

A New Translation Matching Method Based on Autocorrelated Normalized Cross-Power Spectrum

Liangliang Zhu^{ID} and Xiurui Geng^{ID}

Abstract—Translation matching is one of the most fundamental problems in the field of image matching, and the normalized cross-power spectrum (NCPS)-based methods have achieved great success regarding this problem. However, when the images to be matched are seriously corrupted by noise, most current NCPS-based methods cannot obtain satisfactory results. Besides, the 2-D phase extraction of the NCPS, which is required in most NCPS-based methods, may cause an additional error to the final result. In this article, we proposed the concept of autocorrelated NCPS (ANCPS) that is theoretically proved to be able to significantly alleviate the influence of noise and developed a new method based on it. Furthermore, by utilizing the property of equal phase interval of ANCPS, the 2-D phase extraction problem is also naturally avoided in our method. The experiments with simulated and real data demonstrate that the presented method has a better performance in both accuracy and antinoise performance compared with state-of-the-art methods.

Index Terms—Antinoise, autocorrelated normalized cross-power spectrum (ANCPS), translation matching.

I. INTRODUCTION

HIGH-ACCURACY image registration is an important task in many fields [1]–[6], such as remote sensing [7], [8], computer vision [9], and so on [10], [11]. Recently, many methods using the normalized cross-power spectrum (NCPS) [12]–[15] have been developed to address the translation matching problem, which are more accurate and effective than the commonly used correlation methods [16]–[19]. Assume that there are two $M \times N$ images $f(m, n)$, $g(m, n)$, and the displacements between them are m_0 and n_0 . Therefore, the relationship between them can be written as

$$f(m, n) = g(m - m_0, n - n_0). \quad (1)$$

Conducting the discrete Fourier transform (DFT) on both sides of the equation, we have

$$F(u, v) = G(u, v)e^{-j2\pi(um_0/M+vn_0/N)} \quad (2)$$

Manuscript received June 14, 2020; revised September 3, 2020; accepted October 8, 2020. (Corresponding author: Xiurui Geng.)

The authors are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100864, China, also with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100864, China (e-mail: gengxr@sina.com.cn).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2020.3030935

where $F(u, v)$ and $G(u, v)$ are the DFT of $f(m, n)$ and $g(m, n)$, respectively, and u and v are the coordinates in the frequency domain. Then, NCPS is defined as

$$S(u, v) = \frac{F(u, v)G^*(u, v)}{|F(u, v)G^*(u, v)|} = e^{j2\pi(um_0/M+vn_0/N)} \quad (3)$$

where $*$ stands for the complex conjugate. The phase only correlation (POC) function $s(m, n)$ is the inverse DFT (IDFT) of $S(u, v)$

$$s(m, n) = \text{IDFT}(S(u, v)) = \delta(m - m_0, n - n_0) \quad (4)$$

where δ is the impulse function [20]. Typically, we can solve the registration problem by locating the peak of $s(m, n)$. In order to achieve subpixel accuracy, many methods have been developed. All NCPS-based matching methods can be roughly classified into two categories [21].

In the first category [22]–[24], the most commonly used approach to estimate the subpixel offset is based on the determination of (m, n) that maximizes $s(m, n)$. In [25], by padding zeros to NCPS, an upsampled POC function can be obtained. In order to achieve $1/k$ pixel accuracy, the NCPS matrix has to be padded to the size of $kM \times kN$. Therefore, the method becomes much time-consuming when k is large. In [26], an improved method, namely, IDFT-US, is proposed to increase the speed of the upsampling: it first computes the initial estimation of the displacements according to the conventional IDFT method with an upsampling factor $k = 2$, and then, a much larger k is applied to the data in a 1.5×1.5 neighborhood around the initial estimation. In [27], an analytic model of the correlation peak is used to fit the 2-D numerical data array. Because the peak of the POC function is very sharp, usually, $3 \times 3 \sim 9 \times 9$ data points around the peak are enough to achieve high-accuracy function fitting. As is well known, the high-frequency components of NCPS are more easily affected by noise [28]. Therefore, a low-pass-type weighting function can be applied to NCPS for higher accuracy. However, IDFT assumes the noise as additive white Gaussian noise, which is usually inconsistent with the real situation. Thus, the obtained POC function will inevitably deviate from the true value, from which a high-accuracy displacement estimation is almost impossible.

The second category [18], [29], [30] estimates the displacements between two images by linearly fitting the phase of $S(u, v)$, which is a plane in theory after phase unwrapping. However, the 2-D phase unwrapping is much more complicated than the 1-D phase unwrapping. In order to avoid 2-D

phase unwrapping, Stone *et al.* [18] first registered the integer part by using the conventional IDFT method. Then, the least-squares (LS) algorithm was used to fit the phase of $S(u, v)$ filtered by the removal of high-frequency and small-magnitude components, which are more susceptible to the noise. Since the integer part has been registered, the left displacement should be less than one pixel, which means that there is no phase wrap in $S(u, v)$. By introducing the cyclic shift matrix (CSM) to the translation matching problem, Geng and Yang [31] obtained the LS solution of this problem, which is proved to be equivalent to Stone's method. In addition, Geng and Yang [31] also proved that, as there exist unmatched areas between the two images, all NCPS-based methods cannot obtain the exact solution. To solve this problem, the CSM method iteratively applies the CSM to reduce the impact of the unmatched areas. However, the noise in NCPS generally does not obey the Gaussian distribution; thus, when the noise is large, the displacements estimated by LS-based plane fitting will have a large deviation from the exact value. Hoge [32] proposed a straightforward method using singular value decomposition (SVD) to find the dominant rank-one approximation of the NCPS matrix, which can reduce the 2-D plane fitting problem into two 1-D line fitting problems, where the phase unwrapping is more easily solved. However, in large noise cases, the conventional phase unwrapping algorithms may fail to extract an accurate 1-D phase. Ling *et al.* [33] proposed a new phase unwrapping algorithm based on the monotonic and linear variation of the 1-D phase, which can further improve the accuracy of phase unwrapping. In order to deal with the outliers in the unwrapped 1-D phase, Tong *et al.* [21] proposed a method based on Hoge's method and the unified random sample consensus (RANSAC) [34], [35] method, called SVD-RANSAC. By using RANSAC to fit the 1-D phase, SVD-RANSAC can partially eliminate the influence of outliers and had achieved a higher accuracy. However, it should be noted that SVD is also an LS-based method; therefore, in the presence of non-Gaussian noise, the domain rank-one approximation of the NCPS matrix obtained by SVD is not the theoretically optimal one.

