15 (a), (b), and (c), respectively. The keypoints were extracted from two original frames and the matched keypoints were connected by line. As being shown in Fig. 15, the matching results such as the blue area in Fig. 15 are similar to all of the three methods, however, the matching results of the red area provided by the proposed method are missed from the results of SIFT and SURF.

As the quantitative evaluation of the triangle keypoint matching system, this paper compares the processing time of the proposed method with that of SURF and SIFT. The processing time of SURF and SIFT were measured in terms of three processes: feature extraction, description generation, and matching process. Additionally, multiple experiments have been conducted on the same image in order to improve the accuracy of the experiments. Two different test sets are used, and the processing time for each test set is measured ten times, as shown in Fig. 16 (a) and (b). Here, the processing time of the proposed method is less than that of SURF and SIFT. On average, the processing time of the proposed method over SURF is reduced by 11.06% in test set 1 and by 13.27% in test set 2. Then, the processing time of the proposed method over SIFT is reduced by 65.81% in test set 1 and by 67.35% in test set 2.

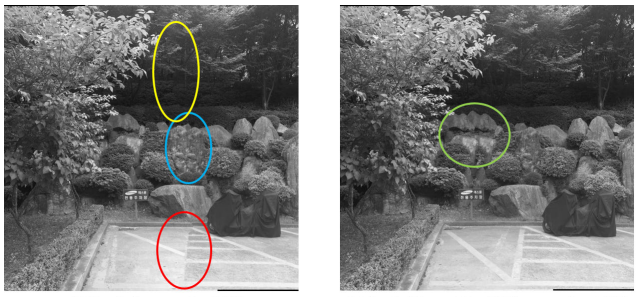

(a) Result of a conventional MES (b) Result of the applying HVW to conventional MES

FIGURE 17. Results of MES.

B. RESULTS OF THE WEIGHTED MINIMUM ERROR SEAM METHOD FOR VIDEO BLENDING

This paper proposes a weighted minimum error seam for the video blending function, which consists of an HSV-based image difference module, parallax adjustment module, and seam consistency module, as explained in section III-B. The HSV-based image difference module produces error matrixes of differences between two images to be stitched in terms of individual HSV domains rather than in terms of only a grayscale domain. The parallax adjustment module yields edge weight matrixes for preventing a faulty seam caused by parallax, which causes objects to disappear in a stitched video sequence. The seam consistency module generates HVW matrixes, which are used to overcome the disadvantage of unexpected object movement in successive frames. The results of the HSV-based image difference module, parallax adjustment module, and seam consistency module are described in section IV-B-1), IV-B-2), and IV-B-3), respectively.

1) RESULT OF THE HSV-BASED IMAGE DIFFERENCE

The HSV-based image difference module, as proposed in section III-B-1), avoids producing unnatural boundary lines brought up by the conventional MES.

As shown in Fig. 17 (a), conventional MES produces the errors indicated in the yellow, blue, and red circles. This is because conventional MES uses only differences of lightness, and thus brings up those errors when those images have similar light information. Since the proposed HSV-based MES considers not only grayscale but also chroma information, it overcomes those errors produced by a conventional MES, as displayed in Fig. 17 (b). Therefore, the proposed HSV-based MES can achieve more accurate stitching results than a conventional MES.

2) RESULTS OF PARALLAX ADJUSTMENT MODULE

The parallax adjustment module proposed in section III-B-2) is designed for preventing a faulty seam caused by parallax, which could cause objects to disappear in a stitched video sequence. In order to prevent object disappearance in a conventional MES, the proposed parallax adjustment module

