UAV-based automatic generation of high-resolution panorama at a construction site with a focus on preprocessing for image stitching

Seongdeok Bang, Hongjo Kim, Hyoungkwan Kim^*

^* Department of Civil and Environmental Engineering, Yonsei University, Seoul, Republic of Korea

1. Introduction

Image processing has been widely studied for construction management applications. Examples of such applications include productivity analysis [1,2], safety management [3–5], facility condition assessment [6–9], and progress monitoring [6,10–12]. However, to automatically monitor a large construction site, more often than not, the entire site or a large portion of it needs to be captured in an image. The relative locations of important construction objects, such as materials, heavy equipment, and temporary structures, can only be well understood in the context of the entire site or a large portion of it. This argument is more valid considering equipment movement analysis, progress monitoring, or material tracking for a large-scale construction site. Thus, a methodology is needed to generate a high-resolution panorama that covers a large construction site for various management purposes.

Image stitching has been developed in the computer vision field for the past few decades [9]. There are some off-the-shelf software packages such as Image Composite Editor (ICE) [13], Photostitch [14], and Autostitch [15] that automatically create a panorama of the multiple images inputted by users. Although these software packages make it easy to conduct the image stitching process, they do not always produce high-quality panorama images. The quality of panorama images is dependent on many factors including overlap ratios between images, lens distortion, and the quality level of the input image itself. In general, effective preprocessing is required to turn raw images into a high-quality panorama.

Advancements in unmanned aerial vehicles (UAVs) enable the efficient collection of image information of construction sites [16,17]. UAVs equipped with a camera can collect images from various perspectives, unlike closed-circuit television (CCTV) cameras. However, using UAVs for construction monitoring is challenging. First, the enormous number of images taken on construction sites must be processed promptly to be meaningful for project management. Second, vibrations in UAVs caused by wind and unstable control of the aircraft are likely to result in the acquisition of low-quality images. Third, the continuous position changes of UAVs makes camera pose estimation difficult.

To address these issues, this paper suggests an automatic method of high-resolution image generation using UAV videos, with a focus on preprocessing, which consists of three modules: blur filtering, key frame selection, and camera correction. The blur filtering module filters out blurred images caused by the vibration of the UAV. The module for key frame selection chooses frames that contain important information from UAV videos by maintaining a constant overlap ratio. Third, the camera correction module corrects camera lens distortion for accurate matching between images. The preprocessed images go through an image stitching algorithm. An experiment involving a real subway construction site in Korea was conducted to evaluate the proposed method. The resultant panorama image of the construction site covered a large portion of the site with high-quality detail. The remaining part...
of the paper includes literature review, methodology, and experiment. Finally, conclusions and recommendations are provided.

2. Literature review

Image stitching has been used in the construction industry mainly to visualize a structure for detecting specific anomalies (e.g., cracks, delamination, and fractures) or to visualize an entire presentation. Some of the studies used existing image stitching techniques as part of their research [7,18–21]. Adhikari et al. [7] stitched together images of cracked segments of bridges as preprocessing for detecting variations in the crack length and width. Klein et al. [18] stitched feature points for image-based 3D reconstruction to automate the updating of as-built documents. Li et al. [19] proposed a methodology to inspect the cracks of bridges using multiple image processing techniques. The detected crack images were stitched and presented as a surface panorama of the bridge substructure, and the locations of all cracks were represented in a panorama. Zhu et al. [20] also proposed a new method that used image stitching techniques to detect bridge columns in a bridge condition evaluation system. Jahanshahi et al. [21] used database technology and image stitching techniques to improve the inspection of structure quality.

On the other hand, some of the studies contributed to the further development of image stitching technologies for the construction industry [8,9,22,23]. Chaiyasarn et al. [8], Zhu et al. [9], and Lee et al. [23] improved image stitching technologies for the panoramic image of a tunnel by assuming the tunnel shape or rectifying the acquired images. Image rectification, which was a geometric correction to be represented in a plane image, was performed using the 3D surface geometry of the tunnel. Hsu [22] used a wavelet analysis to suggest an image registration algorithm for indoor and built environments.