As we have mentioned earlier, the two categories of methods both suffer from the impact of noise in NCPS. First, all the LS-based matching methods require the noise to obey the Gaussian distribution, which often deviates from the real situation. Second, when the images to be matched are heavily corrupted by noise, the performance of all existing methods will inevitably deteriorate. Interestingly, we find that the correlations between the components in NCPS can be used to effectively suppress the noise. In this article, we propose the concept of autocorrelated NCPS (ANCPS), and based on this concept, we propose a new image translation matching method, named the ANCPS method, which is expected to further refine the solution of the image matching problem.

This article is organized as follows. The definition of ANCPS and the implementation of ANCPS are presented in Section II. Experiments with simulated and real images are discussed in Section III. The conclusion is given in Section IV. We provide detailed proofs of the lemmas and theorems in Appendix B.

II. METHOD

In this section, we introduce the ANCPS method from the following three aspects. First, the concept of ANCPS is defined, and the relevant theoretical analysis is elaborated. Then, the matching algorithm based on total least-squares (TLS) to estimate the displacements from ANCPS is presented. Finally, the iterative version of the ANCPS method using the CSM is discussed. Details are stated as follows, and the proofs of the lemmas and theorem in this section are demonstrated in Appendix B.

A. Autocorrelated Normalized Cross-Power Spectrum

Noting that (3) only holds in noise-free circumstances; however, the real image always contains more or less noise. Assume that the real images are affected by zero-mean additive Gaussian noise with a constant variance. Let $\hat{F}(u, v)$ and $\hat{G}(u, v)$ be the DFTs of the images with noise, and $N_f(u, v)$ and $N_g(u, v)$ are the phase noise in $\hat{F}(u, v)$ and $\hat{G}(u, v)$, respectively, and then, the NCPS of the two noisy images can be written as

$$\begin{aligned}\hat{S}(u, v) &= \frac{\hat{F}(u, v)\hat{G}^*(u, v)}{|\hat{F}(u, v)\hat{G}^*(u, v)|} = \frac{\hat{F}(u, v)}{|\hat{F}(u, v)|} \frac{\hat{G}^*(u, v)}{|\hat{G}^*(u, v)|} \\ &= \frac{F(u, v)}{|F(u, v)|} e^{jN_f(u, v)} \frac{G^*(u, v)}{|G^*(u, v)|} e^{-jN_g(u, v)} \\ &= e^{j2\pi(um_0/M+vn_0/N)} e^{jN_f(u, v)-jN_g(u, v)} \\ &= S(u, v) e^{jN_s(u, v)}\end{aligned}\quad (5)$$

where $N_s(u, v) = N_f(u, v) - N_g(u, v)$ is the phase of the multiplicative noise contained in $\hat{S}(u, v)$. All existing NCPS-based methods estimate the offset between two images to be matched directly from $\hat{S}(u, v)$ instead of $S(u, v)$. When the images are seriously corrupted by noise, it can be seen from (5) that $\hat{S}(u, v)$ may greatly deviate from $S(u, v)$, which will inevitably lead to a poor accuracy for most NCPS-based matching methods. Therefore, how to obtain a more accurate NCPS from $\hat{S}(u, v)$ is an important and significant problem.

Although $\hat{S}(u, v)$ and $S(u, v)$ differ in phase, as shown in (5), we find that the phases of the expectation of $\hat{S}(u, v)$ and $\hat{S}(u, v)\hat{S}^*(u - \mu, v - \nu)$ are the same as that of $S(u, v)$ and of $S(\mu, v)$, respectively, which is stated in Lemma 1.

Lemma 1: The expectation of $\hat{S}(u, v)$ has the same phase as $S(u, v)$, and the expectation of $\hat{S}(u, v)\hat{S}^*(u - \mu, v - \nu)$ has the same phase as $S(\mu, v)$.

According to this lemma, it can be found that, for a fixed pair (μ, v) , though the phase of $\hat{S}(u, v)\hat{S}^*(u - \mu, v - \nu)$ varies with different (u, v) due to the existence of the phase noise, their expectations all have the same phase as $S(\mu, v)$. For simplicity, we denote

$$R(\mu, v; u, v) = \hat{S}(u, v)\hat{S}^*(u - \mu, v - \nu). \quad (6)$$

By fixing (μ, v) and varying (u, v) , we can obtain a series of $R(\mu, v; u, v)$, whose expectations all have the same phase as $S(\mu, v)$. This motivates us to pursue a more accurate $S(\mu, v)$ by averaging all the above $R(\mu, v; u, v)$. Hence, we propose

the following formula to obtain an estimation of $S(\mu, v)$:

$$\begin{aligned} R(\mu, v) &= \frac{1}{k(\mu, v)} \sum_{u,v} R(\mu, v; u, v) \\ &= S(\mu, v) \frac{\sum_{u,v} e^{jN_s(u,v)-jN_s(u-\mu,v-v)}}{k(\mu, v)} \end{aligned} \quad (7)$$

where $k(\mu, v)$ is the number of sum terms. Interestingly, we can find that $R(\mu, v)$ is actually the autocorrelation of NCPS, which is why our method is named ANCPS. In addition, we have an important conclusion about ANCPS, which is given in Theorem 1.

Theorem 1: The expectation of $R(\mu, v)$ has the same phase as $S(\mu, v)$.

The significance of Theorem 1 lies in that, by using $R(\mu, v)$ instead of $\hat{S}(u, v)$, we can obtain a more accurate estimation of NCPS, from which more robust and accurate displacements can be evaluated. Furthermore, we give a quantitative analysis of the denoising performance of ANCPS, which is stated in Lemma 2.

Lemma 2: If c and σ^2 are assumed to be the mean and variance of the multiplicative noise in NCPS, respectively, then the variance of the multiplicative noise in ANCPS is $(\sigma^4 + 2c^2\sigma^2)/k(\mu, v)$.

In the following, we use a simple experiment to demonstrate the superiority of ANCPS over NCPS. In the experiment, the image shown in Fig. 1(a) is taken as the reference image. Then, we generate the image to be matched by cyclically shifting [31] the reference image with $(5.5, 5.5)$ pixels, as shown in Fig. 1(b). When there is no noise added to these two images, it can be seen from Fig. 1(c) and (d) that ANCPS is totally equivalent to NCPS, which is consistent with the conclusion in Theorem 1. However, when the two images are manually added by zero-mean Gaussian noise with the standard deviation of 0.1, we can find from Fig. 2(c) that the phase of NCPS is severely affected by the noise. The periodicity of the phase is hard to be distinguished from the high-frequency part of NCPS. However, by introducing the concept of ANCPS given by (7), the required phase is almost magically restored in Fig. 2(d), which obviously provides great convenience for the following matching work.