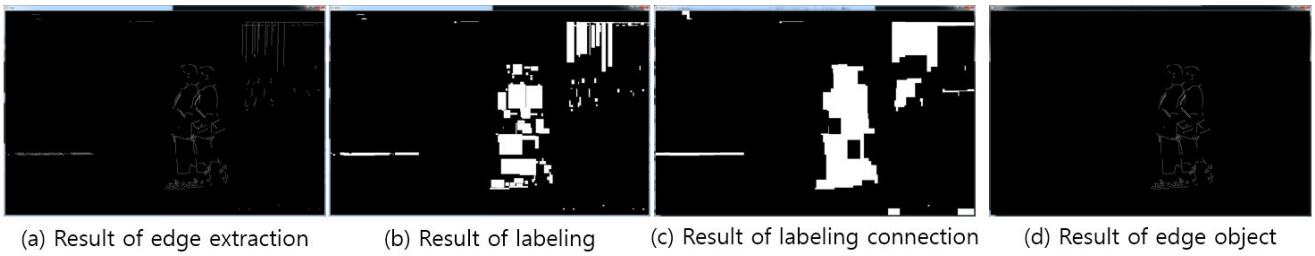


FIGURE 18. Results of the parallax adjustment module.

takes an object edge detection method in order to consider outer line values of objects for an error matrix. The proposed parallax adjustment module consisted of six sub-modules: blur, edge detection, labeling, label connection, object extraction, and edge weight matrix generation, as described in Fig. 10 in section III-B-2).

After the blur and edge detection sub-modules, a stitching region image still has the edges of an object as well as the edges of the background, as displayed in Fig. 18 (a). In order to extract a clear object region, this paper uses the labeling and the label connection sub-modules, as displayed in Fig. 18 (b) and (c), respectively. Here, the labeling sub-module produced rectangular labels to cover continuous edges in Fig. 18 (a), and those adjacent labels were merged by dilation and erosion in the label connection sub-module. Since background edges are diminished by the blur sub-module, the largest edge label can be considered as an object edge region. When this largest edge label region is applied to the edge image, the object region was clearly obtained, as shown in Fig. 18 (d). Finally, the edge weight matrix defined in (6) in section III-B-2), which will be used for compensating for an error matrix of a conventional MES.

Because of conventional MES does not consider an object, there is a problem that the loss of the object occurs as shown in Fig. 19 (a). To overcome this problem, in the result of applying the parallax adjustment module, losses can be avoided as shown in Fig. 19 (b).

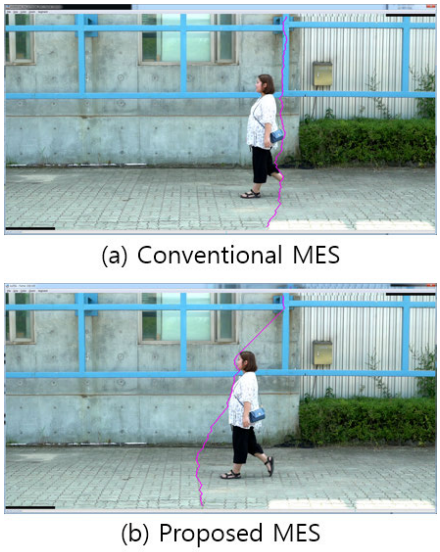


FIGURE 19. Comparison of conventional and proposed MES.

3) RESULTS OF SEAM CONSISTENCY

The seam consistency module suggested in section III-B-3) is designed to overcome the disadvantage of unexpected object movement in successive frames. When a conventional MES for image stitching is applied to video sequences, it is found that objects unexpectedly move in successive stitched frames when the position of MES abruptly moves. In order to prevent sudden movement of MESs between consecutive stitched frames, this paper proposed to produce the HVW matrix for applying the position information of the MES of a previous stitched frame into a current frame stitching, as described in section III-B-3).