UAVs with cameras have received attention as monitoring tools in the construction industry [10,24–28]. UAVs with high mobility and low cost are suitable for the regular monitoring of a large-scale construction project. Han et al. [10] presented a process using a UAV and 4D building information modeling (BIM) to automate construction progress monitoring. Construction site information saved in the BIM was used as a tool for path planning and navigation; thus, the UAV could autonomously collect and update the visual data. Gheisari et al. [24] evaluated and suggested potential applications of an unmanned aerial system (UAS) for the safety management of construction sites, based on real-time visual access to the job site environment. González-Jorge et al. [25] built a 3D model based on images captured by a UAV to monitor a breakwater. Siebert and Teizer [26] used a UAV to monitor an earthwork process by calculating the volume of the cut and fill area. Wen and Kang [27] implemented an augmented reality (AR) technology that integrates a virtual construction scene with an actual construction site scene captured by a UAV. Yeum and Dyke [28] used a UAV as a tool to automatically detect cracks for bridge inspection. As can be seen from the aforementioned studies, the utilization of UAVs with cameras has contributed to vision-based monitoring for the construction industry.

There have been some attempts to use both UAVs and image stitching techniques for facility monitoring. Ellenberg et al. [29] stitched red, green, and blue (RGB) and infrared (IR) images acquired from a UAV to identify bridge deck peeling. Eschmann et al. [30] used a UAV to acquire images and stitch them in order to observe the damage and cracks of the building. Although the approaches of Ellenberg et al. [29] and Eschmann et al. [30] advanced the body of knowledge because they combined image stitching techniques and UAVs. Luckily, they did not encounter problems in image stitching. However, in general, image stitching can cause many problems in its application to construction sites owing to unstable or uncontrollable operation of the UAV. Moreover, the use of video data acquired from UAVs can also be more difficult than still images. It is common to see failures in the stitching of video images obtained from UAVs during monitoring. Thus, this paper aims to contribute to vision-based construction site monitoring by proposing a UAV-based video image stitching method with a focus on preprocessing.

3. Methodology

3.1. Overview of the method

Fig. 1 describes the method for automatically changing a video of a construction site acquired with a UAV into a high-quality panorama that allows the entire construction site to be viewed at a glance. The first module identifies and removes frames that are blurrier than the surrounding frames. Based on the research of Crete et al. [31], to filter out frames over a certain threshold, we compared the blurriness of all frames with the moving average. The second module selects key frames to have a certain overlap ratio with the adjacent frames. By understanding the altitude and speed of the UAV, the module helps to uniformly express the entire area without assigning too many frames to a specific scene. The third module corrects the camera lens distortion of the key frames selected in the second module. The lens distortion of the key frames is corrected using the camera’s intrinsic parameters estimated through the calibration of the camera attached to the UAV. The preprocessed images are used to create a high-quality construction site panorama using an off-the-shelf image stitching program. Details of the proposed method are described in Sections 3.2 to 3.5.
3.2. Blur removal

The vibration and unstable control of a UAV often results in the production of blurred images. The motion of the object during shooting also causes motion blur. Blurred images are the main cause of poor image stitching quality, and should be removed [15,21,32]. A criterion for determining a blurred image among frames extracted from UAV video is needed. Crete et al. [31] proposed a "no-reference blur metric" that quantifies the blur effect in each image without referencing other images. The blur estimation in their approach started with the fact that the blur image lost its original high-frequency content. Thus, they intentionally blurred the image through a low-pass filter, and compared the intensity variations within the image before and after the filtering. A large deviation in the intensity variations between the original and the blurred images means that the original image is sharp. If the deviation is small, the original image is considered to be blurred. However, images cropped from the same image can have different blur metric values depending on the visual objects. Fig. 2(b), (c), and (d) show images cropped at the same resolution of 201 × 201 from the original image shown in Fig. 2(a). Although Fig. 2(b), (c), and (d) are all from the same original image, they turn out to have different blur metric values (0.3147, 0.4600, and 0.4871). In order to solve this problem, a frame whose blur metric value is larger than those of neighboring frames is determined to be a blurred image using the moving average concept.