B. Estimating Displacements

In this section, we investigate how to estimate the displacements from ANCPS defined earlier. Noting that additive noise is much easier to be handled than multiplicative noise; thus, we transform the multiplicative noise in ANCPS into additive noise

$$\begin{aligned} R(\mu, v) &= S(\mu, v) \frac{\sum_{u,v} e^{jN_s(u,v)-jN_s(u-\mu,v-v)}}{k(\mu, v)} \\ &= S(\mu, v) \\ &\quad + S(\mu, v) * \left(\frac{\sum_{u,v} e^{jN_s(u,v)-jN_s(u-\mu,v-v)}}{k(\mu, v)} - 1 \right) \\ &= S(\mu, v) + W(\mu, v) \end{aligned} \quad (8)$$

where $W(\mu, v) = S(\mu, v) * (\frac{\sum_{u,v} e^{jN_s(u,v)-jN_s(u-\mu,v-v)}}{k(\mu, v)} - 1)$, and we have the following lemma:

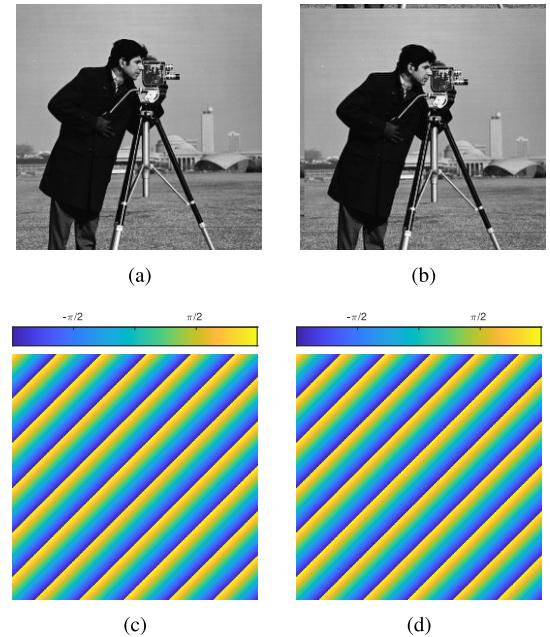


Fig. 1. Comparison of NCPS and ANCPS in absence of noise. (a) Reference image. (b) Image cyclically shifted by $(5.5, 5.5)$ pixels. (c) Phase of NCPS. (d) Phase of ANCPS.

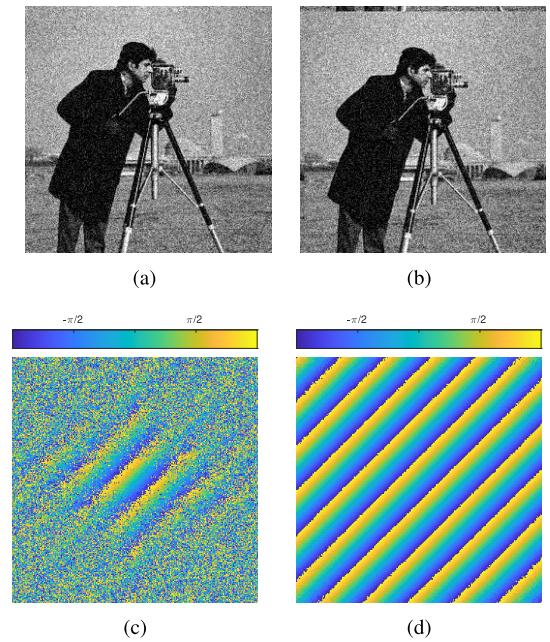


Fig. 2. Comparison of NCPS and ANCPS in presence of noise. (a) Reference image added by noise. (b) Image cyclically shifted by $(5.5, 5.5)$ pixels and added by noise. (c) Phase of NCPS. (d) Phase of ANCPS.

Lemma 3: $W(\mu, v)$ tends to be a Gaussian random variable as the size of images grows larger.

Since $S(\mu, v)$ is a complex number with magnitude of 1, by removing the additive noise from ANCPS, we have the following formula:

$$\begin{aligned} \frac{R(\mu, v) - W(\mu, v)}{R(\mu - 1, v) - W(\mu - 1, v)} &= \frac{S(\mu, v)}{S(\mu - 1, v)} \\ &= S(\mu, v)S^*(\mu - 1, v). \end{aligned} \quad (9)$$

According to (3), it is obvious that $S(\mu, \nu)S^*(\mu - 1, \nu)$ equals $e^{j2\pi m_0/M}$; thus, (9) can be rewritten as

$$R(\mu, \nu) - W(\mu, \nu) = [R(\mu - 1, \nu) - W(\mu - 1, \nu)]e^{j2\pi m_0/M}. \quad (10)$$

From (10), we can find that $e^{j2\pi m_0/M}$ can be estimated by solving a regression problem between two variables $R(\mu - 1, \nu)$ and $R(\mu, \nu)$. Since there exists noise on both variables, we adopt the TLS algorithm instead of the LS algorithm to estimate the coefficient $e^{j2\pi m_0/M}$. The details are given in the following.

Theoretically, the coefficient $e^{j2\pi m_0/M}$ can be evaluated by performing TLS on the observations of $R(\mu, \nu)$ and $R(\mu - 1, \nu)$. However, in practice, it is unnecessary to utilize all pixels in ANCPs to estimate the displacements. Therefore, we assume that l pixels are involved in the estimation of the coefficient, and they are denoted in the form of a vector as follows:

$$\mathbf{q} = [R(\mu_1, \nu_1), R(\mu_2, \nu_2), \dots, R(\mu_l, \nu_l)]^T \quad (11)$$

where (μ_i, ν_i) ($1 \leq i \leq l$) is the coordinate of the i th element in \mathbf{q} . Accordingly, by shifting the coordinate μ , we can obtain another vector \mathbf{p}

$$\mathbf{p} = [R(\mu_1 - 1, \nu_1), R(\mu_2 - 1, \nu_2), \dots, R(\mu_l - 1, \nu_l)]^T. \quad (12)$$