Fig. 20 (a) and (b) show the 75th and 76th stitched frames with relevant MESs by using a conventional MES, respectively. As shown in these two figures, the person is supposed to move to the left in the 75th frame, and thus, the person is expected to appear at the left position in a successive frame compared to the 76th frame. However, the person in the successive 76th frame appeared in a position to the right compared to the 75th frame. This is because the MES in the 76th frame is obtained at the right position compared to the person, as displayed in Fig. 20 (b). In order to resolve this abrupt MES position change, this paper proposed to use the HVW defined in (7) in section III-B-3), where the MES of the 75th stitched frame is applied as the weight to prevent the MES of the 76th stitched frame from suddenly moving. As shown in Fig. 20 (c) and (d), the individual generated MES in both the 75th and 76th stitched frames are located in a

$$Y[\%] = \frac{\textit{frames with object loss or abrupt movement by proposed method}}{\textit{frames with object loss or abrupt movement by conventional MES method}} \times 100 \quad (9)$$

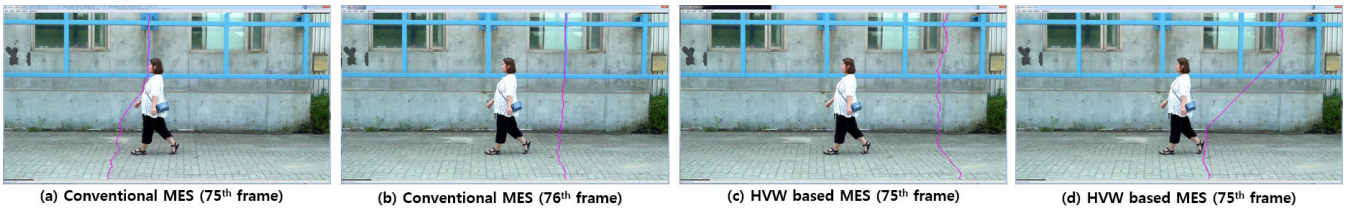


FIGURE 20. Results of stitching in successive frames.

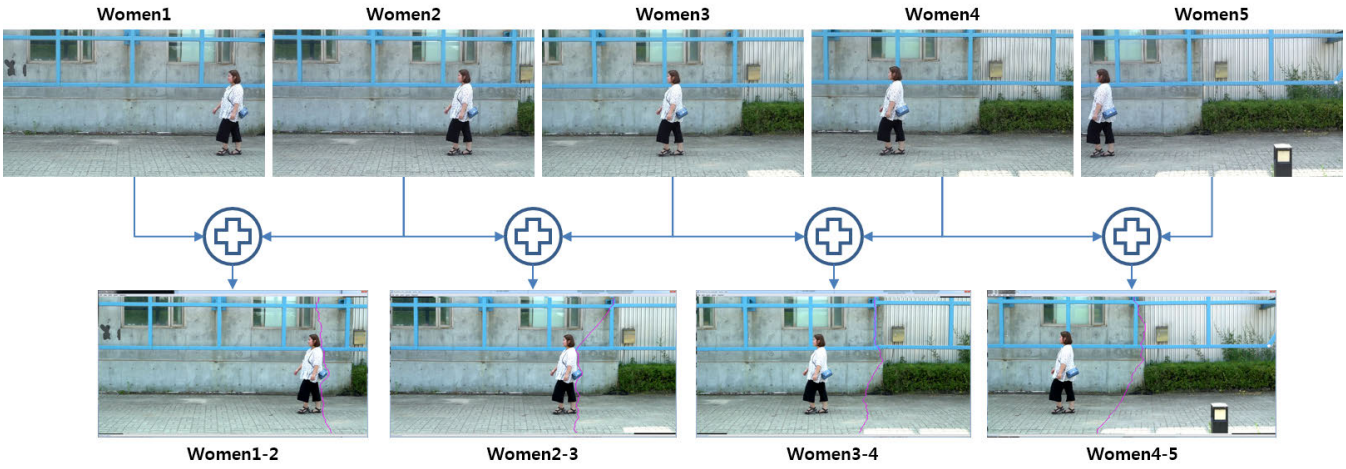


FIGURE 21. Original video sequences and stitched video sequences for test set 1.

similar position, and thus, the person is more smoothly moving in successive video frames after applying the proposed HVW.