The proposed filtering method was implemented using Eqs. (1) and (2). Eq. (1) shows the procedure to calculate the blur metric value of each frame. To properly identify the frames for the moving average filter, we use three cases of relationships between the position of the frame of interest (k) and the length of moving average filter (n). Once the blur value is obtained from Eq. (1), it is subtracted from the blur metric value for the frame (x_k). If the difference exceeds the threshold value (α), the frame is judged as a blurred image, as shown in Eq. (2).

\[
\begin{align*}
\bar{x}_k & = \begin{cases} 
\frac{\sum_{i=1}^{k-1} x_i}{k-1} , & \text{if } k \leq \frac{n-1}{2} \\
\frac{\sum_{i=k-(n-1)}^{k} x_i}{n} , & \text{if } \frac{n-1}{2} < k < m - \frac{n-1}{2} \\
\frac{\sum_{i=m-n}^{m} x_i}{2m-n} , & \text{if } k \geq m - \frac{n-1}{2}
\end{cases}
\end{align*}
\]

(1)

if \( x_k - \bar{x}_k > \alpha \), then the k-th frame is a blurred image

(2)

where \( x_k \) is the blur metric value of the k-th image, \( \bar{x}_k \) is the moving average of the blur metric value of the k-th image, n is the number of
neighboring frames and is an odd number, and \( m \) is the total number of frames in the video.

### 3.3. Key frame selection

Stitching all the frames extracted from the video is not only inefficient in computation time, but can also fail to generate panoramas [30]. When a relatively large number of video frames is allocated to a specific region in the UAV video, misalignment may occur during image stitching because the overlap ratio between frames is not constant. This phenomenon occurs easily when the flight speed and altitude of the UAV are not constant. In this section, we explain the process of selecting key frames with a certain overlap ratio between images using the principle of triangulation.

Eq. (3) represents the number of key frames selected per unit time, and utilizes the concepts of instantaneous horizontal velocity, instantaneous effective horizontal distance, and unit time of the UAV. The instantaneous effective horizontal distance is the difference between the viewing distance of the UAV and the overlap distance, which makes the corresponding key frame have a constant overlap ratio with adjacent key frames (Fig. 3). In this study, we assumed that the instantaneous effective horizontal distance is the same as the difference between the UAV positions that photographed the two frames. This is also the actual distance of the two adjacent key frames’ center points. The unit time refers to the specific time interval of the horizontal velocity recorded by the UAV.

\[
\text{Number of selected key frames per unit time (NKFUT)} = \frac{d_{\text{v}} \text{ of UAV}}{d_{\text{H}}},
\]

\[
= \frac{d_{\text{v}} \text{ of UAV (m/s)}}{2 \times (1 - \text{OR}) \times \tan \left( \frac{\text{FOV}}{2} \right) \times dH \text{ of UAV (m)} \times \frac{1}{dt}},
\]

(3)

where \( d_{\text{v}} \) is the instantaneous horizontal velocity of UAV, \( d_{\text{H}} \) is the instantaneous effective horizontal distance, OR is the overlap ratio set by the user, FOV is the field of view of UAV, \( dH \) is the instantaneous altitude of UAV, \( dt \) is the unit time of the flight data recorded in the UAV, and the instantaneous horizontal velocity and instantaneous altitude are imported from the flight data of the UAV, corresponding to the playback time of the video.

As illustrated in Fig. 3, the real-time horizontal velocity of the UAV and the calculated effective viewing distance of the camera enables the calculation of the time it takes for the UAV to move the instantaneous effective horizontal distance. The result is then multiplied by the conversion factor \( \frac{m}{1000} \) to match the unit time of seconds to the unit time of the UAV, which produces the number of key frames per unit time with the proper choice of an overlap ratio.

\[
f(t) = \sum_{i=1}^{\lfloor t/dt \rfloor} (\text{NKFUT}),
\]

(4)

where \( \text{NKFUT} \) is the number of key frames per unit time, and Eq. (4) shows that \( f(t) \) is the number of frames selected until time \( t \). In this case, the maximum value of the floor function of \( f(T) = \text{the total playtime of the video} \), which is the maximum integer not exceeding \( f(T) \), is the total number of key frames (\( N \)). The value of \( f(t) \) takes a lot of real numbers according to the changing \( t \) value. Among these real numbers, the values closest to natural numbers are selected. Finally, these frames that have the selected numbers are determined to be the key frames.

