A Generalized Tool for Accurate and Efficient Image Registration of UAV Multi-lens Multispectral Cameras by N-SURF Matching

Jyun-Ping Jhan and Jiann-Yeou Rau

***Abstract*—The original multispectral (MS) images obtained from multi-lens multispectral cameras (MSCs) have significant misregistration errors, which require image registration for precise spectral measurement. However, due to the nonlinearity intensity differences among MS images, performing image matching is difficult to find sufficient correct matches (CMs) for image registration, and results in a complex coarse-to-fine solution. Based on the modification of speed-up robust feature (SURF), we proposed a normalized SURF (N-SURF) that can significantly increase the amount of CMs among different pairs of MS images and make one-step image registration possible. In this study, we first introduce N-SURF and adopt different MS datasets acquired from three representative MSCs (MCA-12, Altum, and Sequoia) to evaluate its matching ability. Meanwhile, we utilized three image transformation models—affine transform (AT), projective transform (PT), and an extended projective transform (EPT) to correct the misregistration errors of MSCs and evaluate their co-registration correctness. The results show that N-SURF can obtain 6–20 times more CMs than SURF and can successfully match all pairs of MS images, while SURF failed in the cases of significant spectral differences. Moreover, visual comparison, accuracy assessment, and residual analysis show that EPT can more accurately correct the viewpoint and lens distortion differences of MSCs than AT and PT, and it can obtain co-registration accuracy of 0.2–0.4 pixels. Subsequently, using the successful N-SURF matching and EPT model, we developed an automatic MS image registration tool that is suitable for various multilens MSCs.**

***Index Terms*—Image matching, image registration, multispectral camera (MSC), multispectral (MS) image.**

I. Introduction

MULTISPECTRAL (MS) images can be acquired using a multilens multispectral camera (MSC) that records visible [red (RED), green (GRE), and blue (BLU)] and invisible [red edge (REG) and near-infrared (NIR)] spectral information. The light weight and small size of MSCs make them suitable for mounting on various unmanned aerial vehicle (UAV) platforms to obtain high-spatial-resolution images, and the diversity of the MS band combinations can derive various vegetation indexes [1] that can be widely applied in environment monitoring applications [2], [3]. However, since MSCs adopt a multi-lens structure to record distinct spectral information, the viewpoint and lens distortion differences among each lens lead to significant ghost effects in original images. To recover one-sensor geometry for precise spectral analysis, performing automatic MS image registration is an important task to reduce the band misregistration errors of MSCs [4]–[6].

Image matching is a crucial step in image registration [7]; it is used to find conjugate features between overlapped images and then estimate the coefficients of a geometric transformation model. Therefore, the geometric differences, such as scale, rotation, and translation, between two images can be corrected by a proper image transformation. Various image matching methods have been developed in recent decades and these can be generally classified into area-based and feature-based methods [7]. Compared to area-based methods that compute the cost function of pixel intensities in an image template, feature-based methods extract specific features (i.e., points, corners, or lines) that are better suited for real-time image registration purposes. State-of-the-art feature-based image matching methods such as scale-invariant feature transform (SIFT) [8] and speed-up robust feature (SURF) [9] can extract scale- and rotation-invariant features to match images with geometric distortions and have been widely adopted in 3-D scene reconstruction [10], object recognition [11], and image registration [7]. Feature-based image matching methods generally contain three steps: feature extraction, feature description, and feature matching. Feature extraction extracts the local extremum from a multiscale image space to acquire scale-invariant features, and a feature threshold (FT) is utilized to determine whether it is a robust feature. Next, the main direction is assigned by finding the maximum gradients to achieve rotation invariance and a multi-dimension descriptor is established on each feature by computing the gradient distribution of neighborhood pixels. In the end, features are matched if two points have the most similar invariant descriptors on reference and target images.

To increase the image matching performance, recent improvements in SIFT descriptors, such as principal component analysis-SIFT (PCA-SIFT) [12], colored SIFT (CSIFT) [13], and gradient location-orientation histogram (GLOH) [14] make them more effective and correct. Meanwhile, to increase the number of matching points between viewpoint differences, applying geometric transformation matching, such as Affine-SIFT

Manuscript received November 16, 2020; revised January 28, 2021 and April 24, 2021; accepted May 8, 2021. Date of publication May 12, 2021; date of current version July 2, 2021. This work was supported by the Ministry of Science and Technology (MOST), Taiwan with project number MOST 107-2621-M-006-001. *(Corresponding author: Jyun-Ping Jhan.)*

The authors are with the Department of Geomatics, National Cheng Kung University, Tainan 70101, Taiwan (e-mail: jyunpingjhan@geomatics.ncku.edu.tw; jyrau@geomatics.ncku.edu.tw).

Digital Object Identifier 10.1109/JSTARS.2021.3079404

[15] and fully affine invariant SURF (FAIR-SURF) [16] can overcome distortion among images. Although many improved SIFT-like, SURF-like, and even deep learning-based methods [17] have been proposed to increase the matching robustness in common color images, successful image matching of the non-linearity intensity differences of MS images remains challenging.

A. Related Works on MS Image Matching and Image Registration

Due to it is difficult to find sufficient correct matches (CMs) on MS images, related works on MS image registration have generally adopted complicated coarse-to-fine image registration procedures [4], [18]. Ye and Shan [18] adopted two-stage image matching procedures for satellite MS image registration. They first use scale restriction SIFT [19] to correct the significant translation and rotation among images and then perform a fine co-registration by another area-based matching. Similar approaches of integrating both feature-based and area-based matching for satellite MS image registration can be found in [20] and [21]; however, these approaches are too complicated and time-consuming for practical application when dealing with large amounts of MS image registration. For MSCs, Jhan *et al.* [4] proposed a robust and adaptive band-to-band image transform method for correcting their misregistration errors. They first conducted a camera system calibration to directly obtain the coefficients of an image transformation model and then used SURF matching to optimize the remaining systematic errors caused by the calibration uncertainty. Though co-registration accuracy better than 0.5 pixels were reported, it requires prior knowledge and successful image matching to conduct sensor calibration and errors optimization, respectively.

In order to conduct one-step image registration, three MS image matching strategies can be considered to overcome the nonlinearity intensity differences and to increase the number of CMs. The first one is to construct an MS cube and match neighbor images that have the most similar spectral information [22]. However, this is only suitable for datasets that have various bands of MS images. The second method is to enhance the image contrast and brightness to increase the extracted features [23], [24]. Since more features can be detected, the number of CMs is increased, but this caused an extra image processing effort. The final and most commonly adopted method is to modify the feature descriptor to increase the robustness to intensity change. Researchers have studied the performance of various feature-based matching methods on MS images and proposed their improved feature descriptors [25]–[28]. However, since there was no further improvement in feature extraction, incrementing of CMs is still limited.