Section IV-B explained the results of the proposed three modules to improve conventional MES for video blending. As the quantitative evaluation of a video blending function, this paper proposes (9), which shows the gains obtained by the proposed three modules,as shown in (9) at the bottom of the previous page.

The frames with object loss or abrupt movement is visually confirmed. For the experiments, five original video sequences of 600 frames are captured by five different cameras, and a pair of two adjacent video sequences are going to be stitched. Thus, four different stitched video sequences, Women1-2, Women2-3, Women3-4, and Women4-5, are generated as shown in Fig. 21. The test result of each experiment is shown in Table 1.

For more accurate experiments, this paper also conducted a test set called Men, as shown in Fig. 22, which is captured by five different cameras with 600 frames. The pair of two adjacent video sequences are to be stitched. Thus, four different stitched video sequences, Men1-2, Men2-3, Men3-4, and Men4-5, are generated as shown in Fig. 22. The test result of each experiment is presented in Table 2.

For more accurate experiments, this paper also conducted a test set called Men, as shown in Fig. 22, which is captured by five different cameras with 600 frames. The pair of two adjacent video sequences are to be stitched. Thus, four different stitched video sequences, Men1-2, Men2-3, Men3-4, and Men4-5, are generated as shown in Fig. 22. The test result of each experiment is presented in Table 2.

TABLE 1. Result of weighted MES for test set 1.

	Women 1-2	Women 2-3	Women 3-4	Women 4-5
Frames with object loss or abrupt movement using a conventional MES method	21	34	17	18
Frames with object loss or abrupt movement by the proposed method	4	5	4	4
Y[%] (Error ratio compared to conventional MES)	19	14.7	23.5	22.2

TABLE 2. Result of weighted MES for test set 2.

	Men1-2	Men2-3	Men3-4	Men4-5
Frames with object loss or abrupt movement using a conventional MES method	16	67	19	15
Frames with object loss or abrupt movement by the proposed method	4	6	4	5
Y[%] (Error ratio compared to conventional MES)	25	8.9	21.1	33.3

adjacent video sequences are to be stitched. Thus, four different stitched video sequences, Men1-2, Men2-3, Men3-4, and Men4-5, are generated as shown in Fig. 22. The test result of each experiment is presented in Table 2.

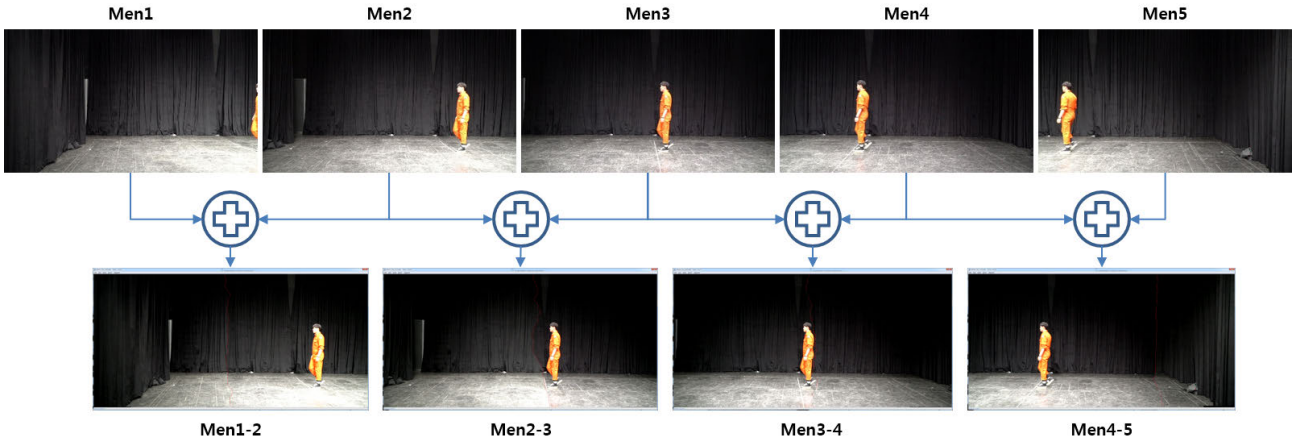


FIGURE 22. Original video sequences and stitched video sequences for test set 2.