The detailed procedure is as follows (Fig. 5):

1. Initialize \( n \) as 1.
2. Find the maximum value of \( f(t) \) that does not exceed the natural number \( n \).
3. Identify two video playback times \( t \) and \( t + dt \). The time \( t \) corresponds to the value found in step (1), and \( t + dt \) is obtained by adding the UAV unit time to the time \( t \). Then, transform the two values into the frame unit by multiplying them and the frame per second (fps) of the video.
4. Using a linear interpolation method with \( f(t) \) and \( f(t + dt) \) as the end points, select the frame (\( n \)) at which the value of \( f(t) \) is closest to \( n \). If the corresponding frame was filtered by the method described in Section 3.2, the frame (\( n \)) at which the value of \( f(t) \) is next closest to \( n \) is selected, and this process is repeated until a frame is selected.
5. Repeat steps (2) to (4) by incrementing \( n \) by 1 until \( n \) reaches the total number of key frames (\( N \)).

Fig. 3 shows an example of the process of selecting a key frame using \( f(t) \). The unit time of the UAV flight data in the example is 0.1 s, and the frame rate of the video is 30. In Fig. 4, the 10th key frame was captured between 37.6 s and 37.7 s based on the video playback time. With the conversion of seconds to frames, the 10th key frame is located between the 1128th and 1131st frames of the video. Using the two points 1128 and 1131, and the corresponding values 9.975 and 10.02 of the function \( f(t) \), the 1130th frame closest to the natural number 10 is
image misalignment in image stitching, which degrades the quality of parallel with the image sensor. The two types of lens distortion causes because the camera lens is decentered or because the lens is not in standard lens. Second, there is a tangent distortion that occurs either graph a wide range, more radial distortion occurs than when using a convex lens. When a camera using a wide-angle lens is used to photo-
there is a radiation distortion caused by the refractive index of the 

3.4. Camera correction

When a three-dimensional real world is projected as a two-dimen-
sional image through a camera, distortion owing to the camera lens 
occurs. There are two main types of camera lens distortion [33]. First, 
there is a radiation distortion caused by the refractive index of the 
convex lens. When a camera using a wide-angle lens is used to photogra-
ph a wide range, more radial distortion occurs than when using a 
standard lens. Second, there is a tangent distortion that occurs either 
because the camera lens is decentered or because the lens is not in 
parallel with the image sensor. The two types of lens distortion causes 
image misalignment in image stitching, which degrades the quality of the 
panorama [34,35]. In order to correct the radial distortion and 
tangent distortion, the intrinsic parameters of the camera (focal length, 
principal point, skew coefficient, and distortion coefficient) must be 
calculated through camera calibration. A distortion correction process 
using the intrinsic parameters of the camera was implemented ac-
cording to the methodology described in [36].

3.5. Image stitching

The image stitching process includes feature extraction, feature 
matching, bundle adjustment, and image blending in order to construct 
a panorama image from a large amount of input images. First, feature 
points are extracted from each image considering the camera pose. In 
image stitching, feature-based methods are preferred over direct-based 
methods for all pixels because the former is robust to differences in 
scale, orientation, and foreshortening [15,37]. Second, the feature 

4. Experiment

4.1. Experiment environment

The proposed methodology was implemented using the MATLAB 
programming language. The UAV used in this experiment was a 
Phantom 3 Professional from DJI Corporation. The fps of the UAV 
camera was 30, and the resolution of the video was 3840 by 2160. The 
camera calibration was conducted using the MATLAB 
cameraCalibration App [36]. A subway construction site and a parking 
lot in the city of Namyangju, Korea, were used as test sites for UAV-
based monitoring. The videos captured from both sites had playtimes of 
84 s and 21 s, respectively.

4.2. Results

Fig. 6 shows the result of the blur measurement for all frames of the 
video, and indicates that a large variation exists in the blur measure-
ment of the frames. As previously mentioned, the blurriness of each 
image was compared with the moving average. If the difference was 
larger than a threshold value of 0.01, the frame was determined to be 
blurred image. Of a total of 2540 frames used in this study, a total of 
163 frames were considered as blurry and were removed, resulting in 
the use of 2377 frames.