Therefore, (10) can be expressed as the following equation:

$$\mathbf{q} + \mathbf{e}_2 = (\mathbf{p} + \mathbf{e}_1)b_m \quad (13)$$

where \mathbf{e}_1 and \mathbf{e}_2 are the unknown noise vectors and b_m is the coefficient to be estimated. The estimation of b_m can be attributed to the following optimization problem [36]:

$$\begin{aligned} & \arg \min_{\mathbf{e}_1, \mathbf{e}_2} \|[\mathbf{e}_1 \ \mathbf{e}_2]\|_F \\ & \text{s.t. } \mathbf{q} + \mathbf{e}_2 = (\mathbf{p} + \mathbf{e}_1)b_m \end{aligned} \quad (14)$$

where $\|\cdot\|_F$ represents the Frobenius norm. Assuming that the SVD of $[\mathbf{p} \ \mathbf{q}]$ is

$$[\mathbf{p} \ \mathbf{q}] = \mathbf{U}\Sigma\mathbf{V} = \mathbf{U}\Sigma\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \quad (15)$$

where \mathbf{V} is a 2×2 matrix. According to [36], b_m can be estimated by the following formula:

$$b_m = -\mathbf{V}_{12}\mathbf{V}_{22}^{-1}. \quad (16)$$

Then, the displacement m_0 can be simply obtained by extracting the phase of b_m

$$m_0 = \frac{M}{2\pi}\text{angle}(b_m). \quad (17)$$

Similarly, we can obtain the displacement n_0 by

$$n_0 = \frac{N}{2\pi}\text{angle}(b_n) \quad (18)$$

where b_n is the coefficient estimated in the other direction.

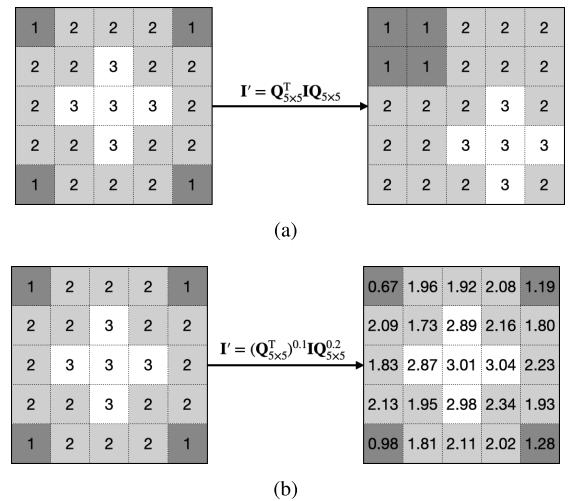


Fig. 3. Simple examples of CSM. (a) Integer pixel displacements with $m = 1$ and $n = 1$. (b) Subpixel displacements with $m = 0.1$ and $n = 0.2$.

C. Eliminating the Influence of Unmatched Areas

In addition to noise, the effect of unmatched areas between two images on the final matching accuracy cannot be ignored. According to the suggestion in [31], we intend to use CSM to mitigate the influence of unmatched areas.

The CSM is a special constructed matrix, which has the form of

$$\mathbf{Q} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & \vdots & \ddots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}. \quad (19)$$

Let \mathbf{I} be an $M \times N$ image, $\mathbf{Q}_{M \times M}$ be an $M \times M$ CSM, and $\mathbf{Q}_{N \times N}$ be an $N \times N$ CSM, and then, the image cyclically shifted by (m, n) pixels can be expressed as

$$\mathbf{I}' = (\mathbf{Q}_{M \times M}^T)^m \mathbf{I} \mathbf{Q}_{N \times N}^n. \quad (20)$$

Fig. 3 shows two examples using the CSM to shift a 5×5 image. Especially, it can be found from Fig. 3(b) that subpixel displacements with any accuracy can be achieved by calculating the corresponding power of the CSM.

In the following, we use a simple example to demonstrate how the CSM iteratively eliminates the influence of the unmatched areas. As shown in Fig. 4, two 5×5 images (\mathbf{I}_1 and \mathbf{I}_2) with displacements of $(1, 1)$ are required to be matched, and their matched areas are framed by red boxes.

It is proven in [31] that due to the existence of unmatched areas, all phase-based subpixel matching methods can only obtain an approximate solution. Assuming that the initial estimation is $(0.6, 0.6)$, then we generate $\mathbf{I}_2^{(1)}$ by cyclically shifting \mathbf{I}_2 by $(-0.6, -0.6)$ pixel. In order to eliminate the impact of unmatched areas, two subimages of \mathbf{I}_1 and $\mathbf{I}_2^{(1)}$, which are framed by the blue boxes, are generated by removing their outermost pixels. The next matching is conducted between these two subimages. It is assumed that the estimated

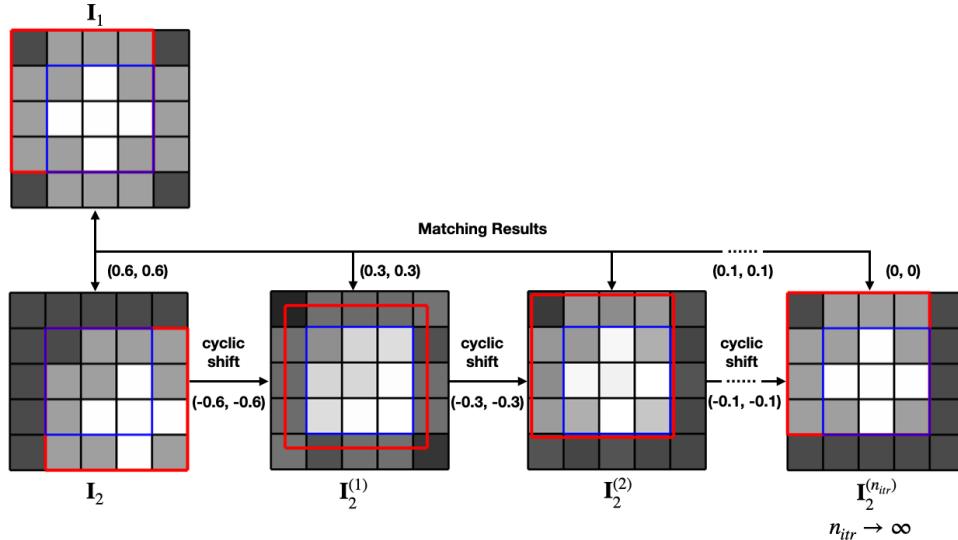


Fig. 4. Workflow of iterative use of CSM, where n_{itr} is the number of iterations.

displacements are $(0.3, 0.3)$, which are apparently smaller than the last estimation, i.e., $(0.6, 0.6)$. By repeating the abovementioned step, we can gradually correct the subpixel displacement estimation, and the total displacement estimation can be obtained by summing up the abovementioned displacement of each step.

