

# Experimental Optimal Choice Of Initial Candidate Inliers Of The Feature Pairs With Well-Ordering Property For The Sample Consensus Method In The Stitching Of Drone-based Aerial Images

Byeong-Chun Shin<sup>1</sup> and Jeong-Kweon Seo\*<sup>1</sup>

<sup>1</sup> Department of Mathematics, Chonnam National University  
Gwangju, South Korea  
[e-mail: bcshin@chonnam.ac.kr]  
[e-mail: seojksc@gmail.com]

\*Corresponding author: Jeong-Kweon Seo

*Received April 22, 2019; revised July 25, 2019; revised December 3, 2019; accepted February 5, 2020;  
published April 30, 2020*

---

## Abstract

There are several types of image registration in the sense of stitching separated images that overlap each other. One of these is feature-based registration by a common feature descriptor. In this study, we generate a mosaic of images using feature-based registration for drone aerial images. As a feature descriptor, we apply the scale-invariant feature transform descriptor. In order to investigate the authenticity of the feature points and to have the mapping function, we employ the sample consensus method; we consider the sensed image's inherent characteristic such as the geometric congruence between the feature points of the images to propose a novel hypothesis estimation of the mapping function of the stitching via some optimally chosen initial candidate inliers in the sample consensus method. Based on the experimental results, we show the efficiency of the proposed method compared with benchmark methodologies of random sampling consensus method (RANSAC); the well-ordering property defined in the context and the extensive stitching examples have supported the utility. Moreover, the sample consensus scheme proposed in this study is uncomplicated and robust, and some fatal miss stitching by RANSAC is remarkably reduced in the measure of the pixel difference.

---

**Keywords:** Computer Algorithms, Scale-invariant feature transform, Feature matching, Stitching images, Drone-based aerial image

## 1. Introduction

**B**y the 1990s, various image registration technologies were developed, starting with the patch-based translational alignment technique developed in the early 1980s [1]. It has been a principal axis of computer vision (see [2] and references therein). In terms of applications, developers apply the techniques of image registration in remote sensing; for example, environmental monitoring, change detection, image mosaicking, and integrating information in geographic information systems (GIS)). They also apply it in medicine, including combining computed tomography (CT) to diagnose tumor, and in computer vision including target localization, automatic quality control, *etc.* [3]. The primary process of image registration mainly constitutes of four parts: feature detection, feature matching, transform estimation, and image remapping (see [2, 3] and references therein).

In general, image registration takes place in the matching scheme depending on the following ways: by feature-based registration using common feature descriptors or by pixel-to-pixel direct matching based on image geometry. Matching by feature descriptors is mainly known to be relatively robust and to exhibit fast computation speed.

In this study, we apply the drone-based aerial image in a mosaic creation for the utility of photovoltaic (PV) panel inspection, which allows the inspectors to observe PV array change and locate broken ones to repair them. One may find similar applications for the computer-vision based various utility of the drone aerial images [4, 5, 6, 7].

For the stitching of images using aerial images, there is commercial software in use: Autopano Giga in [6] and Agisoft Photoscan in [4]. Moreover, some problems have risen with these. One of these is due to the low flight height of the drones, which causes the failure of image matching by misinterpretation of the same object at different angles [6]. For the problem of mismatching, they suggest higher altitudes of the drone's flight height, *e.g.*, 300m above ground, to minimize the perspective disparity. Another problem is that the photographed object may be distorted by the reflection of sunlight [4]. For this problem, it is suggested that the acquisition of images is conducted on a cloudy day or late afternoons to make the amounts of the direct sunlight reflections small.

Thus, the problem of the stitching of the aerial images in computer vision is not trivial, and it is crucial to develop algorithms that work well in various unspecified environments.

The purpose of this study is to propose a new scheme to stitch the drone-based aerial images by means of feature-based registration; we show the efficiency of the new scheme. The main process of our proposed methodology is as follows:

### A. Extracting feature descriptor by SIFT and matching the feature points

We apply the scale-invariant feature transform (SIFT) feature descriptor [8, 9] as the descriptor of the feature-based matching in order to consider the speed of calculation, because the SIFT method is fast and has invariant properties on the scale and has been proven to maintain high repeatability for projection transformations [10, 11].

### B. Estimating hypothesis by sample consensus method

The next consideration is to find the hypothesis of the mapping function to have the resampling of the sensed image. For the calculation of finding the hypothesis, we develop a

new sample consensus method and compare it with the random sample consensus method (RANSAC) given in [12] which is generally known to maintain high accuracy in matching features. On the other hand, there are other methods, such as the M-estimator and Least-median of squares, which are known to maintain good results when the false positiveness is low [13, 14, 15]. In this study, for the possible high false positiveness rates of the matching features of the drone aerial images, the type of sample consensus method is further applied.

Regarding the process of finding hypothesis via feature points, for RANSAC would be sensitively affected by its initial guess there has been developed many modifying schemes as the family of RANSAC, and most of them are based on probability strategies. Performances of these have been verified to some extent [16]. Instead, our newly invented sample consensus method is decisive and intuitive without the need to consider a probability conception, but only the geometric properties; we develop the method considering the geometric congruence given by the affine transformation between the pairs of feature points, which leads to determining some optimally chosen initial candidate inliers (OCICI) in the sample consensus method. In the experimental results, we present the verification of our proposed sample consensus method, OCICI, superior to the general RANSAC, focusing on the process of finding the hypothesis of the mapping function by various tests. Amid the tests, a definition of well-ordering property is given via a test which purposes to illustrate the range of good stitching addressed by the amount of the candidate inliers in OCICI; from the experimental results, we suggest that OCICI appears to have the well-ordering property.

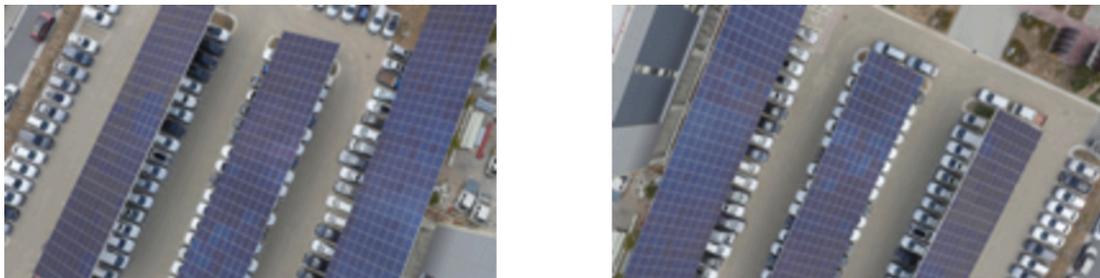
The remaining sections of this article are composed as follows: In Section 2, we introduce the proposed image registration algorithm, and in Section 3 the experimental results of the proposed method are presented; it is compared with two benchmark methods of RANSAC, one by the homographic mapping function and the other by the affine mapping function. In Section 4, we discuss several issues related to the proposed method and future works. In Section 5, we conclude the content of this study.

## 2. Method Description

This section introduces a comprehensive flow chart of the image stitching proposed in this study.

### 2.1 Feature extraction by SIFT

When two drone-based aerial images are given as shown in Fig. 1, for example,



(a) Reference image

(b) Sensed image

**Fig. 1.** Two drone aerial images

we obtain SIFT features as given in Fig. 2 where the red circles represent the detected features on the two; the SIFT feature descriptors are made up of 128-dimensional vectors, that are represented in green lattices according to their geometric orientation, following the idea in [8]. The two drone-aerial images will be matched using the given features. For the matching of features, the nearest rate of distance strategy used in [8] is applied with the preset threshold value  $\epsilon$  using Euclidean norm of the difference between SIFT feature vectors, of which the form of criteria is given as follows:

$$|\tilde{x} - \tilde{y}_2| / |\tilde{x} - \tilde{y}_1| < \epsilon \quad (1)$$

for SIFT feature descriptor  $\tilde{x}$  in the reference image, and  $\tilde{y}_1, \tilde{y}_2$  in the sensed image such that the values of  $|\tilde{x} - \tilde{y}_1|$  and  $|\tilde{x} - \tilde{y}_2|$  are the minimum and the second minimum ones;  $\tilde{y}_1$ 's are gathered to make the list of matches if Eq.(1) is satisfied.

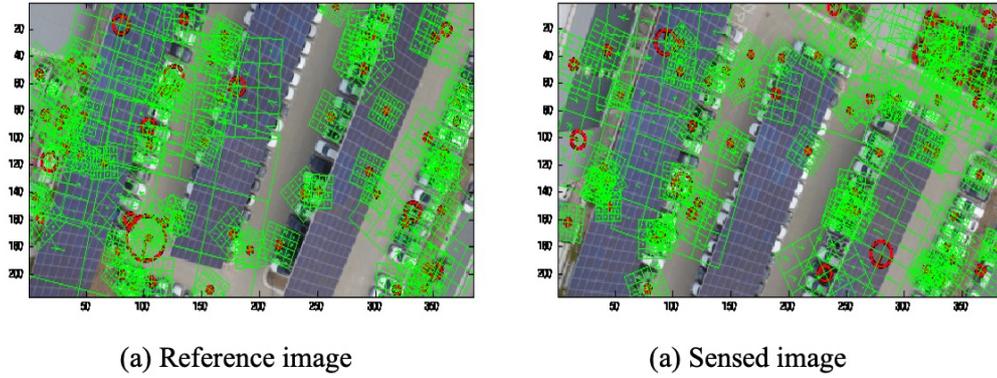


Fig. 2. SIFT feature points (center of the red circles) and descriptors (green lattice)

## 2.2 Hypothesis estimation

In this section, we investigate a technique for the hypothesis estimation, regarding the geometric congruence between the feature points using the characteristic of the affine map.

We sequentially take three distinct pairs of feature points and then apply the least-squares method in order to find the sequence of the linear operators. The set of the centers of the red spots on the sensed image, (b) of Fig. 2, is denoted by  $X$ , while  $Y$  denotes the corresponding set on the reference image, (a). Let  $\hat{x}$  be a triplet of the distinct three points in  $X$  and  $\hat{y}$  be the corresponding triplet in  $Y$  satisfying

- i)  $\hat{x} = \{x_l, x_m, x_n\}$ , where  $x_l, x_m,$  and  $x_n$  are two-dimensional column vectors which represent the geometric positions
- ii)  $x_l \neq x_m$  if  $l \neq m$

for distinct positive integers  $l, m,$  and  $n$ .