B. Motivation and Objective

Although related studies have focused on methods to increase the feature descriptor robustness, increase the candidate features can significantly increase the number of CMs for accurate image registration. However, the intensity differences among MS images afford an inconsistent amount of features and a potential number of CMs are lost when same FT value is adopted. Therefore, to conduct MS image registration of MSCs, we modified the feature extraction and FT criteria of SURF and proposed normalized SURF (N-SURF) [29]. N-SURF can extract more features than SURF to achieve successful MS image matching.

In this article, we evaluated the N-SURF matching ability on different pairs of MS images that were obtained from three representative MSCs. Additionally, we adopted three different image transformation models to correct the misregistration errors and compared their co-registration correctness. Moreover, to automatically co-register a large amount of UAV MS images, we developed an automatic MS image registration tool that is suitable for the accurate and rapid image registration applications of various MSCs.

TABLE I
MS DATASETS

MSC	MCA-12[1]	Altum[2]	Sequoia[3]
Manufacture	Tetracam	Micasense	Parrot
Bands	12	5 MS + 1 TIR	4 MS + 1 RGB
Wavelength (nm)	450–950	475–840	530–810
Bandwidth (nm)	10 ~ 20	10 ~ 40	10 ~ 40
MS Image Resolution	1280 × 1024	2064 × 1544	1280 × 960
Groups of Images	100 × 12	100 × 5	100 × 4

[1]http://www.tetracam.com/
[2]https://www.micasense.com/altum
[3]https://www.parrot.com/us

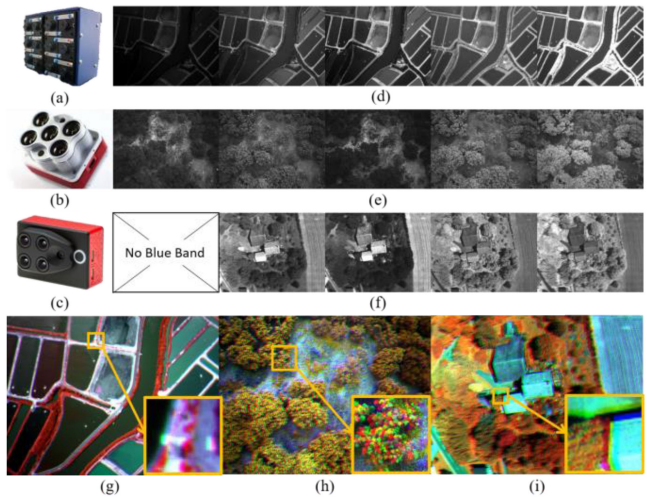

Fig. 1. MSCs, sample MS images, and misregistration errors of sample MS images. (a)–(c) MCA-12, Altum, and Sequoia, respectively. (d)–(f) Sample MS images of (a)–(c) showing BLU, GRE, RED, REG, and NIR images from left to right. Please note that only five images of MCA-12 are demonstrated here and Sequoia has no BLU image. (g)–(i) The significant band misregistration errors of (d)–(e) with false color combinations of REG, GRE, and RED images.

II. MS Datasets

As given in Table I and shown in Fig. 1, three representative MS datasets acquired from MCA-12, Altum, and Sequoia MSCs

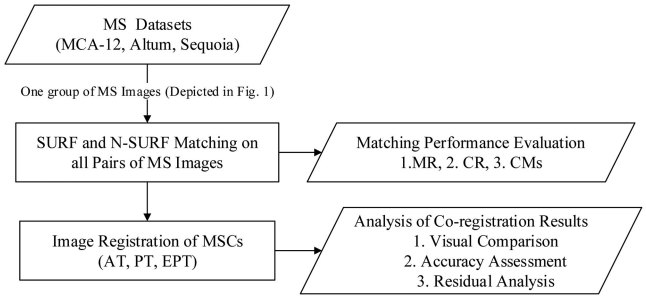

Fig. 2. MS image matching analysis of N-SURF and image registration of MSCs.

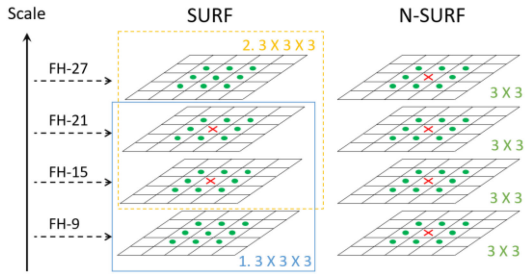

Fig. 3. Differences in local extremum detection structure between SURF and N-SURF.

are used to conduct image matching and image registration analysis. MCA-12 has 12 separate lenses that can record 12 distinct 1.3-megapixel narrowband MS images. Altum and Sequoia are hybrid MSCs, where Altum records five 3.2-megapixel MS images and one very-low-resolution (0.02-megapixel) thermal infrared (TIR) image, while Sequoia only captures four 1.2-megapixel MS images and another high-resolution (16-megapixel) color (RGB) image. Although the numbers of MS bands are unequal, they can be generally classified into five MS categories, BLU, GRE, RED, REG, and NIR; note that Sequoia has no BLU image.

Each camera comprises 100 groups of MS images, where each group of MS images contains all bands of images that were simultaneously captured by the camera. The Sequoia dataset was downloaded from the Pix4D sample dataset [30] and the other two were acquired by a UAV company in Taiwan. Fig. 1 demonstrates one group of sample MS images for each MSC, which shows that the nonlinearity intensity change across different MS bands is significant. For example, the vegetation area is lighter on an NIR image and darker on a RED image. Furthermore, since MCA-12 adopts narrower bandwidth filters, its original BLU image has very low contrast and low illumination to the point that the image contents are barely visible; thus, requiring atmospheric correction to obtain the true reflectance value [31]. In addition, Fig. 1 illustrates that, due to the geometric differences of viewpoint and lens distortion among lenses, the original images are not aligned, which leads to significant misregistration errors. Since the textures, intensity, and illumination differ among MS bands and scenes, performing successful MS image matching is quite challenging. Meanwhile, correcting the geometric differences among the lenses of MSC requires understanding how to choose a proper image transformation model.

III. Methodology

Fig. 2 depicts the MS image matching evaluation of N-SURF and image registration analysis of MSCs. The sample MS images in Fig. 1 are used for N-SURF matching, wherein we analyzed all MS image matching pairs (ex: 5 bands have $C_2^5 = 10$ different pairs) and evaluated the N-SURF matching performance using three indexes: matching rate (MR); number of CMs; and correct rate (CR). Meanwhile, in order to correct the misregistration errors of MSCs, we adopted affine transform (AT), projective transform (PT), and extended projective transform (EPT) for image registration and analyzed their correctness by visual comparison, accuracy assessment, and residual analysis.