In order to find the proper overlap ratio, a range of ratios were 
selected as the key frame, based on the interpolation method. If the 
1130th frame is determined to be a blurred image and is removed, the 
1129th frame, which is the second closest to the natural number 10, is 
selected as the key frame. Algorithm 1 summarizes the steps for se-
lecting key frames. According to this procedure, frames that maintain 

Algorithm 1
Selecting key frames from UAV video with blur removed.

Input: Flight data of UAV (horizontal velocity and altitude), FOV of UAV, and overlap ratio
Output: Key frames

\begin{algorithm}
\begin{algorithmic}
\Procedure{keyFrameSelection}{Flight data of UAV (horizontal velocity and altitude), FOV of UAV, and overlap ratio}
\State \textbf{Input}: Flight data of UAV (horizontal velocity and altitude), FOV of UAV, and overlap ratio
\State \textbf{Output}: Key frames
\State \textbf{Algorithm 1} summarizes the steps for selecting key frames. According to this procedure, frames that maintain a constant overlap ratio are selected from the video regardless of the changing UAV position.

Algorithm 1
Selecting key frames from UAV video with blur removed.

\begin{input}
\textbf{Input:} Flight data of UAV (horizontal velocity and altitude), FOV of UAV, and overlap ratio
\end{input}

\begin{output}
\textbf{Output:} Key frames
\end{output}

\begin{algorithm}
\begin{algorithmic}
\Procedure{keyFrameSelection}{Flight data of UAV (horizontal velocity and altitude), FOV of UAV, and overlap ratio}
\State Temporary = Number of extracted frames per unit time calculated by Eq. (3)
\State Accumulated $\leftarrow$ Accumulated + Temporary
\For{$i = 1$ to $N$ (where $N$ is total number of key frames)}
\State $x_i = \text{argmin}(|f(x_i, y_i)|)$ (i.e., $x_i$ is the point on $f(x, y)$ where its parameter becomes the maximum, and $(x_i, y_i)$ is a point on $f(x, y)$)
\State $g(x) = \frac{x_{i+1} - x_i}{y_{i+1} - y_i}$, $x = x_i + g(x)(y - y_i)$
\State $x_{i+1}$ is the point on $f(x, y)$ where its parameter becomes the minimum
\EndFor
\State $\text{Accumulated obtained here plays the role of}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
\State $\text{Temporary + Accumulated}$
\State $\text{[Temporary + Accumulated obtained here plays the role of]}$
tested, including 50%, 60%, 70%, 80%, and 90%. In previous studies [8,9], the overlap ratio for image stitching was set at 50% for tunnel applications, so the overlap ratio in this study was set to at least 50%. Fig. 7 shows the panoramas produced with different overlap ratios, and the number of selected key frames. With a lower level of overlap ratio, more missing parts occur in the resultant panorama; Fig. 7(c), (d), and (e) show some parts that do not exist in Fig. 7(a) and (b). The reason for this is conjectured to be as follows. First, the actual FOV of the UAV is not as wide as it claims to be. In this experiment, a smaller FOV than the specified FOV of 94° was observed. Second, unlike the previous studies [8,9], the positional change of a UAV occurs in three dimensions, which makes it difficult for the UAV to capture the target object in a stable manner. Fig. 8 represents the horizontal velocity and altitude of the UAV corresponding to the video used in the experiment. This shows a large variation in the horizontal velocity and altitude of the UAV, and shows the importance of key frame selection. In this study, we set the overlap ratio as 70% to consider the quality and efficiency of the image stitching. Overlap ratios of 80 or 90% correspond to a higher number of key frames, leading to more processing time for image stitching, and more misalignment.

4.3. Evaluation and discussion

The validity of the proposed methodology was evaluated in three ways. First, the accuracy of the panorama was evaluated by comparing the generated panorama with the aerial photograph taken at the test site. Second, a blurriness comparison between the generated panorama and an aerial photograph containing the same object was performed to evaluate the ability to represent details. For the same purpose, we compared the blurriness between a portion of the generated panorama and the corresponding original frame containing the same object. Third,
we evaluated the necessity of this methodology by comparing the panorama generated by the proposed method and the panorama produced by the existing image stitching technique without going through the proposed methodology.