For a pair of triplet  $\hat{x}$  and  $\hat{y}$ , we determine the linear operator  $T(\hat{x}; \hat{y})$  in  $\mathbb{R}^{3 \times 3}$  which transforms  $\hat{x}$  to  $\hat{y}$  in the least-squares method for the following linear system:

$$\begin{bmatrix} x_l(1) & x_l(2) & 1 \\ x_m(1) & x_m(2) & 1 \\ x_n(1) & x_n(2) & 1 \end{bmatrix} * T = \begin{bmatrix} y_l(1) & y_l(2) & 1 \\ y_m(1) & y_m(2) & 1 \\ y_n(1) & y_n(2) & 1 \end{bmatrix} \quad (2)$$

where  $x_l = (x_l(1), x_l(2))^{\text{transpose}}$  (for the solution of the least-squares system, one may refer to [17]). Taking pairs of triplet  $\hat{x}$  and  $\hat{y}$  sequentially, we obtain a sequence of linear operators.

Then, from the sequence of the linear operators, we take some candidate optimal ones using the following criteria.

### A. The similarity of the geometric ratio

Let us assume the red points as the best matching three-feature pairs between the sensed image and the reference image in Fig. 3. For the best choice of the pairs of the triplets  $\hat{x}$  and  $\hat{y}$ , we consider the similarity of the triangles  $\Delta ABC$  and  $\Delta A'B'C'$  given in Fig. 3. Hence, as criteria for choosing the best option of the hypothesis of  $T$ , we impose that the ratio of the length of the corresponding sides of  $\Delta ABC$  and  $\Delta A'B'C'$  are the same, *i.e.*, by defining vectors  $\vec{a}$ ,  $\vec{a}'$ ,  $\vec{b}$ ,  $\vec{b}'$ ,  $\vec{c}$ ,  $\vec{c}'$  as the sides of  $\Delta ABC$  and  $\Delta A'B'C'$  in Fig. 3, the following relations hold.

$$\frac{|\vec{a}|}{|\vec{a}'|} = \frac{|\vec{b}|}{|\vec{b}'|} = \frac{|\vec{c}|}{|\vec{c}'|} \quad (3)$$

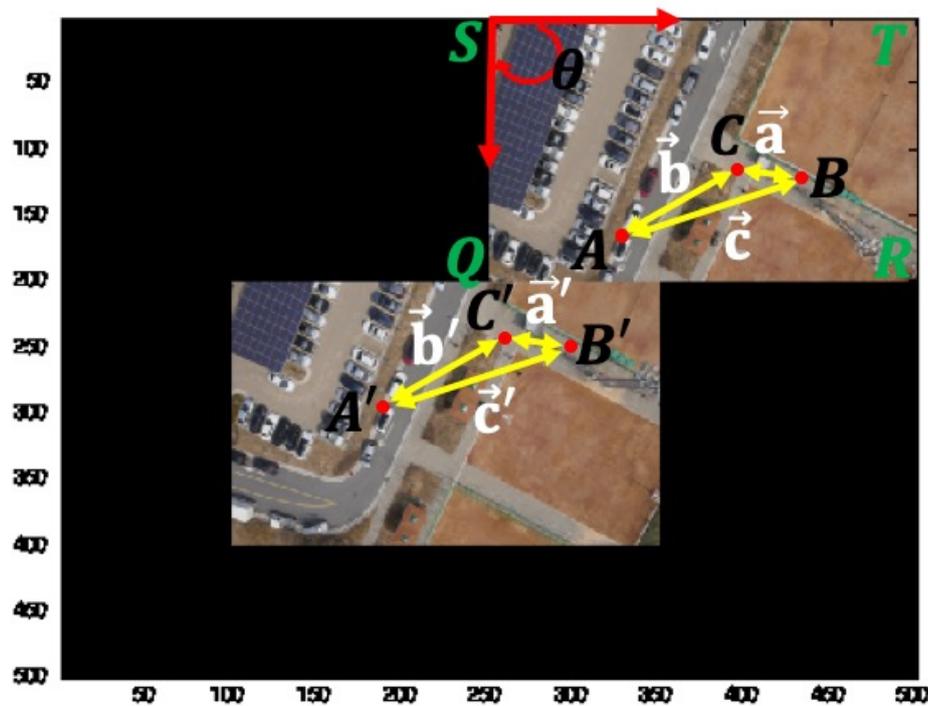


Fig. 3. Illustration of the similarity property of candidate matching feature pairs (the sensed image in upper-right, the reference image in bottom-left)

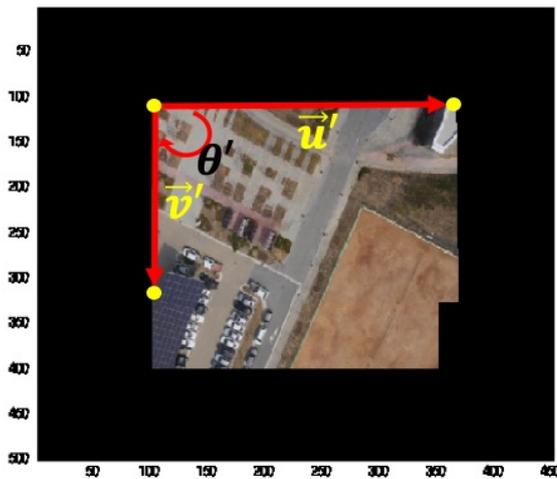
### B. The certification by the orthogonal corner angle

With each linear operator  $T$ , we transform the three corner vertices of the sensed image and measure the included angle between the two vectors, as depicted in Fig. 4. Among the outputs of the angle measurement, we search for the case that the two transformed corner vectors are most perpendicular mutually. This judgment is based on the fact that two given vectors should conserve the angle between them after the transformation by the optimal hypothesis of the affine operator  $T$  from the assumption given in Eq. (3). The judgment of whether the angle is approximate to the perpendicular or not is determined by the value of the inner product between the two corner vectors made by the transformed three corner vertices, *i.e.*, we reassure

the choice of the optimal operator  $T$  if it makes the inner-product value best to approximate to zero for the three transformed corner vertices.



(a) The angle between the two sides of the boundary edge before transformation



(b) The angle between the two sides of the boundary edge after transformation

Fig. 4. Corresponding triplet fitness test by inner-product for the angle measurement

The conditions to be satisfied by the two types of the feature matching criteria enumerated above can be switched to a minimization problem defined by the following measurements.

First, let us define that the length of each side of the triangle formed by the three feature points of the sensed image is  $|\vec{a}|$ ,  $|\vec{b}|$ ,  $|\vec{c}|$  and that the length of each corresponding side of the triangle formed by the three feature points of the reference image is  $|\vec{a}'|$ ,  $|\vec{b}'|$ ,  $|\vec{c}'|$  (see Fig. 3). Then, the function  $K(\hat{x}, \hat{y})$  for selecting the inliers of the feature pairs is given by:

$$K(\hat{x}, \hat{y}) = \left[ \text{abs} \left( \frac{|\vec{a}|}{|\vec{a}'|} - \frac{|\vec{b}|}{|\vec{b}'|} \right) + \text{abs} \left( \frac{|\vec{b}|}{|\vec{b}'|} - \frac{|\vec{c}|}{|\vec{c}'|} \right) + \text{abs} \left( \frac{|\vec{c}|}{|\vec{c}'|} - \frac{|\vec{a}|}{|\vec{a}'|} \right) \right] / \left( \frac{|\vec{a}|}{|\vec{a}'|} \right) \quad (4)$$

where  $\text{abs}(\cdot)$  makes the sign positive.

Second, let  $\vec{u}$  and  $\vec{v}$  be represented as the adjacent vectors forming the side margins of the sensed image, and also let  $\vec{u}'$  and  $\vec{v}'$  be of the reference image (see Fig. 4). Let  $\hat{u}$  and  $\hat{v}$  be calculated as follows.

$$\hat{u} = [\vec{u}, 1] * T, \hat{v} = [\vec{v}, 1] * T \quad (5)$$

Then, using Eq. (5), as a constraint function for qualifying the inliers quantized by  $K(\hat{x}, \hat{y})$  in Eq.(4) we define  $\theta(T(\hat{x}; \hat{y}))$  such that

$$\theta(T(\hat{x}; \hat{y})) = |\hat{u}(1)\hat{v}(1) + \hat{u}(2)\hat{v}(2) - \vec{u}' \cdot \vec{v}'| \quad (6)$$

Now, in multiplying the extra adjustment coefficient  $\alpha$  and  $\rho$ , we define the total measurement function  $J(\theta, K; \hat{x}, \hat{y}, T)$  such that

$$J(\theta, K; \hat{x}, \hat{y}, T) = \alpha\theta(T(\hat{x}; \hat{y})) + \rho K(\hat{x}, \hat{y}) \quad (7)$$

Then, we are faced with the minimization problem:

Determine the optimal components  $(\hat{x}, \hat{y}, T)$  satisfying

$$\min_{\hat{x}, \hat{y}, T} J(\theta, K; \hat{x}, \hat{y}, T) \quad (8)$$

Note that, even if the optimal operator  $T$  which satisfies the model of Eq. (8) has been found, it does not automatically mean that the operator  $T$  is the best one; since the model of the criteria made of up to Eq. (8) is devised to collect some candidate good initial inliers for the sample consensus method to gain the best one of the mapping operator of the stitching, *i.e.*, some pairs of the triplets  $\{\hat{x}, \hat{y}\}$  which approximate the minimizing function of Eq. (8) in the increasing order of the value of  $J$  are supposed as the candidate initial inliers to be fed into the sample consensus method.

### 2.3 Pseudo codes of the proposed algorithm

We present the pseudo-codes of our proposed algorithm.

#### A. Step 1. - Feature matching by Euclidean distance

Here we define the notations; we define the sequence of the feature descriptors assigned to the feature points (the geometric positions of the features) of the sensed image as  $\tilde{X} = \{\tilde{x}_j\}_{j=1}^N$  for the sequence of the feature points  $\{x_j\}_{j=1}^N$ , and the sequence of the feature descriptors of the reference image as  $\tilde{Y} = \{\tilde{y}_j\}_{j=1}^N$  for the sequence of the feature points  $\{y_j\}_{j=1}^N$ . Then, for the first step, for the matching process, we define empty sets  $X_c$  and  $Y_c$ , and obtain the matching pairs constrained by a given constant threshold  $\epsilon$  by the criteria given in Eq. (1).