A. Normalized Speed-Up Robust Feature

In contrast to the original SURF that extracts features on a multi-image scale, we adopted single-image scale to increase the candidate features and compute the cumulative distribution function (CDF) of FT to obtain the required number of features. The modified SURF is named N-SURF since equal amounts of features can be obtained across different scenes and different bands of MS images. N-SURF is a modified form of the open-source code OpenSURF [32], which is almost identical to the original SURF but affords slightly better performance [33]. Details of N-SURF are introduced in the following sections.

1) Feature Extraction: To extract scale-invariant features, SURF constructs a multiscale image space that contains different sizes of fast Hessian (FH) box filter to detect blob-like features. It consists of numbers of octaves, where each octave contains four layers of FH and each one refers to a specific image scale. According to Bay *et al.* [9], the first octave has filter sizes of FH-9, FH-15, FH-21, and FH-27, while the second octave has larger sizes of FH that contain FH-15, FH-27, FH-39, and FH-51. As depicted in Fig. 3, the feature extraction in first octave searches the local extremum (i.e., red cross) within 3 $\times$ 3 neighborhood pixels and three connected image scales of 9-15-21 and 15-21-27. Therefore, if the FH's determinant of the local extremum (3 $\times$ 3 $\times$ 3) is larger than a FT, it is treated as a candidate feature. In contrast to SURF, N-SURF (see the right part in Fig. 3) only searches the local extremum (3 $\times$ 3) in each single image scale to increase the candidate features.

Fig. 4 summarizes the maximum number of extracted features (i.e., where FT = 0) of sample MS images in each individual scale of N-SURF and the first octave of SURF; obviously, a small size for FH can extract more features than a larger one and each scale has significantly more than SURF. Please note that N-SURF is also scale-invariant as the scale factor is interpolated in single scale space and applied to construct feature descriptors. Regarding the scale factor and the descriptor construction, refer to Bay *et al.* [9].

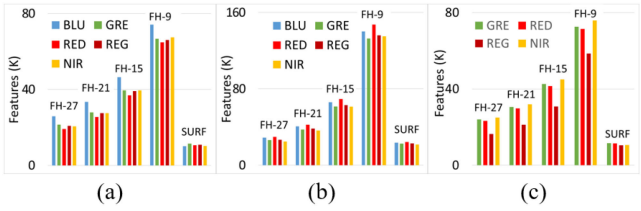

Fig. 4. Comparison of the extracted maximum number of features for different scales of FH and first octave of SURF. (a)–(c) Results of sample MS images for MCA-12, Altum, and Sequoia, respectively.

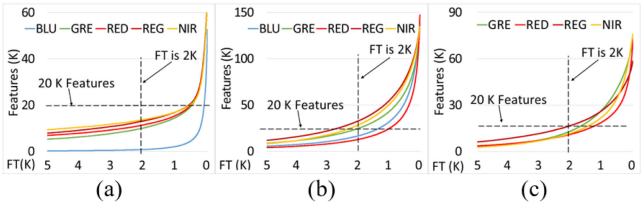

Fig. 5. CDF of the FT on FH-9. (a)–(c) Results of sample MS images for MCA-12, Altum, and Sequoia, respectively.

2) Feature Threshold: The FH determinant is sensitive to image contrast and texture such that a fixed value of FT yields different number of features depending on the MS images and scenes. To obtain required amount of features, we adopt CDF to adaptively select the proper FT. Fig. 5 presents the FT's CDF of sample MS images on FH-9, which clearly shows that how the intensity differences of MS images affect the curve of CDF. Taking MCA-12 as an example, if FT is equal to 2000, then 0 k, 10 k, 11 k, 13 k, and 13 k features will be afforded on BLU, GRE, RED, REG, and NIR images, respectively. Since no features are detected on BLU images, image matching and image registration cannot be performed. Conversely, by querying CDF to obtain 20k features on BLU, GRE, RED, REG, and NIR images, FT is automatically determined as 100, 500, 500, 500, and 500, respectively.

3) Matching and Evaluation Indexes: N-SURF is same as SURF that can construct 64- or 128-dimension descriptors and then match features by comparing the Euclidean distance ratio of the first- and second-closest descriptors on a target image. Even though the higher-dimension descriptor is more robust to find unique features, it causes greater memory usage and higher computation cost. On the other hand, a higher distance ratio can offer more matches but also lead to more errors. To reduce the computation cost and complexity, SURF and N-SURF utilize 64-dimension descriptors for matching features that have a distance ratio $< = 0.8$, and perform matching on all possible MS pairs, for example: NIR versus (REG, RED, GRE, and BLU), REG vs. (RED, GRE, and BLU), RED versus (GRE and BLU), and GRE versus BLU. Meanwhile, as shown in (1)–(2), we introduce two evaluation indexes and use the number of CMs to evaluate the matching performance.

$$MR\,(\%) = 2\ Matches/TotalDetectedFeatures \tag{1}$$

$$CR\,(\%) = CMs/Matches. \tag{2}$$

a) Matching Rate: The MR represents the ability to find conjugate matches between images, which is calculated by two times the number of matches over the summation of extracted features on the reference and target images. Since matched features exist in both the reference and target images, it is multiplied by two. Therefore, a higher value of MR can be expected between two MS images that have fewer spectral differences.

b) Correct Matches and Correct Rate: The number of CMs is the most important consideration for accurately estimate the coefficients of image transformation model. First, the blunders in initial matches are filtered by random sample consensus (RANSAC), and then a recursive error removal scheme is applied to eliminate false matches and preserve the CMs while estimating the coefficients. Hence, the CR is represented by CMs over total matches; for details about the recursive computation and error removal, refer to "*coefficients computation and accuracy assessment*" section below.

B. Image Registration

With successful image matching, we can select an appropriate image transformation model to estimate the coefficients, correct the misregistration errors, and evaluate the correctness of co-registration results through visual inspection and quantitative accuracy assessment.

1) Image Transformation Model: From the geometric accuracy perspective of MS image registration, finding a proper image transformation model is key for accurately correcting the systematic misregistration errors of viewpoint and lens distortion differences of MSCs. For MSCs, we utilized the AT, PT, and EPT models to conduct MS image registration.

a) Affine Transform: As shown in (3) and (4), AT has six parameters (A–F), in which (x, y) and (u, y) are the image coordinates of the target and reference images, respectively. It has been widely used on most image registration applications for two parallel scenes that have translation, rotation, and anisotropic linear distortions

$$\mathrm{u} = A \times x + B \times y + C \tag{3}$$

$$\mathrm{v} = D \times x + E \times y + F. \tag{4}$$

b) Projective Transform: Unlike AT, PT as shown in (5) and (6) is suitable for two non-parallel images that describes the map projection of a quadrangle and a square using eight parameters ($A_1 - A_3$, $B_1 - B_3$, and $C_1 - C_2$). Due to the slight rotation differences among each lens of the MSC results in an incompletely parallel images, it is expected that PT would have better co-registration results than AT

$$\mathrm{u} = \frac{A_1 \times x + A_2 \times y + A_3}{C_1 \times x + C_2 \times y + 1} \tag{5}$$

$$\mathrm{v} = \frac{B_1 \times x + B_2 \times y + B_3}{C_1 \times x + C_2 \times y + 1}. \tag{6}$$

c) Extended Projective Transform: Although PT is suitable to correct the non-parallel images of MSCs, there are still lens distortion differences among each lens. As depicted in Fig. 6, the radial lens distortion curve of each lens of MCA-12, Altum, and Sequoia that were calibrated using an indoor camera