It can be found from Fig. 4 that the repeated use of the CSM can effectively reduce the influence of the unmatched areas between the images to be matched. Therefore, in this study, we integrate the CSM into the presented ANCPS algorithm, and the complete pseudocode of the algorithm is given in Algorithm 1.

III. EXPERIMENTS

Simulated and real images are used to evaluate the performance of our method. In the simulated case, image pairs with known offsets are generated from a large image captured by the Ziyuan-3 (ZY-3), which is a Chinese Earth observation satellite launched in January 2012. In addition, multiple other commonly used image registration methods based on NCPS (IDFT-US, Stone, SVD-RANSAC, and CSM) are performed to be compared. For the real image case, a hyperspectral image acquired by a push-broom imaging (PHI) spectrometer is selected, which has a misalignment between the sensors. Thus, there may exist subpixel displacements between different bands in the hyperspectral image, which will lead to spectral distortion. Since there is no ground truth of the displacements between the bands of the image, the displacements estimated with different bands selected as the reference image are compared to demonstrate the performance of our method.

A. Simulated Images

In this section, we generate multiple simulated image pairs to evaluate the performance of our method, namely, ANCPS(n_{itr}), where n_{itr} is the number of iterations. For comparison, we also conduct four other methods on the

Algorithm 1 Pseudocode to Implement the ANCPS-Based Method

Input: the reference image \mathbf{I}_1 , the matching image \mathbf{I}_2 , and the number of iterations n_{itr}
Output: the displacements (m, n)

- 1: Use the DFT based method to calculate the integer part of the displacements (m_0, n_0) , and crop the two images into $\hat{\mathbf{I}}_1$ and $\hat{\mathbf{I}}_2$ according to (m_0, n_0) , so that the remaining displacements between $\hat{\mathbf{I}}_1$ and $\hat{\mathbf{I}}_2$ are smaller than one pixel (prepossessing the integer part of the displacements for faster convergence speed).
- 2: Initialize the decimal part of the displacements as $(m_d = 0, n_d = 0)$.
- 3: **for** $i = 1 : n_{itr}$ **do**
- 4: Calculate ANCPS according to Eq. (7) with the outermost pixels of $\hat{\mathbf{I}}_1$ and $\hat{\mathbf{I}}_2$ removed.
- 5: Generate the corresponding vectors in both μ and ν directions, estimate the parameters according to Eq. (16), and obtain the sub-pixel displacements (m_i, n_i) according to Eqs. (17) and (18).
- 6: Cyclically shift $\hat{\mathbf{I}}_2$ with $(-m_i, -n_i)$ according to Eq. (20).
- 7: $(m_d, n_d) = (m_d, n_d) + (m_i, n_i)$.
- 8: **end for**
- 9: $(m, n) = (m_0, n_0) + (m_d, n_d)$.

simulated image pairs, including IDFT-US, SVD-RANSAC, Stone, and the CSM method [CSM(n_{itr})]. In the IDFT-US method, the upsampling factor k is set to 100 in the following experiments since we find that the performance of IDFT-US does not improve anymore as k is further increased. For the SVD-RANSAC method, a threshold to determine whether a data point is an outlier is needed. In the original article, the threshold was set to 0.01–0.03. However, in our experiments, the noise is much larger, where a small threshold could make the RANSAC algorithm fail to converge. Hence, through

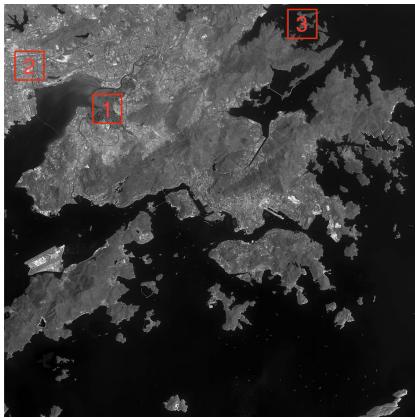


Fig. 5. Hong Kong data (the locations of the three reference images are framed by the red boxes).

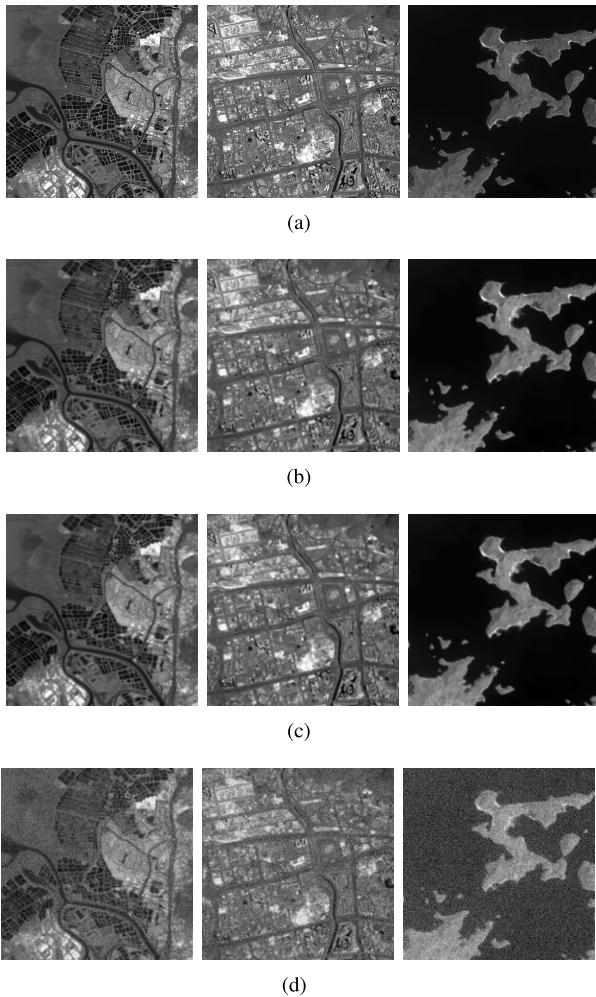


Fig. 6. Procedure of generating the three reference images. (a) Subimages cropped from the original image (1400 × 1400 pixels). (b) Images blurred by a 15 × 15 pixels Gaussian function kernel with $\sigma_g = 5$ (1400 × 1400 pixels). (c) Images downsampled (DDS) by a factor of 7 (200 × 200 pixels). (d) Images added with the Gaussian noise with $\sigma_n = 0.05$ (200 × 200 pixels).