#### B. Step 2. - Estimation of the affine map and the calculus of the function $J$

We define the notations; we define the sequence of the feature pairs obtained by Step 1 as  $X_c = \{x_k\}_{k=1}^M$  and  $Y_c = \{y_k\}_{k=1}^M$  for the sensed and reference images, respectively. We also organize an ordered set of the unordered triples  $\hat{X} = \{(x_l, x_m, x_n)_j: x_l, x_m, \text{ and } x_n \in X_c \text{ are mutually distinctive, } 1 \leq j \leq \binom{M}{3}\}$  and  $\hat{Y} = \{(y_l, y_m, y_n)_j: y_l, y_m, \text{ and } y_n \in Y_c \text{ are mutually distinctive, } 1 \leq j \leq \binom{M}{3}\}$ . Fig. 5 presents the estimation of  $T_j$  for  $\hat{x}_j = (x_l, x_m, x_n)_j \in \hat{X}$  and  $\hat{y}_j = (y_l, y_m, y_n)_j \in \hat{Y}$  and the calculus of the function  $J(\theta, K; \hat{x}_j, \hat{y}_j, T_j)$  (see Pseudo-Code 1 in Fig. 5).

**Pseudo-Code 1:** The estimation of  $T_j$  and the calculus of the function  $J(\cdot)$

```

For  $L = \binom{M}{3}$ ,
1: For  $j = 1$  to  $L$ 
2:   Estimate  $T_j$  satisfying Eq.(2)
3:   Using  $T_j$ , calculate  $J(\theta, K; \hat{x}_j, \hat{y}_j, T_j)$  defined in Eq.(7)
4:   Save as  $J_j = J(\theta, K; \hat{x}_j, \hat{y}_j, T_j)$ 
5: End

```

**Fig. 5.** Pseudo-code for the estimation of  $T_j$  and the calculus of function  $J$

#### C. Step 3. - Determine the hypothesis of the affine map from candidates

Finally, among  $J_j$  measured in Step 2, we choose minimal  $r$  values in increasing order, starting from the least one of  $\{J_i\}_{i=1}^r$ . Then, for those corresponding candidate  $T_j$  estimated from Step 2 such that  $J_j = J(\theta, K; \hat{x}_j, \hat{y}_j, T_j)$  and  $J_j \in \{J_i\}_{i=1}^r$ , we evaluate for a whole set of  $X_c$

(of sensed images); here the set of pairs of triples  $\{\hat{x}_j, \hat{y}_j\}_{i=1}^r$  we call the optimally chosen initial candidate inliers (OCICI). Additionally, if the difference between the evaluated output  $T_j(x_k)$  for  $x_k \in X_c$  and the corresponding expected output  $y_k \in Y_c$  is less than the predefined constant threshold  $\delta$ , we give one count of ‘credit’ to that estimated  $T_j$ . Finally, among those candidates, the estimated  $T_j$  which has the highest credit counts is finetuned as the hypothesis of the resulting affine map (see Pseudo-Code 2 in Fig. 6).

**Pseudo-Code 2:** Finding the resulting hypothesis of the affine map from candidates

```

1: Among  $\{J_j\}_{j=1}^L$ , Choose  $r$  minimal ones in increasing order from the least one such as  $\{J_i\}_{i=1}^r$ .
2: Find the candidate  $\{T_j\}_{j=1}^r$  estimated from Step 2 such that  $J_j = J(\theta, K; \hat{x}_j, \hat{y}_j, T_j)$  and  $J_j \in \{J_i\}_{i=1}^r$ .
3: For j = 1 to r
4:   Create variable  $C_j$ 
5:   For k = 1 to L
6:     Evaluate  $T_j(x_k)$ 
7:     If  $|T_j(x_k) - y_k| < \delta$ 
8:        $C_j = C_j + 1$ 
9:     End
10:  End
11: End
12: Find  $T_j$  which has the highest value of  $C_j$  and set  $T = T_j$ .
13: Create variable  $C_{old} = -1$  and  $C_{new} = 0$ 
14: While ( $C_{old} < C_{new}$ )
15:    $C_{old} = C_{new}$ 
16:    $C_{new} = 0$ 
17:   Create empty sets  $x = \emptyset$  and  $y = \emptyset$ 
18:   For k = 1 to L
19:     Evaluate  $T(x_k)$ 
20:     If  $|T(x_k) - y_k| < \delta$ 
21:        $C_{new} = C_{new} + 1$ 
22:       Put  $x_k$  in  $x$ 
23:       Put  $y_k$  in  $y$ 
24:     End
25:   Estimate  $T$  with  $x$  and  $y$ 
26:   End
27: End
28: Set  $T$  as the hypothesis of the mapping function.

```

**Fig. 6.** Pseudo-code for determining the resulting hypothesis of the affine map from candidates

### 3. Experimental results

#### 3.1 Data composition of the drone-aerial images

The data was obtained from the company - Korea plant service and engineering (KPS) of the Korea electric power corporation (KEPCO). The drone aerial images of this study were produced from the specific area adjacent solar photovoltaic panels where the images were photographed so that at least 50 to 70% of the areas of adjacent images were overlapping for the efficiency of feature matching by SIFT descriptor [4, 6, 18]; The images taken by the drone are made up of vertical photographs for easy panoramic stitching.

On the other hand, in the numerical experiments, we test for 63 pairs of images; each image has a resolution of  $2160 \times 3840$ . In order to improve the experimental computation speed, we reduce the resolution of the original image to  $200 \times 250$ . In Table 1, the composition of the images for the application is listed where 38 in total made by 7, 7, 8, 8, and 8 is applied in the stitching for the 5 courses of the drone’s flight paths, and also, the images of the garbage flights such as shots taken during take-off, landing, and short- or long-around turning are included. For reference, the drone flight height ranges 60~100m from the ground by the scale of the photograph. The illustration of the drone’s flight paths is given in Fig. 7.

**Table 1.** The composition of the drone-aerial images applied

Course	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	Garbage	Total
Amount	7	7	8	8	8	26	64

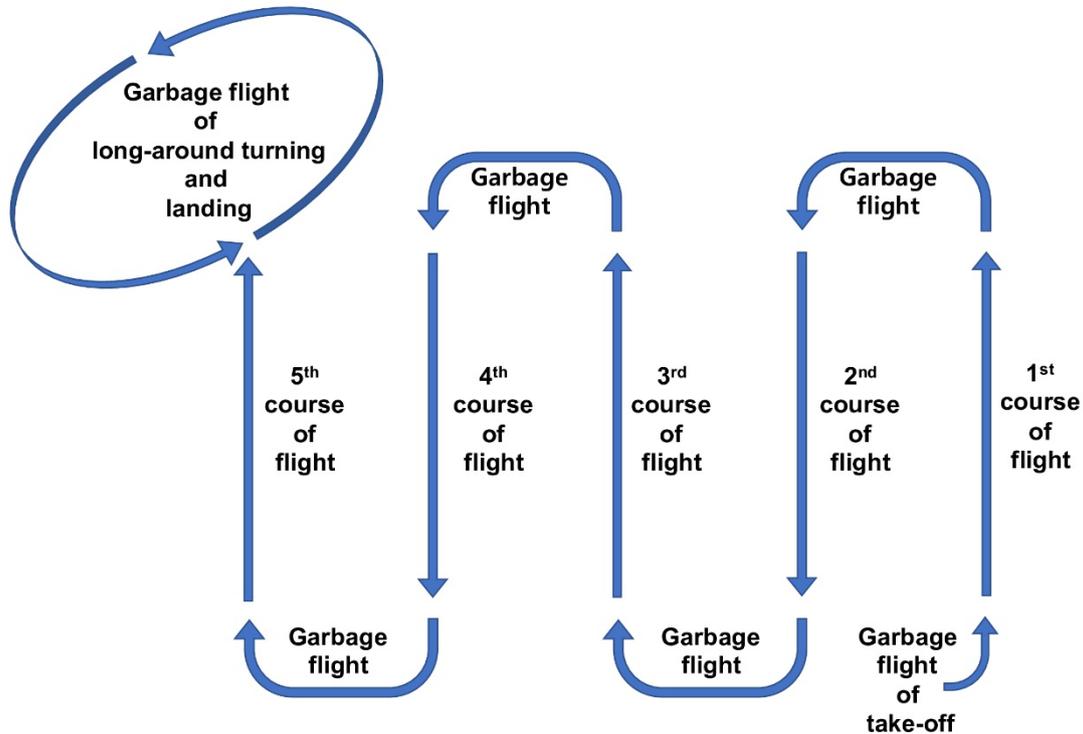


Fig. 7. Drone's flight paths in 5 courses

### 3.2 Experimental result of the proposed method for the drone-aerial images

In this section, we investigate a test of validation for the proof of the effectiveness of the proposed method, for which we examine a test to define a characteristic so-called well-ordering property of the OCICI (from now on, we denote the proposed method as OCICI, for the brevity). Next, we compare the efficiency of the stitching based on the experimental results obtained through the proposed method versus benchmark methodologies of the standard RANSAC. For the actual test of the performance of RANSAC, we have randomly given the initial guess of the hypothesis 10 times for each stitching of images, and we estimate the statistical results in the form of box-plot. Note that, for the discretization and rounding error which can make holes *and/or* overlaps in the output image, we employ the backward transform of the estimated mapping function in the resampling and transforming of the sensed images [3]. The numerical simulation has been implemented by Matlab R2014a. Notice that the SIFT and its feature descriptor implementation in our experiments is given from the VLFeat package [19, 20].

#### A. Well-ordering property of OCICI

For the actual use of OCICI estimating the hypothesis of the mapping function, it needs to be certified that the choice of the initial inliers is optimal in the real problem. In our experiment to prove the optimality of the initial choice of inliers, two ways of proof of test are carried on as follows.

- ① Enlarging the amount of the candidate inliers from the first pair of features which is proved as the smallest one by the proposed measure of geometric congruent similarity introduced in the previous section; in abbreviation, we call this test as EAC (enlarging the amount of the candidate) test.
- ② Discarding the amount of the candidate inliers from the first pair of features, hence the hypothesis estimation is carried on with the remaining pairs; we call this test as DAC (discarding the amount of the candidate) test.