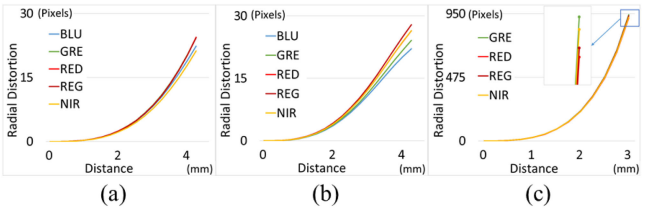

Fig. 6. Radial lens distortion curve of MSCs. (a)–(c) Results of MCA-12, Altum, and Sequoia, respectively.

calibration field [34], the radial lens distortion curves of different lenses in each MSC are very similar, and only slightly different (only 5–10 pixels) at the image borders. Therefore, without considering the lens distortion differences, it is expected that misregistration errors remain at the borders of images after PT. To correct both the geometrical differences of viewpoint and lens distortion, EPT is formed by adding the additional lens distortion parameters [35] of (7) and (8) to the original PT model. $(K_1 - K_3)$ and $(P_1 - P_2)$ are coefficients of the radial and decentering lens distortion coefficients, respectively, and r represents the distance to the image perspective center

$$\Delta_x = x \times \left(K_1 \times r^2 + K_2 \times r^4 + K_3 \times r^6\right) + P_1 \times \left(r^2 + 2x^2\right) + 2P_2xy \quad (7)$$

$$\Delta_y = y \times \left(K_1 \times r^2 + K_2 \times r^4 + K_3 \times r^6\right) + P_2 \times \left(r^2 + 2y^2\right) + 2P_1xy. \quad (8)$$

2) Coefficients Computation and Accuracy Assessment: The coefficients of each image transformation model are estimated using the least-square adjustment of matches, and the co-registration accuracy is represented by the root-mean-square errors (RMSEs) of the residuals of matches. However, since matches still comprise errors after RANSAC, the folds of RMSEs are used as the threshold, and matches that have residuals larger than the threshold are removed. To accurately estimate the coefficients, we adopt a recursive error removal scheme that conducts least-square adjustment and estimates RMSEs for removing errors; then, the procedure is repeated until no matches are removed. According to our experiences, 2.5 times the RMSE is an optimal value that can obtain reasonable accuracy and CMs.

3) Analysis of Co-Registration Results: To evaluate the correctness of the co-registration results among different image transformation models, visual comparison can provide direct evidence of the correctness, while RMSEs can provide numerical accuracy assessment. Moreover, by inspecting the distribution of CMs residuals, we can select the best image transformation model with the smallest systematic errors suitable for MSC image co-registration.

IV. Results and Analysis

This section evaluates the matching performance of SURF and N-SURF, compares the different image registration results, and introduces the generalized tool for MS image co-registration of MSCs.

TABLE II
Sample of Matching Evaluation Table

Matching Method / Features of Target Image			Features of Reference Image					
			MR (%)				CR (%)	
			CM (Pts)				RMSE (Pixels)	
SURF-10	GRE(1,370)		RED(1,772)		REG(2,006)		NIR(2,152)	
BLU (17)	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*
GRE (1,370)	#		36	74	10	52	*	*
			362	0.31	68	0.29	*	*
RED (1,772)	#				10	54	*	*
					79	0.26	*	*
REG (2,006)	#						21	71
							239	0.30

TABLE III
Parameter Settings for Matching Evaluation

Method	SURF-MAX	N-SURF-2%			N-SURF-MAX
Scale	First octave	FH-9			
Features	Max. features	2 % of the image resolution.			Max. features
		MCA-12	Altum	Sequoia	
		26000	64000	24000	
Matching parameters					
Descriptor dimensions		64			
Matching threshold		0.8			
Image transformation model		EPT			
Error removal threshold		2.5 times RMSEs			

A. Matching Performance Evaluation

Table II gives the matching evaluation table of MCA-12 that utilizes SURF-10 (i.e., FT = 10) and EPT for the image registration of all MS image matching pairs, which summarizes the utilized matching method, the extracted number of features, evaluation indexes, and RMSEs of image registration. In the table, the symbols "$*$" and "#" denote that the image matching or image registration has failed and that the matching pair already exists, respectively. Therefore, the differences between different pairs can be clearly observed. For example, it shows that SURF-10 leads to different amounts of features on different MS images, in which only a few features are detected on the BLU image and is not possible to conduct image registration. On the other hand, since the spectral response between NIR versus RED and NIR versus GRE are significantly different, the SURF-10 image registration failed.

To reduce the complexity of matching analysis in this article, only SURF-MAX, N-SURF-2% (feature amounts are equal to 2% of the image resolution), and N-SURF-MAX are used for comparison; Table III gives the adopted matching parameters. Image resolution of 2% is an empirical threshold for ensuring that sufficient features are obtained between different scenes and image sizes. Note that if the threshold is greater than the number of extracted features, N-SURF-2% will be equal to N-SURF-MAX (i.e., FT = 0).

1) Overview: Three major findings can be observed from the highlighted areas of matching evaluation tables (Tables I–III in the appendixes). First, N-SURF-MAX can clearly obtain more CMs (indicated in bold numbers) than SURF-MAX. Second,

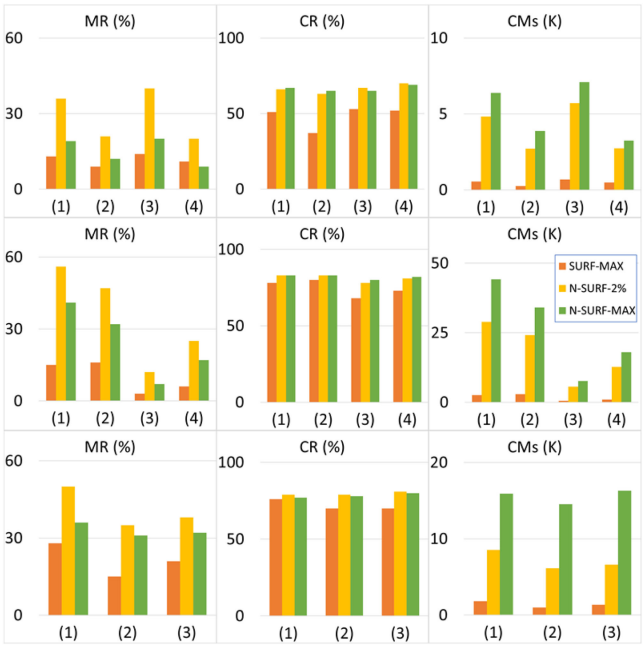

Fig. 7. Summary of the matching performance evaluation of robust MS image matching pairs. From left to right are MR, CR, and CMs; from top to bottom are the results of MCA-12, Altum, and Sequoia. Labels (1)–(4) represent REG versus NIR, REG versus GRE, GRE versus. RED, and GRE versus BLU, respectively.