multiple experiments, we find that 0.2 is a relatively good choice. The CSM(n_{itr}) method estimates the displacements by linear fitting the 2-D plane of the phase of NCPS, which is the same as the Stone method. Thus, if the iteration number

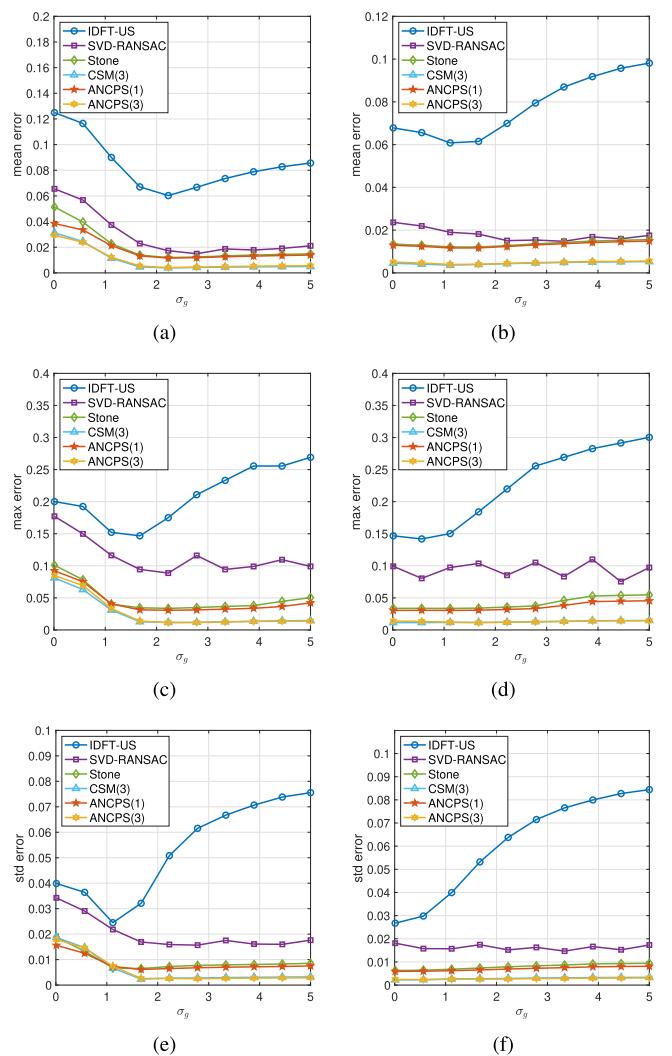


Fig. 7. Results of the aliasing experiment. (a) Mean value of the error in the DDS case. (b) Mean value of the error in the MDS case. (c) Max value of the error in the DDS case. (d) Max value of the error in the MDS case. (e) Standard deviation of the error in the DDS case. (f) Standard deviation of the error in the MDS case.

of CSM(n_{itr}) is set to 1, it is exactly equivalent to the Stone method. In addition, as recommended in SVD-RANSAC, Stone, and CSM(n_{itr}), a frequency mask should be added to NCPS to remove high-frequency components. In the original version of SVD-RANSAC and Stone, the radius of the frequency mask is set to $0.3 \times \min(M, N)$, while, in CSM(n_{itr}), the radius is set to $0.25 \times \min(M, N)$. Through experiments, we find that SVD-RANSAC shows better performance when the radius is set to $0.3 \times \min(M, N)$, and both Stone and CSM(n_{itr}) show better performance when the radius is set to $0.25 \times \min(M, N)$. Hence, in the following, the radius of the frequency mask in SVD-RANSAC is set to $0.3 \times \min(M, N)$, and in Stone and CSM(n_{itr}), it is set to $0.25 \times \min(M, N)$. A raised-cosine window is applied to the images to be matched, and a phase fringe filter with a size of 5×5 is applied to NCPS in SVD-RANSAC as recommended. In Stone and CSM(n_{itr}), it is recommended to remove abnormal pixels before linear fitting; thus, a 5×5 median filter is applied to NCPS. For fair comparison, a frequency mask with a radius

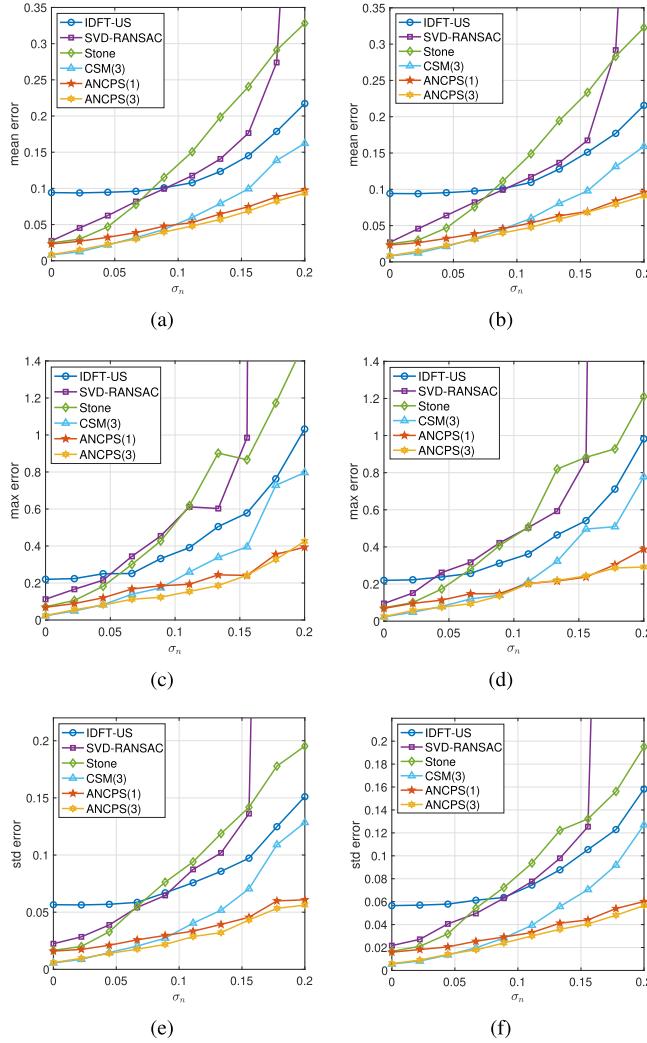


Fig. 8. results of the antinoise experiment. (a) Mean value of the error in the DDS case. (b) Mean value of the error in the MDS case. (c) Max value of the error in the DDS case. (d) Max value of the error in the MDS case. (e) Standard deviation of the error in the DDS case. (f) Standard deviation of the error in the MDS case.

of $0.25 \times \min(M, N)$ is also applied in NCPS, and the pixels within $(\mu^2 + v^2)^{1/2} \leq \min(M, N)/8$ are selected to estimate the displacements. Finally, the iteration numbers of CSM(n_{itr}) and ANCPS(n_{itr}) are both set to 3. In addition, ANCPS(1), which does not use the CSM, is also conducted to intuitively show the effect of ANCPS.