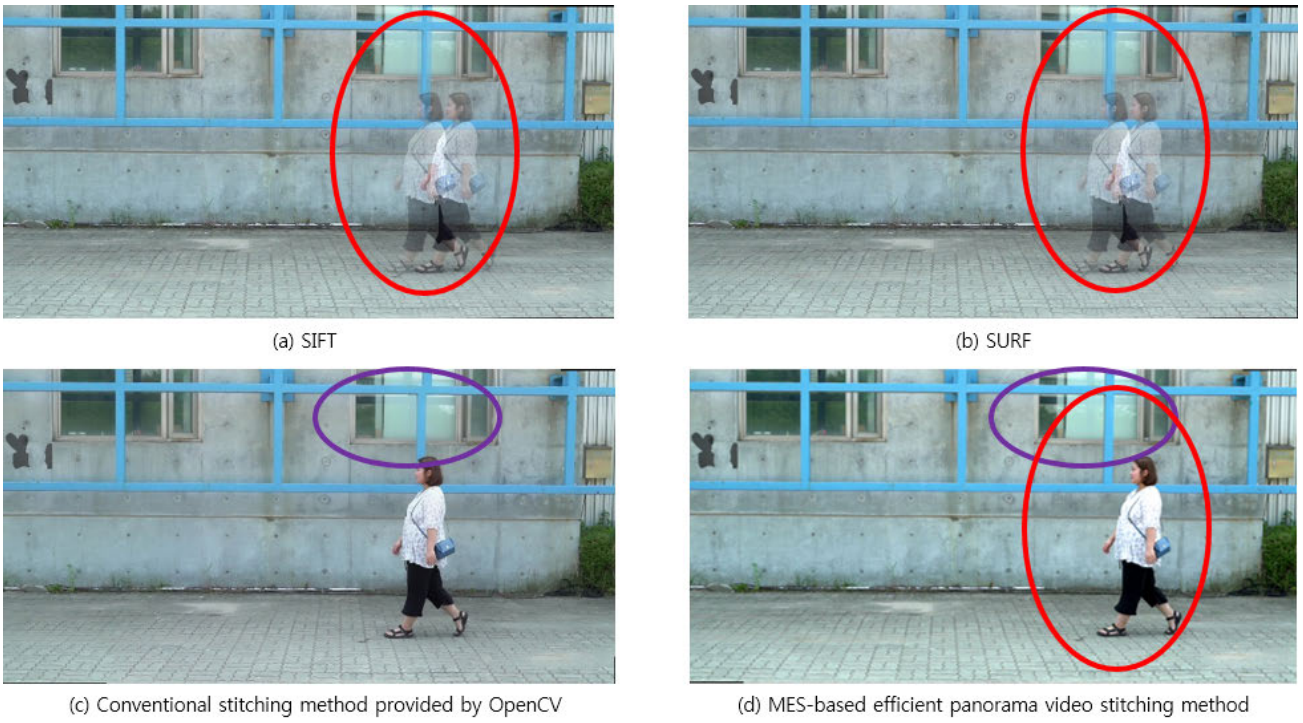


FIGURE 23. Comparison of SIFT, SURF, conventional stitching method and proposed method.

In the two test sequences, the error ratio compared to conventional MES was 8.9% $\sim$ 33.3%. Errors compared to existing technologies have decreased to 17.3% on average.