During the whole experiment, the adoption of the parameters  $\theta, K$  found in the measure in Section 2.2 is given as  $\theta=1$  and  $K=1$ ; the threshold value  $\epsilon$  of the SIFT defined in Section 2.1 is set as  $\epsilon=1.2$ ; the total of the images applied to be stitched pairwise is 38 for the region of interest (ROI) where 7, 7, 8, 8, and 8 are given for, from first to fifth flight, respectively. The illustrations of the results of the above tests are plotted in Figures. For the measure of the difference of the output stitching of images, a norm is defined as follows.

$$|I_1 - I_2| = |I_1^r - I_2^r| + |I_1^g - I_2^g| + |I_1^b - I_2^b| \tag{9}$$

where the sup-scripts  $r, g,$  and  $b$  denote the red, green, and the blue channels of the given color images  $I_1$  and  $I_2$ , respectively;  $|\cdot|$  represents the Euclidean norm.

In Fig. 8, the result of the EAC test is given as the plot where the error which represents the actual difference of pixel values compared to the manual stitching of images applied in the tests; the manual stitching is set by the gathering of inliers composed of the geometric corner points judged by human eyes. Seeing the plot in the top in Fig. 8 which represents the mean of error, as the value of EAC test in x-axis increases, the actual error diminishes to a stable state; the bottom represents the boxplot of the results of error along with the number of candidates denoted by  $N$ .

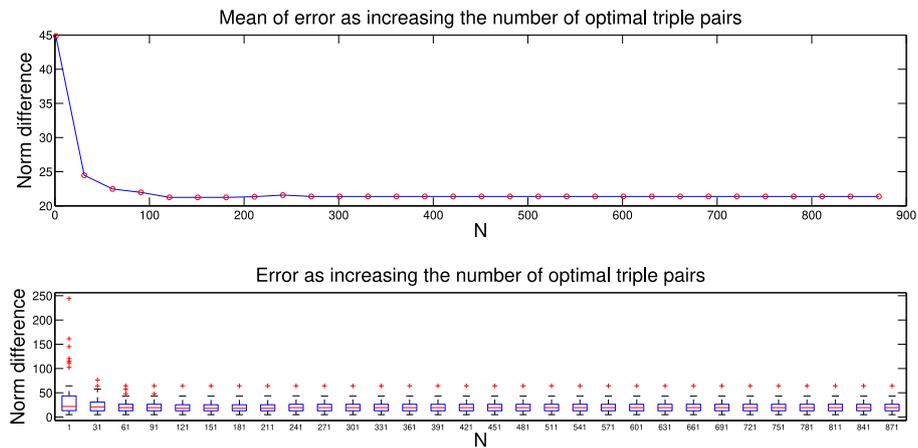


Fig. 8. Plots of results of EAC tests for the drone-aerial images

In Fig. 9, the results of the DAC tests are given as plots. Seeing the plot in the top in Fig. 9 which represents the mean of error, as the value of DAC test in x-axis increases, the actual error diverges; the bottom represents the boxplot along with the number of candidates denoted by  $K$ .

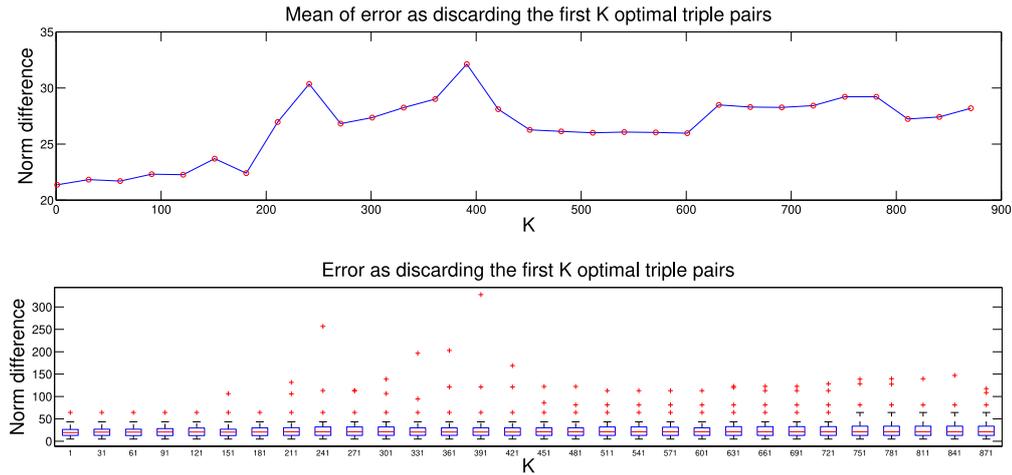


Fig. 9. Plots of results of DAC tests for the drone-aerial images

In overall view of the above certification tests, we have found out that the inclusion and exclusion tests of the candidate initial inliers of the feature pairs with small values in the measure  $J$  in (7), which are evaluated by the plots of results in the EAC and DAC tests, have well proved of the optimality of the measure  $J$ , by experiment. Hence, we define the OCICI performs having well-ordering property when the short range of the candidate inliers gives a good stitching output.

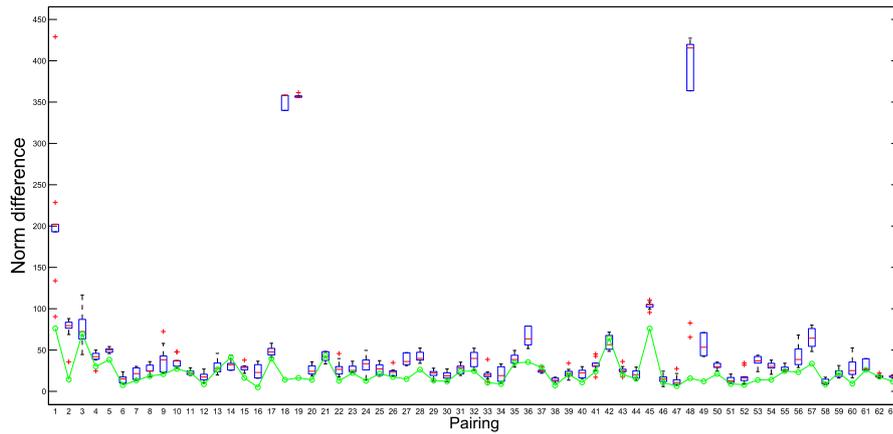
However, since the test of the well-ordering property still has to be proved, for the lack of the actual comparison results, we conduct several benchmark experiments versus models of RANSAC as follows.

## B. Experimental result of the proposed method vs. RANSAC of homography model

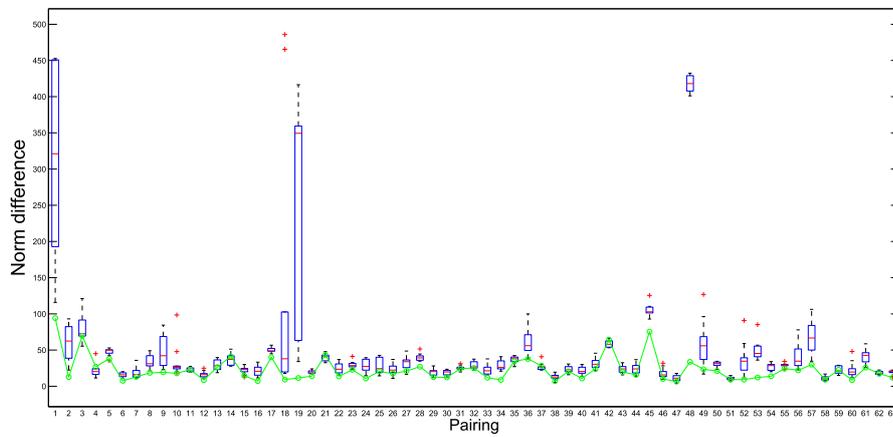
For the benchmark comparison of experiments, we compare the result of OCICI, with standard RANSAC of homography approximation (H-RANSAC) of which the implementation is given from [19].

The experiment is carried out by varying the threshold value  $\epsilon$  in Eq. (1) from 1.1 to 1.5 with 0.1 gaps, and the candidate inliers of OCICI are set in 30, 100, 200, 400, 600, and 800; the number of random choice of RANSAC of H-RANSAC is set in up to 3000. From the experimental results, we have found that the stitching performance of OCICI looks best at threshold 1.2 with 600 candidate inliers and threshold 1.1 with 800 candidate inliers; the results have been improved as the threshold values increased. However, as for the H-RANSAC, fatal miss stitching has occurred at all the thresholds defined above. In Fig. 10 and 11, we render the results between the two methods. In Fig. 10, the green line-plot represents the pixel error between OCICI at threshold 1.2 with 600 candidate inliers and the

manual solution for each pair of images; the box plots represent the error between the manual solution and the H-RANSAC with 10 trials of stitching at threshold 1.2 for each pair. In **Fig. 11**, the green line-plot represents the pixel error between OCICI at threshold 1.1 with 800 candidate inliers and the manual solution for each pair of images; the box plots of the error between the manual solution and the H-RANSAC with 10 trials of stitching at threshold 1.1.



**Fig. 10.** Plots of the error of OCICI (green-plot) vs. H-RANSAC (box plot) at  $\epsilon=1.2$



**Fig. 11.** Plots of the error of OCICI (green-plot) vs. H-RANSAC (box plot) at  $\epsilon=1.1$

From the experimental results, we see that OCICI stitching performs better than H-RANSAC at those given parameters for the estimation. In Fig. 12, the picks appeared in Fig. 10 and 11 made by H-RANSAC are depicted by actual stitching results; the 18<sup>th</sup>, 19<sup>th</sup>, and 48<sup>th</sup> stitching pairs by H-RANSAC are compared to those of OCICI, at threshold 1.2. Also, in Fig. 13, a course of flight view by both method of stitching of images is given, for a large area stitching performance comparison.