N-SURF-2% and -MAX are perform significantly better than SURF-MAX; they successfully match all pairs, while SURF-MAX fails (shown as $*$) in some cases. All cases show that MS image matching pairs (highlighted as gray shading) with smaller spectral difference can obtain higher MR and better RMSEs, where the top four matching pairs (i.e., REG versus NIR, REG versus GRE, GRE versus RED, and GRE versus BLU) can construct robust MS matching pairs to connect each image. Therefore, we can chose REG as a master image to perform matching and registering on NIR and GRE images and then use a corrected GRE image to connect with RED and BLU images. Further detailed analysis is based on these four MS matching pairs.

2) Analysis of Robust MS Image Matching Pairs: Fig. 7 shows the MR, CR, and CMs of the robust MS image matching pairs for each camera. Compared to SURF-MAX, N-SURF-(2% and MAX) have higher values of MR and CR. There are no significant differences in CR between N-SURF-(2% and MAX), but since N-SURF-2% extracts a consistent amount of features between images, it has a better MR. On the other hand, it is obvious that the total number of CMs for N-SURF-(2% and MAX) are respectively increased by 4–13 and 6–20 times compared to SURF-MAX. As for the image registration accuracy (indicated in italic numbers and green color), we can also observe that the RMSEs of N-SURF-2% is smaller than SURF-MAX; since N-SURF-MAX extracts the maximum number of features, its accuracy is slightly poorer than N-SURF-2%. In summary, N-SURF is more efficient and accurate than SURF and can obtain more CMs for MS image matching and registration, where N-SURF-2% can achieve balanced CMs and co-registration accuracy.

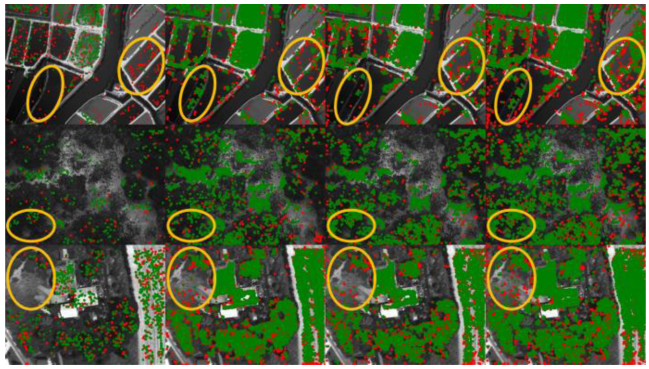

Fig. 8. Comparison of CM distributions. From left to right are SURF, N-SURF-FH-9, N-SURF-FH-15, and N-SURF-FH-9-15; from top to bottom are results of MCA-12, Altum, and Sequoia. The green and red points represent correct and false matches, respectively.

TABLE IV
CMs AND IMAGE REGISTRATION ACCURACY OF SURF AND THREE DIFFERENT SCALES OF N-SURF

REG vs. RED		SURF-MAX	N-SURF-MAX		
		First Octave	**FH-9**	FH-15	FH-9-15
MCA-12	RMSE (Pixels)	0.47	**0.45**	0.58	0.49
	CMs (Pts)	311	**2891**	2170	4911
Altum	RMSE (Pixels)	0.49	**0.34**	0.58	0.42
	CMs (Pts)	282	**3241**	2712	5736
Sequoia	RMSE (Pixels)	0.31	**0.27**	0.35	0.30
	CMs (Pts)	694	**5044**	3455	8325

TABLE V
PROCESSING EFFICIENCY OF INDEPENDENT AND BATCH MODE UNDER MULTITHREAD AND GPU ACCELERATING

Camera	MCA-12	Altum	Sequoia
Independent[1] (min.)	65	62	17
Independent[2] (min.)	**13**	**14**	**3**
Batch[1] (min.)	2	2	2
Batch[2] (min.)	**2**	**2**	**2**

[1]CPU: Intel(R) Core(TM) i7-8700 3.2 GHz
[2]GPU: NVIDIA GeForce GTX TITAN X

[1]CPU: Intel(R) Core(TM) i7-8700 3.2 GHz.
[2]GPU: NVIDIA GeForce GTX TITAN X.

3) Different Scales of N-SURF: To better understand the differences between SURF and N-SURF, Fig. 8 compares the CMs' distributions of SURF (first octave only) and three different scales of N-SURF (i.e., FH-9, FH-15, and FH-9-15), while Table IV gives the number of CMs and image registration accuracy. The matching pair utilized here is REG vs. RED that has significant spectral differences and we used the maximum number of features for matching analysis. Focusing on the yellow circle in Fig. 8, SURF can only obtain a limited amount of CMs, while N-SURF in different scales can all obtain a denser and better distribution of CMs. Meanwhile, different scales of N-SURF in Table IV shows that FH-9 has the best accuracy; the larger filter size of FH-15 slightly reduces the accuracy and amount of CMs, and multiscale FH-9-15 can nearly obtain the total amount of CMs between two independent scales and the image registration accuracies in between them. Therefore, N-SURF both proves that it can acquire more CMs and has a better distribution, which is crucial for accurately estimating the coefficients of the image transformation model.

TABLE AI
MATCHING EVALUATION TABLE OF MCA-12

Matching Method / Features of Target Image	Features of Reference Image							
	MR (%)				CR (%)			
	CM (Pts)				RMSE (Pixels)			
SURF-MAX	GRE(11,522)		RED(10,608)		REG(10,775)		NIR(10,140)	
BLU (10,053)	11	52	11	56	*	*	*	*
	481	0.66	481	0.85	*	*	*	*
GRE (11,522)	#		14	53	9	37	9	31
			678	0.43	256	0.45	171	0.43
RED (10,608)	#				9	42	*	*
					308	0.42	*	*
REG (10,775)	#						13	51
							555	0.46
N-SURF-2%	GRE(26,000)		RED(26,000)		REG(26,000)		NIR(26,000)	
BLU (26,000)	20	70	15	63	8	58	7	55
	2,733	*0.80*	1,945	0.80	848	0.89	641	0.90
GRE (26,000)	#		40	67	21	63	17	61
			5,722	*0.42*	2,710	*0.42*	1,952	0.43
RED (26,000)	#				21	62	21	69
					2,635	0.42	2,362	0.41
REG (26,000)	#						36	66
							4,827	*0.45*
N-SURF-MAX	GRE(66,683)		RED(64,773)		REG(66,053)		NIR(67,386)	
BLU (74,097)	9	69	7	67	4	64	3	59
	3,257	0.89	**2,417**	0.89	**1,202**	0.99	**849**	0.96
GRE (66,683)	#		20	65	12	65	9	62
			7,092	0.46	**3,867**	0.50	**2,794**	0.48
RED (64,773)	#				12	60	12	61
					3,617	0.47	3,525	0.50
REG (66,053)	#						19	67
							6,383	0.52