Also, in terms of a video stitching algorithm, the proposed method has been subjectively compared with widely used stitching methods such as SURF, SIFT and conventional stitching method provided by OpenCV. As being shown in the Fig. 23, subjective test results tell that SIFT and SURF produces distortions in object and blue frame as being shown in Fig. 23 (a) and (b), and the conventional stitching method provided by OpenCV generates distortions in window and blue frame as being shown in Fig. 23 (c). However, distortions has not been occurred when the proposed method is applied, as being shown in Fig. 23 (d).

This section explains the performance of the MES-based efficient panorama video stitching method robust to parallax, which is composed of two functions, ''generation of location information'' and ''video blending'' and describes the superiority of the proposed stitching method through comparison with conventional stitching method in terms of objective and subjective qualities.

V. CONCLUSION

This paper describes the computational burden and the accuracy degradation caused by parallax in a traditional stitching process. In order to overcome this problem, this paper proposed a minimum error seam-based efficient panorama video stitching method robust to parallax, which consists of two

functions, ''generation of location information'' and ''video blending''. The function of generating location information is to conduct a matching process by triangle keypoints classification based on triangle similarities, which reduces complexity and processing time. On average, the processing time of the proposed method over SURF is reduced by 12.17%. Additionally, in this paper, the video blending function is realized by the weighted minimum error seam system, which consists of a HSV-based image difference module, parallax adjustment module, and seam consistency module. The error ratio compared to a conventional MES method has decreased to 16.2% on average. It is expected that the proposed method can achieve a more accurate panorama stitching result with less complexity and be effectively adapted for not only a traditional stitching system but also recently emerging consumer electronics, such as smartphones and panorama devices.

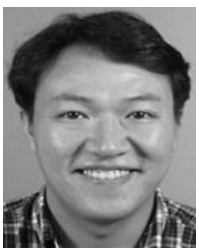
REFERENCES

[1] D. H. Seo, S. H. Kim, H. G. Park, and H. D. Ko, ''Real-time panoramic video streaming system with overlaid interface concept for social media,'' *Multimedia Syst.*, vol. 20, no. 6, pp. 707–719, Nov. 2014.
[2] Y. Deng and T. Zhang, ''Generating panorama photos,'' *Proc. SPIE*, vol. 5242, pp. 270–279, Nov. 2003.
[3] F. Huang, S.-K. Wei, and R. Klette, ''Comparative studies of line-based panoramic camera calibration,'' in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 7, Jun. 2003, pp. 16–22.
[4] J. C. A. Fernandes and J. A. B. C. Neves, ''Using conical and spherical mirrors with conventional cameras for 360°; Panorama views in a single image,'' in *Proc. IEEE Int. Conf. Mechatronics*, Budapest, Hungary, Jul. 2006, pp. 157–160.
[5] R. Szeliski and H.-Y. Shum, ''Creating full view panoramic image mosaics and environment maps,'' in *Proc. 4th Annu. Conf. Comput. Graph. Interact. Techn.*, Aug. 1997, pp. 251–258.
[6] 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.
[7] D. G. Lowe, ''Distinctive image features from scale-invariant keypoints,'' *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[8] B. He, G. Zhao, and Q. Liu, ''Panoramic video stitching in multi-camera surveillance system,'' in *Proc. 25th Int. Conf. Image Vis. Comput. New Zealand (IVCNZ)*, Nov. 2010, pp. 1–6.
[9] W. Xu and J. Mulligan, ''Panoramic video stitching from commodity HDTV cameras,'' *Multimedia Syst.*, vol. 19, no. 5, pp. 407–426, Oct. 2013.
[10] E. Karami, M. Shehata, and A. Smith, ''Image identification using SIFT algorithm: Performance analysis against different image deformations,'' in *Proc. Newfoundland Elect. Comput. Eng. Conf.*, St. John's, Canada, Nov. 2015, pp. 1–4.
[11] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, ''Speeded-up robust features (SURF),'' *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
[12] Y. Xiong and K. Pulli, ''Fast panorama stitching for high-quality panoramic images on mobile phones,'' *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 298–306, May 2010.
[13] S. Shin, S. Yoon, H. Jung, J. Kim, and K. Kim, ''Video blending based on minimal error seam,'' in *Proc. Korea Broadcast Media Eng. Conf.*, 2018, pp. 301–304.
[14] C. Herrmann, C. Wang, R. S. Bowen, E. Keyder, M. Krainin, C. Liu, and R. Zabih, ''Robust image stitching with multiple registrations,'' in *Proc. ECCV*, Sep. 2018, pp. 53–67.
[15] E. Adel, M. Elmogy, and H. Elbakry, ''Image stitching based on feature extraction techniques: A survey,'' *Int. J. Comput. Appl.*, vol. 99, no. 6, pp. 1–8, Aug. 2014.
[16] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, ''Brief: Binary robust independent elementary features,'' in *Proc. ECCV*, 2010, pp. 778–792.
[17] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, ''ORB: An efficient alternative to SIFT or SURF,'' in *Proc. ICCV*, Nov. 2011, p. 3.
[18] A. Alahi, R. Ortiz, and P. Vandergheynst, ''FREAK: Fast retina keypoint,'' in *Proc. CVPR*, Jun. 2012, pp. 510–517.
[19] J. W. Park, Y. T. Kim, and D. K. Shin, ''Feature points matching based on the condition of triangle similarity,'' *J. Korean Inst. Inf. Technol.*, pp. 91–100, 2010.
[20] Y. Deng and T. Zhang, *Generating Panorama Photos*. 2003.
[21] M. H. Hetal and A. P. Patel, ''Comprehensive Study and Review of Image Mosaicing Methods,'' *Int. J. Eng. Res. Technol. (IJERT)*, vol. 1, no. 9, pp. 1–7, Nov. 2017.
[22] M F. Abdul-Halim and N. A. Ibraheem, ''Utilizing genetic algorithms for 2D texture synthesis,'' in *Proc. Int. Conf., Comput. Graph. Artif. Intell.*, 2015.
[23] C. Harris and M. Stephens, ''A combined corner and edge detector,'' in *Proc. Alvey Vis. Conf.*, 1988.