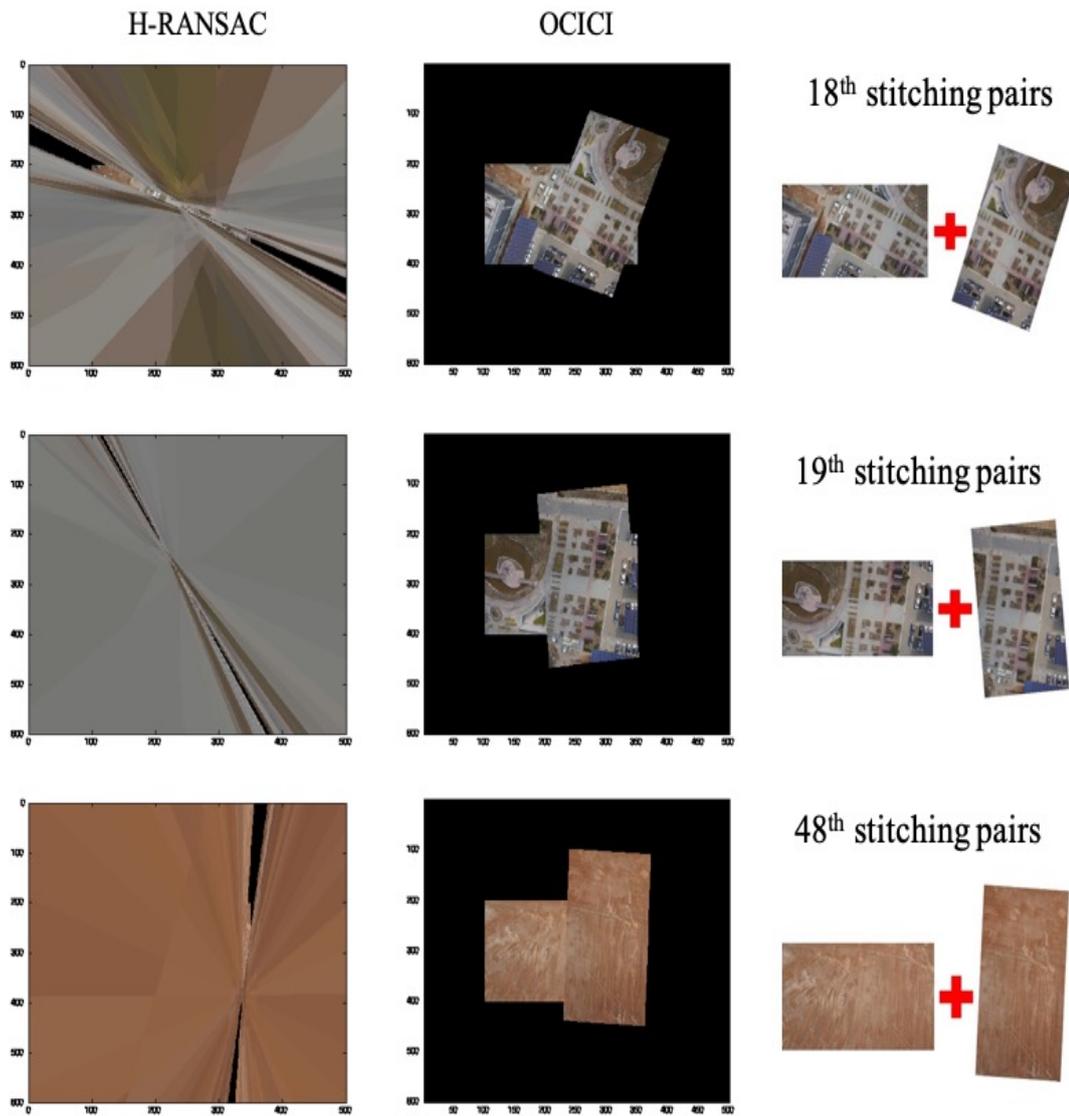
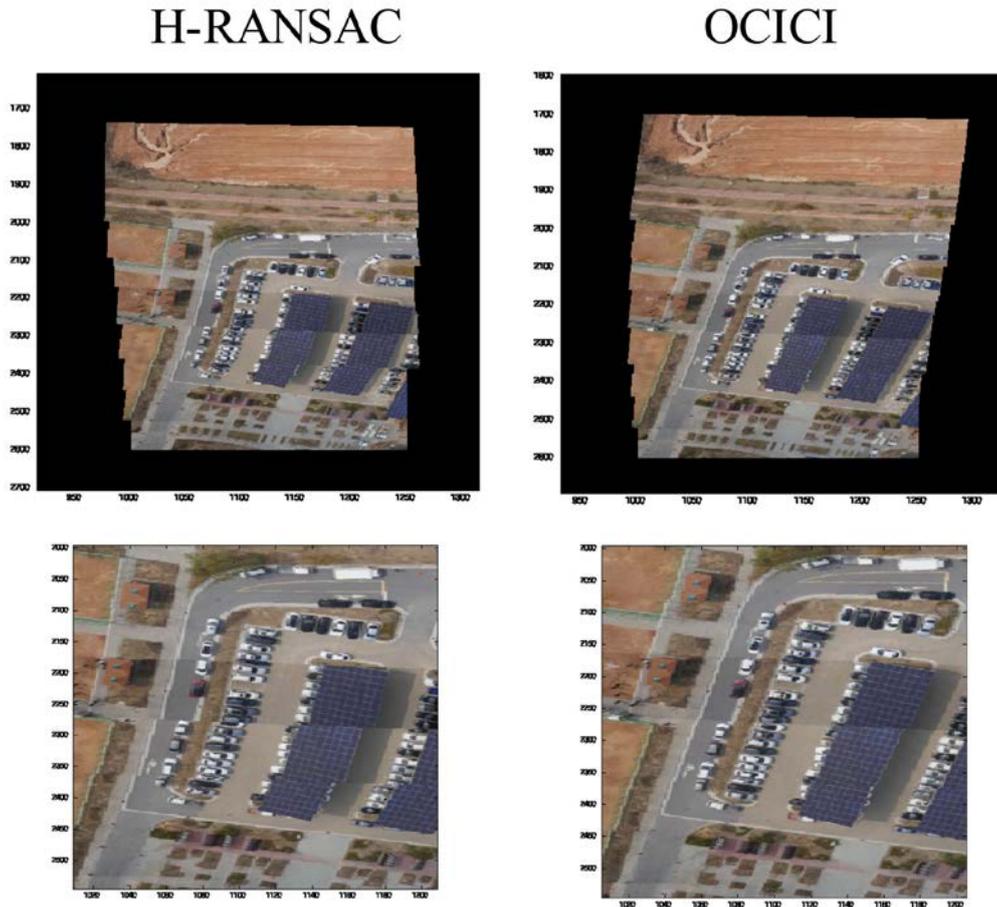


Fig. 12. Actual results compared between both methods at  $\epsilon=1.2$ , for the 18<sup>th</sup>, 19<sup>th</sup>, and 48<sup>th</sup> stitching pairs



**Fig. 13.** A course of flight view by both method of stitching of images (H-RANSAC in left-side and OCICI in right-side)

### C. Experimental result of the proposed method vs. RANSAC of affine-map model

As another benchmark comparison of experiments, we compare the result of OCICI with standard RANSAC of the affine-map approximation (A-RANSAC) of which the implementation is given as similar as that of [19] except the estimation of the affine map, the mapping function, is iteratively calculated by newly added inliers starting from the initial guesses randomly fed for each trial.

In Fig. 14, the green line-plot represents the pixel error between OCICI at threshold 1.2 with 600 candidate inliers and the manual solution for each pair of images; the box plots represent the error between the manual solution and the A-RANSAC with 10 trials of stitching at threshold 1.2 for each pair, with 600 initial guesses of the triple pairs for each trial; as shown in the plots, the results of both methods do not vary in the significant difference. In Fig. 15, the green line-plot represents the pixel error between OCICI at threshold 1.1 with 800 candidate inliers and the manual solution for each pair of images; the box plots of the error between the manual solution and the A-RANSAC with 10 trials of stitching at threshold 1.1, with 800 initial guesses of the triple pairs for each trial; we see that OCICI stitching performs better than A-RANSAC at threshold 1.1.

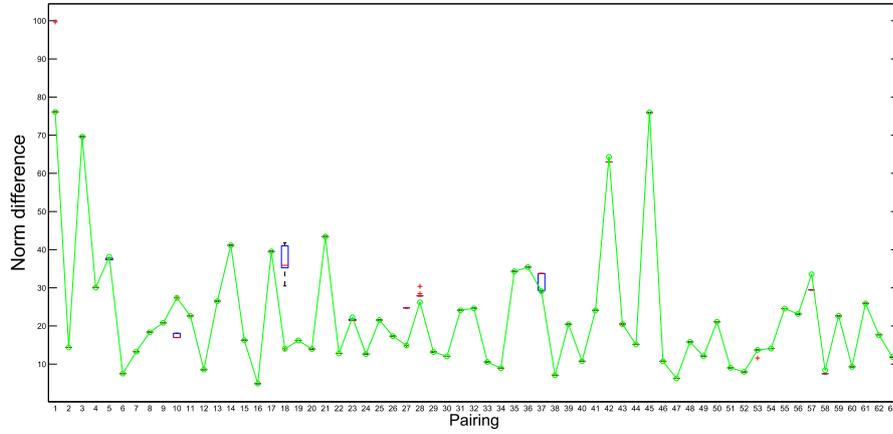


Fig. 14. Plots of the error of OCICI (green-plot) vs. A-RANSAC (box plot) at  $\epsilon=1.2$

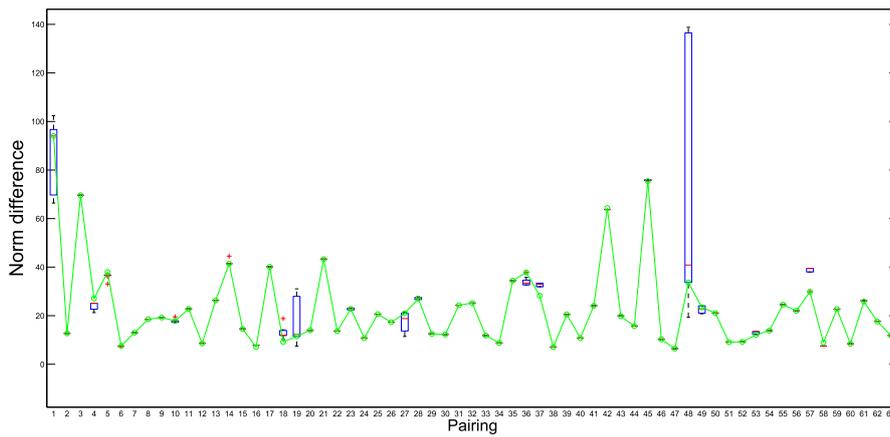


Fig. 15. Plots of the error of OCICI (green-plot) vs. A-RANSAC (box plot) at  $\epsilon=1.1$

In Fig. 16, the picks appeared in Fig. 15 made by A-RANSAC are depicted by actual stitching results; the 48<sup>th</sup> stitching pairs by A-RANSAC are compared to those of OCICI, at threshold 1.1.