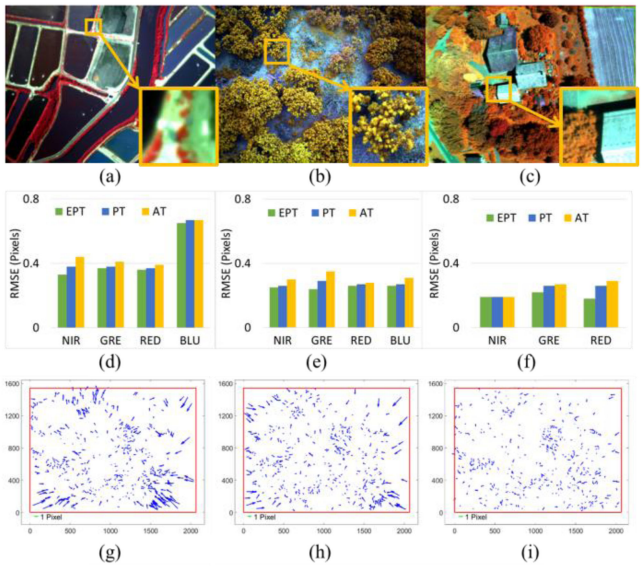

Fig. 9. The image registration results of sample MS images and their accuracy and residuals of different image transformation models. (a)–(c) Co-registered results of sample MS images by EPT. (d)–(f) Accuracy of different image transformation models of (a)–(c), respectively. (g)–(i) Residuals of AT, PT, and EPT, respectively, that are acquired from the co-registered BLU image of Altum.

B. Comparisons of Different Image Registration Results

Fig. 9 depicts the image registration results of sample MS images in Fig. 1, the co-registration accuracy, and the residuals of CMs for different image transformation models. First, it shows that all images are well co-registered, and the accuracy of EPT is better than that of PT, and PT is better than AT. We can observe that the accuracy improvement is not significant since the original images are very close to parallel and the differences in lens distortion among lenses are very small and only exist at the image boundary. Except for the low intensity of MCA-12's BLU image that leads to larger errors, it shows that using AT is suitable for most image registration applications, while EPT can achieve the best results with 0.2–0.5 pixels accuracy.

TABLE AII
MATCHING EVALUATION TABLE OF ALTUM

Matching Method / Features of Target Image	Features of Reference Image							
	MR (%)						CR (%)	
	CM (Pts)						RMSE (Pixels)	
SURF-MAX	GRE(22,611)		RED(24,354)		REG(22,728)		NIR(21,620)	
BLU (23,623)	6	73	10	76	2	63	*	*
	957	0.42	1,785	0.40	318	0.56	*	*
GRE (22,611)	#		3	68	16	80	3	66
			489	0.47	2,858	0.32	469	0.50
RED (24,354)	#				2	57	*	*
					282	0.49	*	*
REG (22,728)	#						15	78
							2,558	0.34
N-SURF-2%	GRE		RED		REG		NIR(32,000)	
BLU (32,000)	25	81	28	82	10	84	3	75
	12,699	*0.35*	14,108	0.38	5,177	0.52	1,253	0.56
GRE (32,000)	#		12	78	47	83	14	87
			5,649	*0.33*	24,086	*0.35*	7,685	0.50
RED (32,000)	#				8	79	2	69
					3,772	0.36	964	0.29
REG (32,000)	#						56	83
							28,894	*0.29*
N-SURF-MAX	GRE(132,800)		RED(147,315)		REG(135,976)		NIR(135,976)	
BLU (140,021)	17	82	17	81	7	84	2	78
	18,033	0.38	**19,972**	0.41	**7,251**	0.55	**2,013**	0.58
GRE (132,800)			7	80	32	83	10	85
			7,667	0.35	**34,005**	0.37	**10,987**	0.51
RED (147,315)					6	80	2	75
					6,356	0.38	**2,226**	0.32
REG (135,976)							41	83
							44,215	0.31

TABLE AIII
MATCHING EVALUATION TABLE OF SEQUOIA

Matching Method / Features of Target Image	Features of Reference Image					
	MR (%)				CR (%)	
	CM (Pts)				RMSE (Pixels)	
SURF-MAX	RED(11,481)		REG(10,451)		NIR(10,632)	
GRE (11,578)	21	70	15	70	11	61
	1,348	0.29	992	0.31	599	0.36
RED (11,481)	#		11	67	9	65
			691	0.31	478	0.47
REG (10,451)	#				28	76
					1,816	0.29
N-SURF-2%	RED(24,000)		REG(24,000)		NIR(24,000)	
GRE (24,000)	38	81	35	79	22	77
	6,608	*0.29*	6,140	0.26	3,475	0.29
RED (24,000)	#		15	71	9	71
			2,477	*0.23*	1,355	*0.27*
REG (24,000)	#				50	79
					8,509	0.20
N-SURF-MAX	RED(71,388)		REG(58,535)		NIR(75,989)	
GRE (72,566)	32	80	31	78	17	78
	16,307	0.33	**14,536**	0.30	**8,842**	0.35
RED (71,388)	#		18	76	10	76
			8,320	0.29	**4,714**	0.33
REG (58,535)	#				36	77
					15,915	0.25

However, when looking to the residuals of co-registered images, their differences are much clearer. Fig. 9(g)–(i) is the CMs' residuals of AT, PT, and EPT acquired from co-registered BLU image of Altum; since the images are not exactly parallel, AT has larger and anisotropy residuals, while PT gets better results but larger residuals remain at the boundary of the image as the lens distortion differences are not considered. However, since EPT considered the lens distortion compensation, it shows that all residuals are <1 pixel and randomly distributed, proving that this is the best model for properly correcting the systematic errors of MSCs.

C. Automatic MS Image Registration of MSCs

With the successful N-SURF matching and correct EPT model, an automatic MS image registration tool was developed for various kinds of MSCs. To process a large amount of MS images, batch or independent image processing can be considered. For MSCs, the systematic misregistration errors are assumed to be consistent; thus, one group of MS images can be selected to estimate the EPT coefficients. The coefficients can be fixed for co-registering the remaining groups of MS images with efficient rapid processing. However, batch processing is not recommended if synchronization errors are present among MS images. In contrast, independent processing can perform image matching and separately use EPT for image registration on each group of MS images, which is adaptable for dealing with inconsistent systematic errors to obtain the best co-registration results. The performances of the automatic MS image registration tool are discussed in the following section.