JEONHO KANG received the B.S. and M.S. degrees in electronics engineering from Kyung Hee University, Yongin, South Korea, in 2010 and 2012, respectively, where he is currently pursuing the Ph.D. degree in electronics engineering. His current research interests include MPEG systems, interactive media processing, and image processing.

JUNSIK KIM received the B.S. and M.S. degrees in electronics engineering from Kyung Hee University, Yongin, South Korea, in 2017 and 2019, respectively, where he is currently pursuing the Ph.D. degree in electronics engineering. His current research interests include point cloud compression, MPEG systems, digital broadcasting technologies, and image processing.

INHONG LEE received the B.S. degree in electronics engineering from Kyung Hee University, Yongin, South Korea, in 2019, where he is currently pursuing the M.S. degree in electronics engineering. His current research interests include MPEG systems and image processing.



KYUHEON KIM received the B.S. degree in electronic engineering from Hanyang University, Seoul, South Korea, in 1989, and the M.Phil. and Ph.D. degrees in electrical and electronic engineering from The University of Newcastle, Newcastle upon Tyne, U.K., in 1996. From 1996 to 1997, he was with Sheffield University, U.K., as a Research Fellow. From 1997 to 2006, he was with the Electronics and Telecommunications Research Institute, South Korea, as the Head of the Interactive Media Research Team, where he standardized and developed T-DMB specification, and conducted the Head of Korean delegates for MPEG standard body, from 2001 to 2005. Since 2006, he has conducted research at Kyung Hee University, Seoul. He has published numerous technical articles. His current research interests include interactive media processing, digital signal processing, and digital broadcasting technologies. He was a recipient of the Ministry Award from the Ministry of Information and Communication, in 2003 and the Prime Minister Award, in 2005.