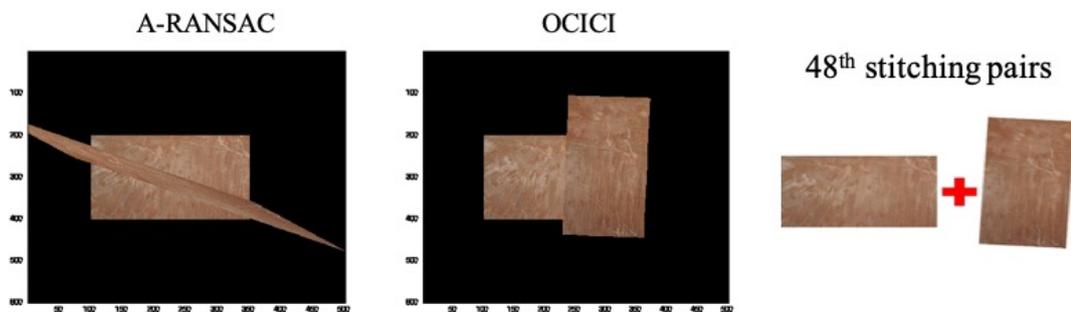


Fig. 16. Actual results compared between both methods at  $\epsilon=1.1$ , for the 48<sup>th</sup> stitching pairs

### 3.3 Experimental results of miscellaneous stitching

#### A. Tearing-and-stitching test

In this section, we present experimental results of some miscellaneous stitching of images. The experiments constitute of so-called the tearing-and-stitching job; for a given image, by tearing off in an arbitrary direction we obtain a separate two images, and as we have the stitching of the teared-off images to be bound again, the tearing-and-stitching is done. For the example images, we gather several cut-and-cropped images. Then, we compare the stitching performance between the methods of H-RANSAC, A-RANSAC, and OCICI. In these examples, we set the candidate initial inliers 3 in OCICI and the number of random choice of RANSAC of H-RANSAC 3000, 3 random initial choice of guesses for A-RANSAC. In Fig. 17, the actual graphical output of stitching is compared between H-RANSAC and OCICI at the matching threshold 1.5; the red box is indicating the spot where the stitching performance of both methods is well compared.

In our examples of the tearing-and-stitching tests, the OCICI outperforms the H-RANSAC; the acknowledgeable point of the merit is that the OCICI method performs excellently for most of the given images for the tearing-and-stitching tests with just 3 candidate inliers. On the other hand, in Fig. 18, it is shown that OCICI with 3 candidate inliers outperforms A-RANSAC with 3 initial random guesses where the matching threshold 1.2 is taken for both methods; from our experiments, as the threshold  $\epsilon$  decreases, the quality difference between the results of both methods tends to increase (in the Figure, fatal miss stitching is shown, given by A-RANSAC).

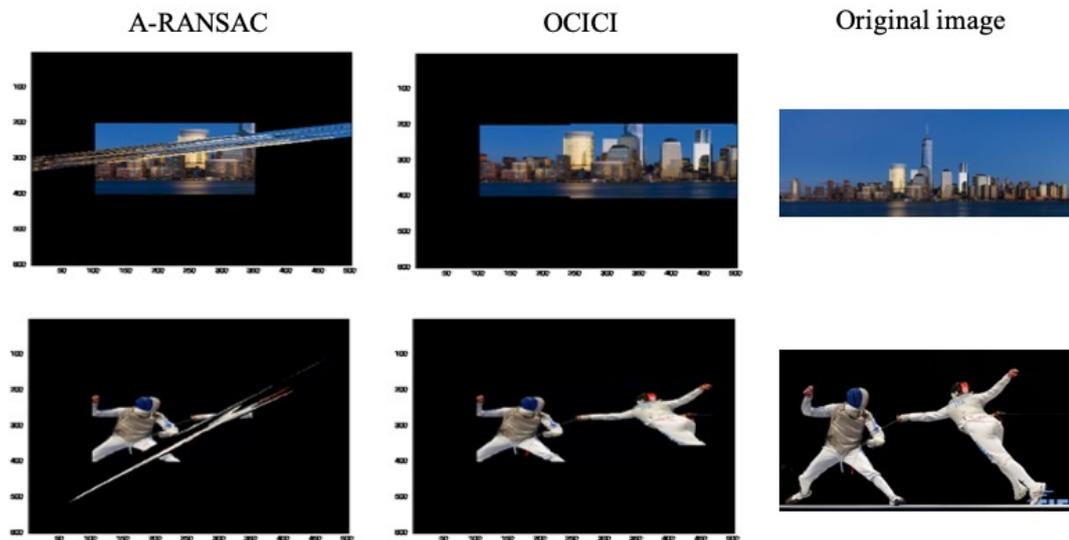


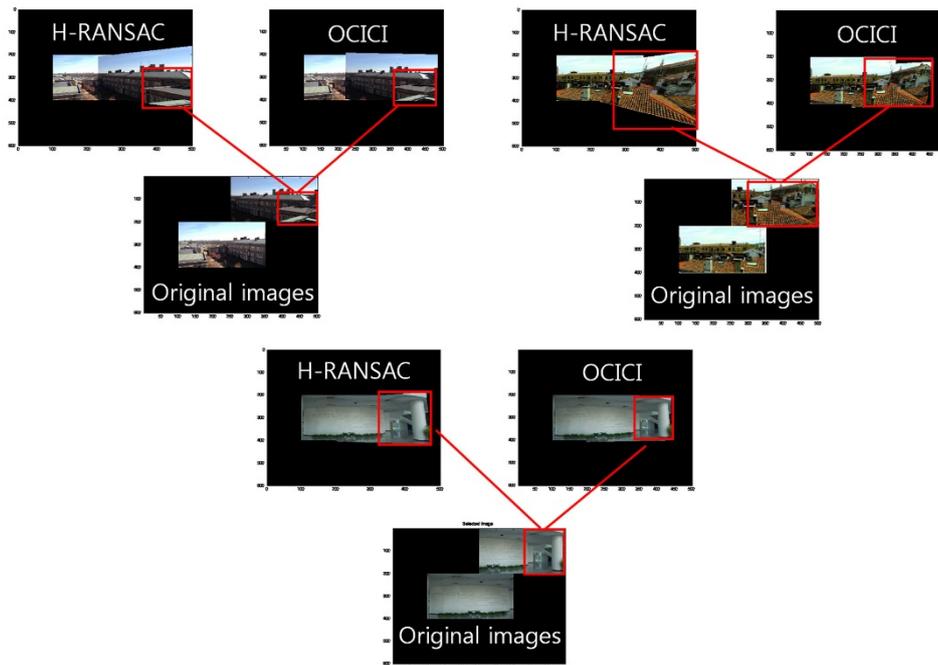
Fig. 18. Actual graphical output of stitching of A-RANSAC vs. OCICI, for  $\epsilon=1.2$



Fig. 17. Actual graphical output comparison of stitching (H-RANSAC in 1<sup>st</sup> column and OCICI in 2<sup>nd</sup> column), for  $\epsilon=1.5$

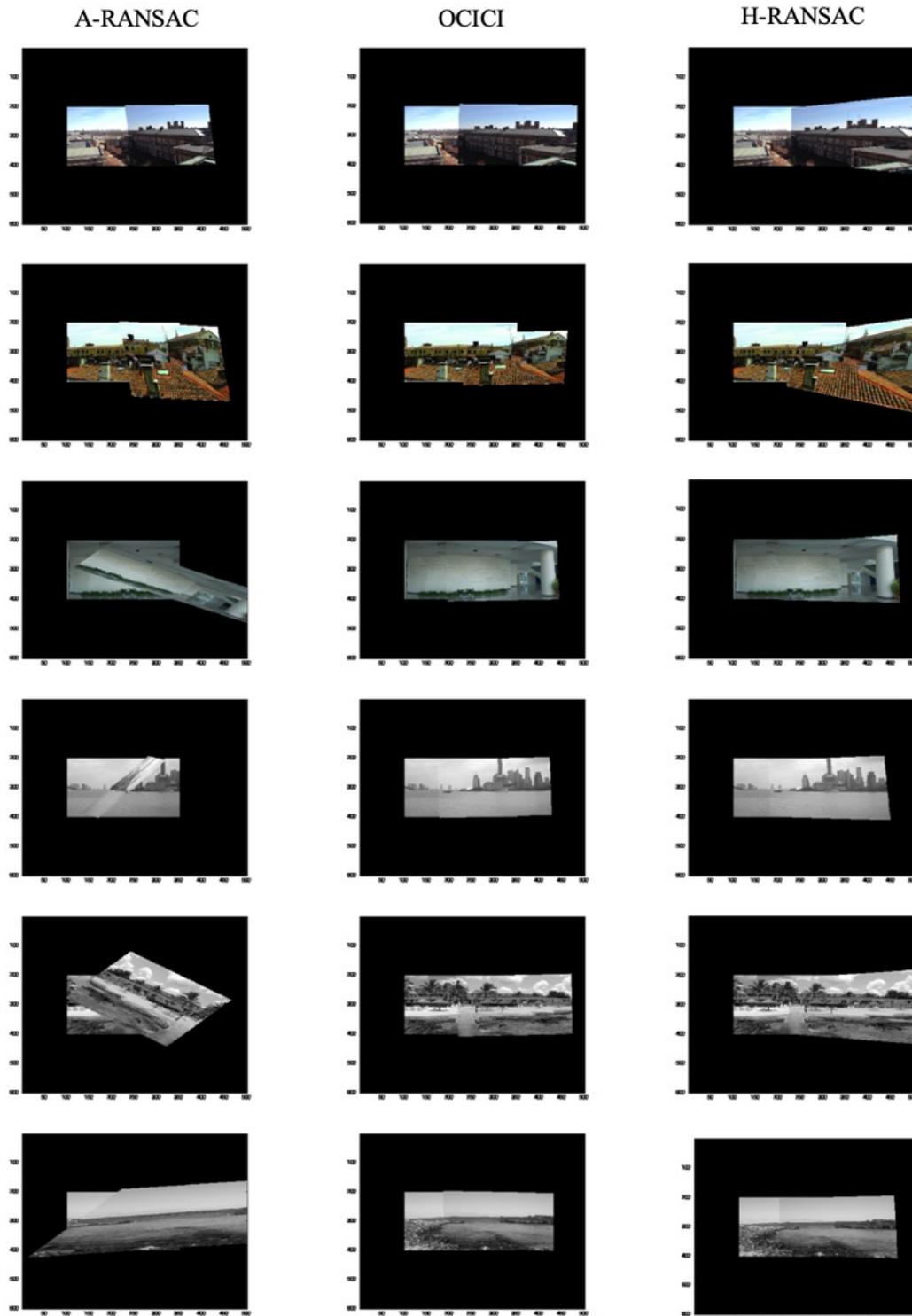
### B. Tests by other general image data

In this section, we present some other miscellaneous stitching results by a general image data; here we apply the methods A-RANSAC, OCICI, and H-RANSAC for several images including those of Adobe Panoramas Dataset which is tested in [19]. As in the previous example tests, we set the candidate initial inliers 3 in OCICI and the number of random choice of RANSAC of H-RANSAC 3000, and 3 random initial choices of guesses for A-RANSAC.