1) Efficiency Analysis: The automatic MS image registration tool is programmed in C#, and it adopts a multi-thread CPU and third-party Alea GPU [32] to increase the processing efficiency. Users can adjust the default amount of N-SURF features required to increase the matching efficiency or number of CMs. Table V gives the time consumed by N-SURF-2% for co-registering 100 groups of MS images (see Table I) under independent and batch processing. In the independent mode, we can observe that GPU is five times faster than the multi-thread CPU. In addition, since there are no image matching efforts in batch mode, it achieves an ultrahigh processing efficiency where both multithread CPU and GPU computing can co-register all images in 2 min. Nevertheless, this shows that the developed automatic MS image registration tool is efficient for co-registering one hundred groups of images under independent mode within 14 min if the GPU is adopted.

2) Comparison of the Results: Fig. 10(a) compares the co-registration results of MCA-12 under batch and independent processing where the coefficients of EPT for batch processing are estimated from sample MS images. Although the results are very similar, we can observe slight differences at the image border, and independent processing performs better since MCA-12 encounters synchronization errors.

On the other hand, although independent processing is sufficiently adaptable to correct inconsistency errors, image matching might fail on homogeneous textures or blurred images. Therefore, when the co-registration accuracy is >0.8 pixels, a failsafe procedure that uses the same coefficients as previous successful matching results is adopted to ensure the accuracy. Fig. 10(b) presents a low-texture image of Sequoia that induces a larger error at the image boundary but it is properly corrected after applying the failsafe procedure. Fig. 11 also shows the image registration accuracies of all MS datasets that are processed under the independent mode. Evidently, the accuracy trend is stable within 0.2–0.6 pixels. However, since the BLU image of MCA-12 has a very low image contrast, it has worse accuracy than the others but still has successful image matching and acceptable registration results with accuracy <0.8 pixels. In addition, we can observe the horizontal trend as the failsafe is automatically applied to ensure accuracy.

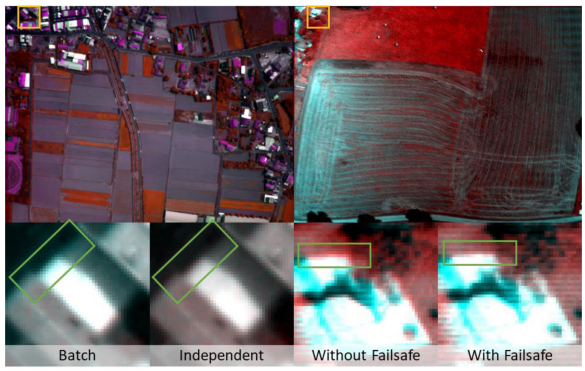

Fig. 10. Analysis of Batch and Independent Processing. (a) Co-registration differences between batch and independent processing. (b) Compare the co-registration results of independent processing with and without failsafe procedure.

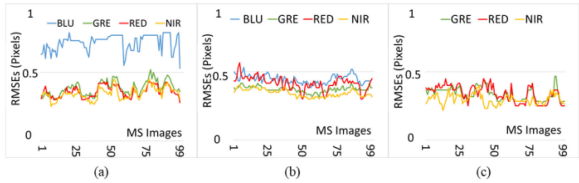

Fig. 11. Image registration accuracies of all MS datasets. (a)–(c) Results of MCA-12, Altum, and Sequoia, respectively.

3) Limitations: Though the developed tool is adaptable for various MSCs, we should be aware of the parallax effects of multilens cameras since they become significant when focusing on closer objects. When mounting an MSC on a UAV where the imaging distance to the ground object is >20 m, the parallax effects between lenses can be ignored and reliable image registration results can be obtained [36]. Otherwise, the parallax effects of MSC become serious when taking ground images due to the parallax value in the image becoming various for object imaging distances, thus performing image registration will result in larger errors.

V. Conclusion

MS image matching is difficult to find sufficient CMs for image registration. In this article, based on the modification of

SURF feature extraction, we present novel N-SURF matching for MSCs image registration. N-SURF can extract more features using a single image scale and it can obtain the required number of features by querying the CDF of FT. The MS image matching evaluation of three representative state-of-the-art MSCs shows that N-SURF has better correctness, can obtain more and well-distributed CMs, and can achieve successful matching on all MS image matching pairs, while SURF failed in some case studies. Meanwhile, to correct the misregistration errors in viewpoints and lens distortion between each lens of MSC, the co-registration results of AT, PT, and EPT are compared through visual comparison, accuracy assessment, and residual analysis. The results show that EPT is the best model for MSC image registration and that it can obtain accurate results within 0.2–0.6 pixels accuracy and have random distributed residuals all <1 pixel.

On the other hand, based on N-SURF and EPT, we have also developed a generalized tool for various MSCs image registration purposes. Through efficiency and visual analysis, it proves that it can rapidly and accurately co-register a large amount of MS images under GPU acceleration and independent processing. Therefore, it has the advantages of accuracy, efficiency, and no requiring prior knowledge for conducting further sensor calibration. Since N-SURF can extract more features and has proven its suitability in MS image matching, the future analysis will focus on more challenging heterogeneous images such as color versus thermal and color vs. synthetic-aperture radar images.

VI. Data Availability Statement

Automatic MS images registration tool and MS evaluation dataset are available at github: https://github.com/jpjhan/RABBIT/tree/master

MS evaluation datasets in google drive: https://reurl.cc/Mdk944

Acknowledgment

The authors were respectively grateful to Prof. Cho-Ying Huang of Department of Geography, National Taiwan University, and Rdata system Company Ltd., for providing a MCA-12, and a RedEdge Altum multispectral camera. They were also grateful to Rdata system Company Ltd., for providing the UAV platform and assisting the sensor integration and data collection.

References

[1] A. Agapiou, D. Hadjimitsis, and D. Alexakis, "Evaluation of broadband and narrowband vegetation indices for the identification of archaeological crop marks," *Remote Sens.,* vol. 4, no. 12, p. 3892, Dec. 2012.

[2] S. Sankaran *et al.*, "Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review," *Eur. J. Agronomy,* vol. 70, pp. 112–123, Oct. 2015.

[3] H. Aasen, A. Burkart, A. Bolten, and G. Bareth, "Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance," *ISPRS J. Photogramm. Remote Sens.,* vol. 108, pp. 245–259, Oct. 2015.

[4] J.-P. Jhan, J.-Y. Rau, and N. Haala, "Robust and adaptive band-to-band image transform of UAS miniature multi-lens multispectral camera," *ISPRS J. Photogramm. Remote Sens.,* vol. 137, pp. 47–60, Mar. 2018.

[5] J. Kelcey and A. Lucieer, "Sensor correction of a 6-band multispectral imaging sensor for UAV remote sensing," *Remote Sens.,* vol. 4, no. 12, pp. 1462–1493, Dec. 2012.