The image with the size of 8000×8000 pixels and with a spatial resolution of 2.5 m is used in the following experiments, which is captured by the nadir sensor of ZY-3. The scene of the image is located in Hong Kong of China; thus, it is referred to as the Hong Kong data. The following steps are adopted to generate image pairs with subpixel displacements. First, two 1400×1400 subimages are cropped from the 8000×8000 image at intervals of (s_x, s_y) pixels. Then, the cropped images are blurred by a 2-D Gaussian function with the parameter σ_g . Then, the images are downsampled by a factor of t so that the downsampled images have the displacements of $(s_x/t, s_y/t)$. Two downsampling methods are used in the following experiments: the first is directly downsampling,

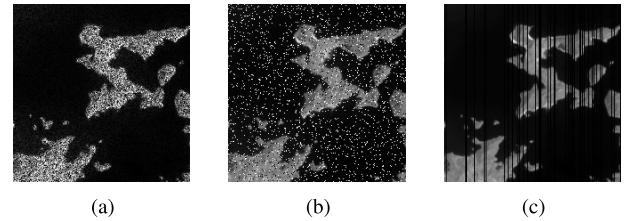


Fig. 9. Examples of images with different types of noise added. (a) Multiplicative Gaussian noise with standard deviation of 0.6. (b) Salt and pepper noise with 0.1% pixels corrupted. (c) Strip noise with 0.2% columns corrupted.

namely, DDS, and the second is mean downsampling, namely, MDS, where the mean value of a region with the size of $t \times t$ pixels is used to represent the region. Finally, the two images are normalized to $[0, 1]$, and the Gaussian noise with a standard deviation of σ_n is added to the images.

The estimation error e is defined as the two-norm of the offset between the estimation and the ground truth, which is given by $e = (e_x^2 + e_y^2)^{1/2}$, where e_x and e_y are the error in the x - and y -directions, respectively. In order to obtain a result in the statistical sense, multiple experiments with different σ_g 's, σ_n 's, and displacements are conducted for each method, and three metrics were used to evaluate their performance: the mean value of e , the max value of e , and the standard deviation of e .

The locations of three reference images are shown in Fig. 5, and Fig. 6 shows how the reference images are generated. In the following experiments, t is set to 7, and the size of the 2-D Gaussian function is set to 15×15 pixels. We select 0, 5, 10, 15, and 20 as integer displacements, and each integer displacement is combined with all possible decimal displacements. Therefore, for each reference image, $5 \times (7 - 1) \times (7 - 1) = 180$ images with different subpixel displacements are generated.

1) *Aliasing Experiment:* In this experiment, σ_n is set to 0, σ_g ranges from 0.01 to 5, and the results are shown in Fig. 7. When σ_g is small, it can be seen from Fig. 7(a), (c), and (e) that the three metrics of all methods gradually decrease as σ_g increases in the DDS case. When σ_g is large, except IDFT-US, all other methods show a stable performance. The performance of ANCPS(1) is apparently better than that of IDFT-US and of SVD-RANSAC and is slightly better than that of Stone. Due to the existence of unmatched areas, ANCPS(1) has a larger estimation error than CSM(3). However, by utilizing the CSM, the accuracy of ANCPS(3) can reach the same level as that of CSM(3). From Fig. 7(b), (d), and (f), we can find that when σ_g is larger than 2, the curves of the three metrics in the MDS case are similar to those in the DDS case.

2) *Antinoise Experiment:* In this experiment, σ_g is set to a fixed value ($\sigma_g = 5$), σ_n ranges from 0 to 0.2, and the results are shown in Fig. 8. It can be seen from Fig. 8(a), (c), and (e) that with an increase of the standard deviation of the noise, the estimation error of all methods inevitably becomes larger in the DDS case. We can also find that the performance of SVD-RANSAC deteriorates a lot when σ_g is close to 0.2, which is found to be caused by the failure of the 1-D phase unwrapping in large noise cases. From Fig. 8(a), (c), and (e),