**Fig. 19.** Actual graphical output comparison of stitching for Adobe Panoramas Dataset (H-RANSAC in 1<sup>st</sup> column and OCICI in 2<sup>nd</sup> column), for  $\epsilon=1.5$

In the tests, the overall stitching performance looks equivalent between H-RANSAC and OCICI or the H-RANSAC performs slightly better in the graphical naturality, especially, for images having the quantity of perspective. However, for images having the perspective view, the H-RANSAC makes warped parts in sensed image; the red boxes compared in **Fig. 19** represent the warped area in the test output images, indicating that the homographic transform of H-RANSAC which performs in better naturality for the perspective images may have inferiority compared to the rigid-affine transform of OCICI in the sense of warping images; the red boxes denote the regions where the distinctive warped areas have occurred as the results of the mapping of the sensed images resampled by H-RANSAC.



**Fig. 20.** Actual graphical output comparison of stitching for Adobe Panoramas Dataset (A-RANSAC in 1<sup>st</sup> column, OCICI in the 2<sup>nd</sup> column, and H-RANSAC in the 3<sup>rd</sup> column), for  $\epsilon=1.2$

In **Fig. 20**, we compare the actual stitching performance between the methods of A-RANSAC, OCICI, and H-RANSAC. In these examples, we set the candidate initial inliers 3 in OCICI, 3 initial guesses in A-RANSAC, and the number of random choice of RANSAC of H-RANSAC 3000, at the matching threshold 1.2; as the threshold  $\epsilon$  decreases, the quality difference between A-RANSAC and OCICI tends to increase (for A-RANSAC, in 10 trials at lots of the examples fatal miss stitching happened as shown in the Figure).

## 4. Discussion

In our experiments of the stitching of the drone-based images, we tested for 63 pairs of images which are obtained by a drone inspecting certain PV panel artifacts. In applying the feature-based method, the SIFT feature descriptor was used, because the SIFT method has invariant properties on the scale and has proven that it still maintains high repeatability for projection transformations within a certain range. In applying the images, in order to improve the experimental computation speed, we reduced the resolution of the original image to  $200 \times 250$  (although the aspect ratio of the reduced images is not the same as that of the original images, from our experiments, we have found out that the SIFT descriptor functioned properly to compare the features, except that the images were reduced by too much; for the images reduced to be  $135 \times 240$ , although the aspect ratio is equal to that of the original images ( $2160 \times 3840$ ), from our preliminary experiments, the quality of stitching tended to be significantly deteriorated). For the proposed method (OCICI) of stitching images, the sample consensus method was applied in order to identify the correct hypothesis of the mapping function given by the  $3 \times 3$  affine transforming matrix. In order to identify the hypothesis, we experimentally compared the performances between the proposed method and RANSAC by homographic transform (H-RANSAC) and affine transformation (A-RANSAC); for H-RANSAC the number of feeds of random initial guess for RANSAC is set 3000 (from our preliminary test, over 3000 the actual performance was not improved any more).

The OCICI is intuitively developed by considering the sensed image's characteristics, such as the image's angle and the geometric congruence ratio between the feature points in the images; for the first-step matching which selects the candidate true SIFT feature pairs we have adapted the parameter  $\epsilon$  appeared in Eq. (1), which is set to be tested from 1.1 to 1.5; for the estimation of the hypothesis and the quantification of the inlier, we set the threshold parameter  $\delta$  appeared in Pseudo-Code 2 as  $\delta=3$  (see **Fig. 6**). For the parameters related to  $\epsilon$  and  $\delta$ , the best choice of the measurement may be controversial for each problem, and for this problem, one may refer to [2].

The performance of RANSAC may be sensitively affected by the initial guess fed into the calculus of finding the inliers, and furthermore, seeing the results in **Fig. 10, 11** and **12** (by H-RANSAC), and **Fig. 15** and **16** (by A-RANSAC), for several sample pairs of test images, the methods of RANSAC has made several fatal false stitching. Instead, OCICI has been proven to perform relatively stable robustness among all the test images at the given set of parameters. From this, we found out that OCICI could provide a solution to such a problem that might not be solved by the RANSAC-like sample consensus method; moreover, from the EAC and DAC tests, we have well proved of the optimality of the measure  $J$ , by experiment, from which, hence, we define the OCICI performs in having well-ordering property.

On the other hand, among the combinations of the parameter functions given in Eq. (7) the scale of the coefficients in the model had not yet been sufficiently investigated. We still keep in mind the possibility of other scales of the coefficients so that the performances of the proposed method be worsened or improved compared to others; only in our experiments, even

from the most straightforward combination of the parameter functions and their coefficients it shows an excellent performance, and so it has proven its convenient utility.

For the miscellaneous example of stitching tests, we have done the tearing-and-stitching tasks. In the examples, we found out that the flat-images such as made by cropping given image to be stitched are better stitched by OCICI than the methods of RANSAC (see Fig. 17 and 18). For the test made of those including Adobe image data, the H-RANSAC outperforms, especially, for the perspective images in the graphical naturalness. However, for the warping property in the direction of the perspective, the H-RANSAC is not guaranteed as the best choice of method to be applied in the task of stitching of the drone-aerial images, since the drone-aerial images have been taken in the vertical focus to reduce the perspective of the aerial view for the easiness of the panoramic stitching.

As for the range of the excellent stitching expressed in the amount of the candidate inliers, for the drone-aerial images it would be said that the range of the stitching of OCICI is set in 600 and it functions well, and for the miscellaneous stitching of images the range is set in just 3 and it is enough to perform well, while the range of the random trials of RANSAC set in 3000 for H-RANSAC does not satisfy the full success of stitching for the given drone-aerial images (see Fig. from 10 to 12). Now, nevertheless, an issue that may arise in the study of OCICI is how the range of stitching could be determined in the use of anonymous images. From our approach of the concept of the well-ordering property which is investigated in the tests of the EAC and DAC test, we could make a rigorous but natural suggestion for that issue; if there are given an ideally large amounts of images to be stitched as much as possible, the EAC and DAC tests should be done and as the diameter where the sharpest slopes in the plot of the EAC and DAC tests happen is searched; then the actual range of the stitching may be determined with several tests in problems to know if the stitching of OCICI performs well for arbitrary pairs of images. Of course, if this suggestion to find the globally optimizing diameter would be realized, it might be the real optimal methodology of the model of OCICI.

For the computation cost, in Fig. 21, we give the result of the ratio of the elapsed time taken between the methods implemented in the serial programming by Intel(R) Xeon(R) 5160 CPU with 3.00 GHz and 24GB RAM in 64bit Windows OS for the examples in Section 3.2.B and C; the plot in the left illustrates the ratio taken for the process of estimation of the mapping functions at  $\epsilon=1.2$  and the plot in the right illustrates the ratio taken at  $\epsilon=1.1$ ; the blue-color plots depict the rate of the seconds taken by OCICI over those by H-RANSAC and the red-color for OCICI over A-RANSAC, respectively. The min and max of the seconds taken by each method are shown in Table 2.

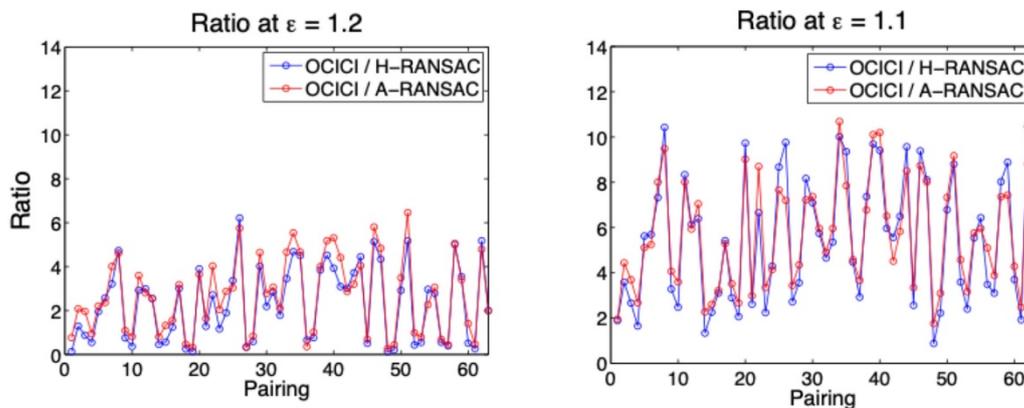
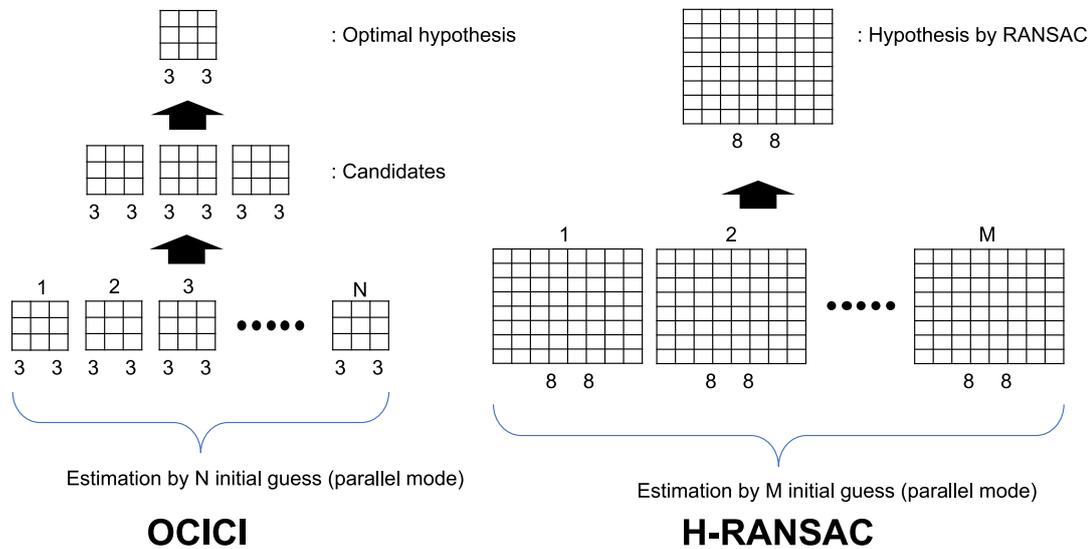