[6] J.-P. Jhan, J.-Y. Rau, and C.-Y. Huang, "Band-to-band registration and ortho-rectification of multilens/multispectral imagery: A case study of MiniMCA-12 acquired by a fixed-wing UAS," *ISPRS J. Photogramm. Remote Sens.,* vol. 114, pp. 66–77, Apr. 2016.

[7] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.,* vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[8] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.,* vol. 60, no. 2, pp. 91–110, Nov. 2004.

[9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up robust features (SURF)," *Comput. Vis. Image Understanding,* vol. 110, no. 3, pp. 346–359, Jun. 2008.

[10] M.-D. Yang, C.-F. Chao, K.-S. Huang, L.-Y. Lu, and Y.-P. Chen, "Image-based 3D scene reconstruction and exploration in augmented reality," *Automat. Construction,* vol. 33, pp. 48–60, Aug. 2013.

[11] P. Loncomilla, J. Ruiz-del-Solar, and L. Martínez, "Object recognition using local invariant features for robotic applications: A survey," *Pattern Recognit.,* vol. 60, pp. 499–514, Dec. 2016.

[12] K. Yan and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.,* 2004, pp. 506–513.

[13] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 1978–1983.

[14] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[15] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM J. Imag. Sci.,* vol. 2, no. 2, pp. 438–469, Jan. 2009.

[16] Y. Pang, W. Li, Y. Yuan, and J. Pan, "Fully affine invariant SURF for image matching," *Neurocomputing,* vol. 85, pp. 6–10, May 2012.

[17] Z. Yang, T. Dan, and Y. Yang, "Multi-Temporal remote sensing image registration using deep convolutional features," *IEEE Access,* vol. 6, pp. 38544–38555, 2018.

[18] Y. Ye and J. Shan, "A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences," *ISPRS J. Photogramm. Remote Sens.,* vol. 90, pp. 83–95, Apr. 2014.

[19] Z. Yi, C. Zhiguo, and X. Yang, "Multi-spectral remote image registration based on SIFT," *Electron. Lett.,* vol. 44, no. 2, pp. 107–108, 2008.

[20] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang, "A novel Coarse-to-Fine scheme for automatic image registration based on SIFT and mutual information," *IEEE Trans. Geosci. Remote Sens.,* vol. 52, no. 7, pp. 4328–4338, 2014.

[21] X. Chang, S. Du, Y. Li, and S. Fang, "A Coarse-to-Fine geometric scale-invariant feature transform for large size high resolution satellite image registration," *Sensors,* vol. 18, no. 5, Apr. 2018. [Online]. Available: https://www.mdpi.com/about/announcements/784

[22] M. Vakalopoulou and K. Karantzalos, "Automatic descriptor-based co-registration of frame hyperspectral data," *Remote Sens.,* vol. 6, no. 4, pp. 3409–3426, 2014.

[23] M. Aladem, S. Baek, and S. A. Rawashdeh, "Evaluation of image enhancement techniques for vision-based navigation under low illumination," *J. Robot.,* vol. 2019, 2019, Art. no. 5015741.

[24] O. Akcay and O. Avsar, "The effect of image enhancement methods during feature detection and matching of thermal images," in *Proc. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 2017, pp. 575–578.

[25] Q. Li, H. Zhang, and T. Wang, "Multispectral image matching using rotation-invariant distance," *IEEE Geosci. Remote Sens. Lett.,* vol. 8, no. 3, pp. 406–410, Jun. 2011.

[26] Q. Li, G. Wang, J. Liu, and S. Chen, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett.,* vol. 6, no. 2, pp. 287–291, Apr. 2009.

[27] W. Ma *et al.*, "Remote sensing image registration with modified SIFT and enhanced feature matching," *IEEE Geosci. Remote Sens. Lett.,* vol. 14, no. 1, pp. 3–7, Jan. 2017.

[28] S. Saleem and R. Sablatnig, "A robust SIFT descriptor for multispectral images," *IEEE Signal Process. Lett.,* vol. 21, no. 4, pp. 400–403, Apr. 2014.

[29] J. P. Jhan, and J. Y. Rau, "A normalized surf for multispectral image matching and band co-registration," in *Proc. ISPRS Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 2019, pp. 393–399.

[30] "Pix4D sample dataset," Pix4D, Johannesburg, South Africa, Jan. 2021. [Online]. Available: https://support.pix4d.com

[31] C.-Y. Huang, H.-L. Wei, J.-Y. Rau, and J.-P. Jhan, "Use of principal components of UAV-acquired narrow-band multispectral imagery to map the diverse low stature vegetation fAPAR," *GISci. Remote Sens.,* vol. 56, no. 4, pp. 605–623, May 2019.
[32] "Alea GPU," Nvidia, Santa Clara, CA, USA, Jan. 2021. [Online]. Available: https://developer.nvidia.com/alea-gpu
[33] D. Gossow, P. Decker, and D. Paulus, "An evaluation of open source SURF implementations," in *Proc. RoboCup 2010: Robot Soccer World Cup XIV*, 2010, pp. 169–179.
[34] J.-Y. Rau and P.-C. Yeh, "A semi-automatic image-based close range 3D modeling pipeline using a multi-camera configuration," *Sensors,* vol. 12, no. 8, Aug. 2012, Art. no. 11271.
[35] C. S. Fraser, "Digital camera self-calibration," *ISPRS J. Photogramm. Remote Sens.,* vol. 52, no. 4, pp. 149–159, Aug. 1997.
[36] J.-P. Jhan, J.-Y. Rau, N. Haala, and M. Cramer, "Investigation of parallax issues for multi-lens multispectral camera band co-registration," in *Proc. ISPRS Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 2017, pp. 157–163.

Jyun-Ping Jhan received the Ph.D. degree in engineering from the National Cheng Kung University, Tainan, Taiwan, in 2017.

He is currently a Post-Doctoral Researcher with the Department of Geomatics, National Cheng Kung University, Tainan, Taiwan. His research interests include in mobile mapping system, sensor integration and calibration, digital image processing, and structural damage investigation.

Jiann-Yeou Rau received the Ph.D. degree in civil engineering from the National Central University (NCU), Taoyuan, Taiwan, in 2002.

He was with the Center for Space and Remote Sensing Research, NCU, as a Research Scientist. He is currently a Professor with the Department of Geomatics, National Cheng Kung University, Tainan, Taiwan. His research activities are mostly concentrated in the domain of digital photogrammetry and mobile mapping technology through different platforms, such as aircraft, UAV, and land vehicle. His research interests include sandbank monitoring and bridge inspection through UAV images, LOD-2 building modeling, 3-D modeling of close-range objects, and other UAV and LiDAR applications.

Dr. Rau is a member of the Chinese Society of Photogrammetry and Remote Sensing and of Image Processing and Pattern Recognition, Taiwan.