Fig. 9 shows an aerial photograph taken by the UAV at an altitude of 93 m, and the corresponding panorama. The resolutions of Fig. 9(a) and (b) are 3840 by 2160 and 8345 by 10,531, respectively. Fig. 9(b) shows the panorama composed of 14 image frames with an overlap ratio set at 70%. The comparison shows that dislocation or misalignment of objects in the panorama is not significant compared with the aerial photograph. Therefore, the panorama generated by the proposed methodology can be considered to have sufficient accuracy for construction site monitoring.

Fig. 10 compares the blurriness of aerial (Fig. 10(a) and (d)),
panoramic (Fig. 10(b) and (e)), and original images (Fig. 10(c) and (f)) with the same scale (1:550). Fig. 10(a), (b), and (c) are cropped images showing reinforcing bar objects, whereas Fig. 10(d), (e), and (f) are cropped images with tarpaulin objects. Fig. 10(a), (b), and (c) have blurriness values of 0.6597, 0.3914, and 0.3178, respectively. The blurriness was measured using the blur metric mentioned in Section 3.2. The blurriness value (0.3914) of the panorama (Fig. 10(b)) was measured to be slightly higher than that (0.3178) of the original frame (Fig. 10(c)) owing to an error occurring in the image stitching process. However, the blurriness value of the panorama is much lower than that (0.6597) of the aerial image (Fig. 10(a)). The same analysis result was derived for the tarpaulin objects. This comparison indicates that the
visual information of the object in the panorama was better preserved than in the aerial photograph.

Fig. 11(a) and (b) show a comparison of the image stitching method with and without the proposed method based on a UAV video of a parking lot. Compared with the image result from the proposed method (Fig. 11(a)), the result without the method (Fig. 11(b)) is a distorted image with many misalignments and blurry regions. This comparison demonstrates the need for the key frame selection, blur removal, and camera correction of the method. However, a comparison of Fig. 11(a) and (b) shows that the top portion of Fig. 11(b) is missing in Fig. 11(a), indicating that the panorama misses some objects in the video. This error is estimated to be caused by the fact that the UAV horizontal velocity, one of the parameters of Eq. (3) for selecting key frames, generally converges to 0 at the beginning and end of the video. Thus, the scenes corresponding to the start and finish times are not selected.

To avoid this problem, it is necessary to capture a wider area than the target area with the UAV for the proposed method. Fig. 12 shows examples of misalignments in the panorama shown in Fig. 9(b). Small-scale misalignments that cannot easily be seen are also believed to exist in Fig. 9(b). This error occurs for two main reasons. First, the estimation of homography and the lens distortion coefficient may not be accurate. Second, it is difficult to adjust images for the large parallax that occurs at the edge of an image that is stitched together with adjacent images.

5. Conclusion

This paper presented an image stitching method with a focus on the preprocessing part for images collected from UAVs. The preprocessing
method is composed of three modules: blur removal, key frame selection, and camera correction. The blur removal module eliminates blurred images, whereas the module for key frame selection maintains a constant overlap ratio between adjacent frames. These modules, coupled with the camera correction module, allows for the production of high-quality panoramas of construction sites. The proposed method was evaluated by a comparison with other methods such as aerial photography and image stitching without preprocessing.

The main contribution of this study is a preprocessing method to improve the quality of image stitching. To the best of our knowledge, this is a unique attempt to apply the blur removal and key frame selection techniques to images obtained by UAVs. Specifically contributions are twofold. First, the blur removal module employs the moving average of the blur values of the adjacent images, unlike existing techniques. Second, the key frame selection module uses the altitude and horizontal velocity information in order to choose key frames. This improvement allows the proposed method to produce a panorama, which is conjectured to be accurate and detailed enough for construction management applications. However, future studies are required to bring this study to its full fruition. The aforementioned deficiencies such as misalignment and information loss suggest a direction for future studies.

Acknowledgement

This research was supported by grants (17CTAP-C133290-01) from Infrastructure and Transportation Technology Promotion Research Program funded by Ministry of Land, Infrastructure and Transport (MOLIT) of Korean government and Korea Agency for Infrastructure Technology Advancement (KAIA) and by National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (No. 2011-030040). The authors would like to thank SK Engineering and Construction for granting access to the construction site of the subway project.

References