Fig. 21. The ratio of the elapsed time implemented in the serial programming

**Table 2.** Min and Max of the seconds taken by each method

	[Min Max] (Sec.)	
	$\epsilon=1.2$	
	$\epsilon=1.1$	
H-RANSAC	[0.4750 3.3941]	[1.3144 2.9668]
A-RANSAC	[0.1497 3.8535]	[1.0239 4.0019]
OCICI	[0.1152 17.0530]	[1.9983 29.4920]

Seeing the above results in the elapsed time, we admit that OCICI is the slowest one among the methods in serial programming. However, fast-parallel computing in various memory architecture may well be employed for the efficiency in the calculation. In the sense of the algorithmic structure of the tested methods in this study, we illustrate the computing mechanism in the parallel mode of the computing in Fig. 22, especially for OCICI and H-RANSAC; the controversial point may be that in OCICI the affine estimation in parallel mode is composed in two steps and in H-RANSAC it is done by one step structure in the parallel mode; nevertheless, the smaller costs of solving  $3 \times 3$  matrices in the affine-estimation of OCICI rather than the more massive  $8 \times 8$  matrix calculations to obtain the  $(3 \times 3)$  homographic mapping functions and the repetitively applied parallel mode in the estimation of the candidate affine-maps may well nullify the issues of the computing costs raised from the complexed two-step methodology of OCICI. Moreover, the numerical results in the previous section thoroughly recommend the utility of the OCICI.



**Fig. 22.** Illustration of the parallel mode of the computing mechanism of both methods

For future works, we can consider several points of view to improve the techniques of the state-of-the-art in the area of the image stitching. The proposed method in this article, i.e. OCICI, is to find the optimal feature pairs in the images via the measurement derived by an intuition of the geometric congruence of the simplex. In view of the measuring of the geometric distance between two sets of positions of the feature points, the thresholding measurement may be defined in other ways. For example, the patches of spots detected as the candidate matching positions of the feature points can be extracted so that a local correlation filter [21] is examined to search the best region of stitching. In the sense of mathematical

methodology, a topological feature [22] may be developed to detect some common regions, or an energy minimization method could be developed for the use in the image stitching such as used in the area of the image reconstruction [23].

## 5. Conclusion

In this study, we proposed a method for optimizing initial guess by comparing direct measurements rather than probabilistic methods such as RANSAC; our proposed methodology (OCICI) was developed through consideration of the image's feature-based geometric inherent characteristics such as the image's angle and the geometric congruence ratio. Based on various experiments, the optimality of the guess for the true features of the proposed method had been observed along with the stability of the proposed method to find the estimation of the mapping function in the stitching of images (see Fig. from 8 to 16).

The error estimation of the stitching results was determined based on the pixel norm difference of the compared outputs, and it was confirmed that the accuracy was high when the combination of the parameters was well determined.

From the example test of our experiments, the proposed method outperforms for such images as the flat-images which may be obtained by cropping given image and making several separate ones and the drone-aerial images made up of vertical photographs for easy panoramic stitching.

The model proposed in this study have contributed in the choice of the good estimation of the mapping function, as an initial guess of the sample consensus method is filtered in two-step robust estimation process by our geometric measure of congruence with a property of well-ordering, which is verified by examples to reduce the ambiguity of RANSAC that is conducted in a rigorous way for the choice of the good initial guess.

## Acknowledgments

The authors are thankful to Prof. Shuai Liu, Prof. Samuel Cheng and the anonymous reviewers for their valuable comments and suggestion. And we give special thanks to Prof. ByungRae Cha of Gwangju Institute of Science and Technology for his courageous suggestion to submit this article. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (2017R1E1A1A03070059).

## Reference

- [1] B. D. Lucas, T. Kanade, "An iterative image registration technique with an application in stereo vision," in *Proc. of Seventh International Joint Conference on Artificial Intelligence(IJCAI-81)*, (Vancouver), pp. 674-679, 1981. [Article \(CrossRef Link\)](#)
- [2] R. Szeliski, "Image alignment, and stitching: a tutorial," *Foundation and Trends in Computer Graphics and Computer Vision*, Vol. 2, No. 1, pp. 1-104, 2006. [Article \(CrossRef Link\)](#)
- [3] B. Zitová, J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, Vol. 21, pp. 977-1000, 2003. [Article \(CrossRef Link\)](#)
- [4] N. Tyutyundzhiev, K. Lovchinov, F. Martinez-Moreno, J. Leloux, L. Narvarte, "Advanced PV modules inspection using multirotor UAV," in *Proc. of 31th European Photovoltaic Solar Energy Conference and Exhibition, Hamburg*, September 2015. [Article \(CrossRef Link\)](#)

- [5] J. A. Besada, L. Bergesio, I. Campana, D. Vaquero-Melchor, J. Lopez-Araquistain, A. M. Bernardos and J. R. Casar, "Drone Mission Definition and Implementation for Automated Infrastructure Inspection Using Airborne Sensors," *Sensors*, 18, 1170, 2018. [Article \(CrossRef Link\)](#)
- [6] L. P. Koh, S. A. Wich, "Dawn of drone ecology: low-cost autonomous aerial vehicles for conservation," *Tropical Conservation Science*, Vol. 5 (2), 121-132, 2012. [Article \(CrossRef Link\)](#)
- [7] P. Lottes, R. Khanna, J. Pfeifer, C. Stachniss, R. Siegwart, "UAV-based crop and weed classification for future farming," in *Proc. of 2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017. [Article \(CrossRef Link\)](#)
- [8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004. [Article \(CrossRef Link\)](#)
- [9] R. Hess, "An Open-source SIFT library," *ACM International Conference on Multimedia*, pp. 1493-1496, 2010. [Article \(CrossRef Link\)](#)
- [10] K. Mikolajczyk, and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.1615-1630, 2005. [Article \(CrossRef Link\)](#)
- [11] O. Miksik, K. Mikolajczyk, "Evaluation of local detectors and descriptors for fast feature matching," in *Proc. of 21<sup>st</sup> International Conference on Pattern Recognition(ICPR 2012)*, pp. 2681-2684, 2012. [Article \(CrossRef Link\)](#)
- [12] M. A. Fischler, R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, Vol. 24, No. 6, pp. 381-395, 1981. [Article \(CrossRef Link\)](#)
- [13] C. V. Stewart, "Robust parameter estimation in computer vision," *SIAM Review*, Vol. 41, No. 3, pp. 513-537, 1999. [Article \(CrossRef Link\)](#)
- [14] P. Meer, D. Mintz, A. Rosenfeld, D. Y. Kim, "Robust regression methods for computer vision: a review," *International Journal of Computer Vision*, Vol. 6, No. 1, pp. 59-70, 1991. [Article \(CrossRef Link\)](#)
- [15] P. J. Rousseeuw, A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley, 1987. [Article \(CrossRef Link\)](#)
- [16] S. Choi, T. Kim, W. Yu, "Performance evaluation of RANSAC family," *British Machine Vision Conference*, 2009. [Article \(CrossRef Link\)](#)
- [17] E. H. Moore, "On the reciprocal of the general algebraic matrix," *Bulletin of the American Mathematical Society*, 26 (9), 394-95, 1920. [Article \(CrossRef Link\)](#)
- [18] H. Sadikin, A. Saptari, R. Abdulharis, A. Hernandi, "UAV System With Terrestrial Geo-referencing For Small Area Mapping," *FIG Congress 2014, Engaging the Challenges-Enhancing the Relevance, Kuala Lumpur*, 7146, 2014. [Article \(CrossRef Link\)](#)
- [19] A. Stoica, "Image mosaicing, version 1.0.0.0," [https:// sites.google.com/ site/ computervisionadinastoica/ project – 2 – image – mosaicing – workflow](https://sites.google.com/site/computervisionadinastoica/project-2-image-mosaicing-workflow), 2014. [Article \(CrossRef Link\)](#)
- [20] A. Vedaldi, B. Fulkerson, <http://www.vlfeat.org/overview/sift.html>, Copyright (c) 2007-11/ The VLFeat Team, Copyright (c) 2012-13. [Article \(CrossRef Link\)](#)
- [21] S. Liu, G. Liu, H. Zhou, "A Robust Parallel Object Tracking Method for Illumination Variations," *Mobile Networks and Applications*, 24(1), 5-17, 2019. [Article \(CrossRef Link\)](#)
- [22] W. Wei, X.-L. Yang, P.-Y. Shen, and B. Zhou, "Holes detection in anisotropic sensor net: Topological methods," *International Journal of Distributed Sensor Networks*, 8(10), 135054, October 23, 2012. [Article \(CrossRef Link\)](#)
- [23] W. Wei, X.-L. Yang, B. Zhou, J. Feng, and P.-Y. Shen, "Combined energy minimization for image reconstruction from few view," *Mathematical Problems in Engineering*, Art. ID:154630, October 31, 2012. [Article \(CrossRef Link\)](#)



**Dr. Byeong-Chun Shin** is presently working as a professor at Department of Mathematics, Chonnam National University, Gwangju, Korea. He received the B.S., M.S. and Ph.D. degrees at department of Mathematics from Kyungpook National University, Daegu, Korea, in 1986, 1988 and 1996, respectively. His main research is developing numerical algorithms to solve partial differential equations and optimal control problems for fluid flows. He is recently focusing on the development of new algorithm including the artificial neural network and deep learning with computer vision.



**Dr. Jeong-Kweon Seo** received his Ph.D. in Mathematics at Chonnam National University in Gwangju, South Korea. He is currently working as an algorithm designer in the field of the scientific computing, collaborating with several engineers and researchers in various areas of the computational sciences. His research interest is ranging over the solver of the partial differential equations, computer vision, and deep-learning, in the fields of applied mathematics, vision-based bio-medical engineering, computational bio-sciences and a little of the brain-computer-interface technologies.