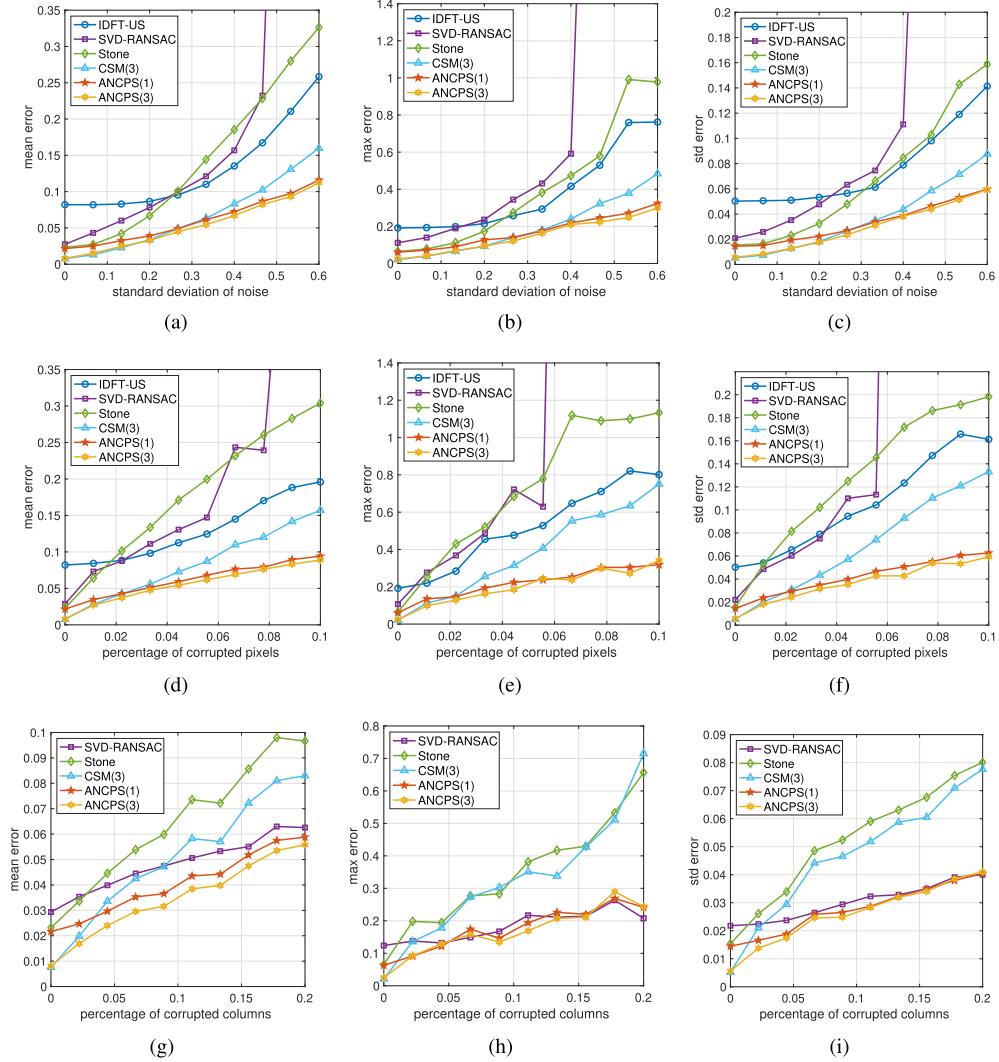


Fig. 10. Antinoise experiment with different types of noise added. (a)–(c) Multiplicative Gaussian noise. (d)–(f) Salt and pepper noise. (g)–(i) Strip noise.

it can be found that in comparison with IDFT-US, SVD-RANSAC, and Stone, ANCPS(1) can obtain a higher matching accuracy. However, in a small noise case, ANCPS(1) is inferior to CSM(3) in the matching accuracy. It indicates that, in this case, the unmatched areas have a larger impact on the result than the noise. As the standard deviation of the noise increases, ANCPS(1) gradually shows its superiority over CSM(3). It is worth noting that because ANCPS(3) considers the impact of both noise and the unmatched areas, its accuracy is always higher than other methods. From Fig. 8, we can find that the three metrics of all methods share a similar tendency with different downsampling methods adopted.

Except for the additive Gaussian noise, we also add other types of noise to the images, including multiplicative Gaussian noise, salt and pepper noise, and strip noise. For multiplicative Gaussian noise, the mean value of it is set to 1 and the standard deviation of it ranges from 0 to 0.6. For the salt and pepper noise, the percentage of corrupted pixels ranges from 0 to 0.1. For the strip noise, we randomly remove some columns of an image, and the percentage of corrupted columns ranges from 0 to 0.2. The examples of images added by different types of

noise are shown in Fig. 9, and the quantitative comparisons are shown in Fig. 10. Noting that, in Fig. 10(g)–(i), the error of IDFT-US is much larger than other methods; thus, its curves are not included in the figures. In addition, because different downsampling methods have a small influence on the antinoise performance, only DDS is used in this experiment.

From Fig. 10, we can find that the trends of the three metrics of all methods with different types of noise are similar to that in Fig. 8. It is worth noting that, in the strip noise case, SVD-RANSAC shows a stable performance. As can be seen from Fig. 10 that ANCPS(1) and ANCPS(3) are both robust for different types of noise, especially for the salt and pepper noise. Noting that, comparing with ANCPS(1), ANCPS(3) also considers the unmatched areas; therefore, it always has the highest accuracy.

In addition, the computational time of all these methods is listed in Table I. All the abovementioned algorithms are programmed in MATLAB on a computer with a 2.0-GHz CPU and 8-GB memory. As can be seen from Table I, among all these methods, ANCPS(3) has the highest computation complexity. However, considering its powerful performance

TABLE I
COMPUTATION TIME OF ONE RUN FOR ALL METHODS

Methods	Total time (ms)
IDFT-US	6.75
SVD-RANSAC	8.74
Stone	7.63
CSM(3)	18.46
ANCPS(1)	17.13
ANCPS(3)	46.74



Fig. 11. True-color image of the farmland data.

(especially in terms of noise immunity), such a time cost is acceptable in many cases.

B. Real Hyperspectral Image

For hyperspectral images acquired by the PHI spectrometer, there may exist subpixel displacements between different bands; thus, it is necessary to estimate and calibrate the displacements to prevent spectral distortion. In this section, the hyperspectral image acquired by the PHI spectrometer developed by the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, is used to evaluate the performance of our method.

The scene in the hyperspectral image is covered by a variety of farmland; thus, it is referred to as farmland data. The farmland data have a size of 350×570 pixels and 80 bands ranging from 411.9 to 832.79 nm. The true-color image of this data is shown in Fig. 11.

In this experiment, the first and last few bands are removed because of their low signal-to-noise rate, i.e., only bands 10–60 are selected. To evaluate the performance of the methods, each band is selected as the reference image, and the displacements between the reference band and left bands are estimated. In order to compare the displacements estimated with different reference images, the displacement vector of band 10 is always set to $[0, 0]^T$ by a simple shifting operation. For example, let band i be the reference image, and then, the displacement vector between bands i and j can be estimated, which is denoted $\mathbf{d}_{i,j}$. Thus, for this reference image, a 2×51 displacement matrix $[\mathbf{d}_{i,10}, \mathbf{d}_{i,11}, \dots, \mathbf{d}_{i,60}]$ can be obtained. Then, an adjusted displacement matrix $[\mathbf{0}, \mathbf{d}_{i,11} - \mathbf{d}_{i,10}, \dots, \mathbf{d}_{i,60} - \mathbf{d}_{i,10}]$ is generated by subtracting $\mathbf{d}_{i,10}$ from all columns. Finally, by calculating the norm of each column of the adjusted matrix, we can obtain a displacement vector $[\mathbf{0}, ||\mathbf{d}_{i,11} - \mathbf{d}_{i,10}||, \dots, ||\mathbf{d}_{i,60} - \mathbf{d}_{i,10}||]$ for band i . Noting that one of the 51 bands is selected as the reference image at a time;

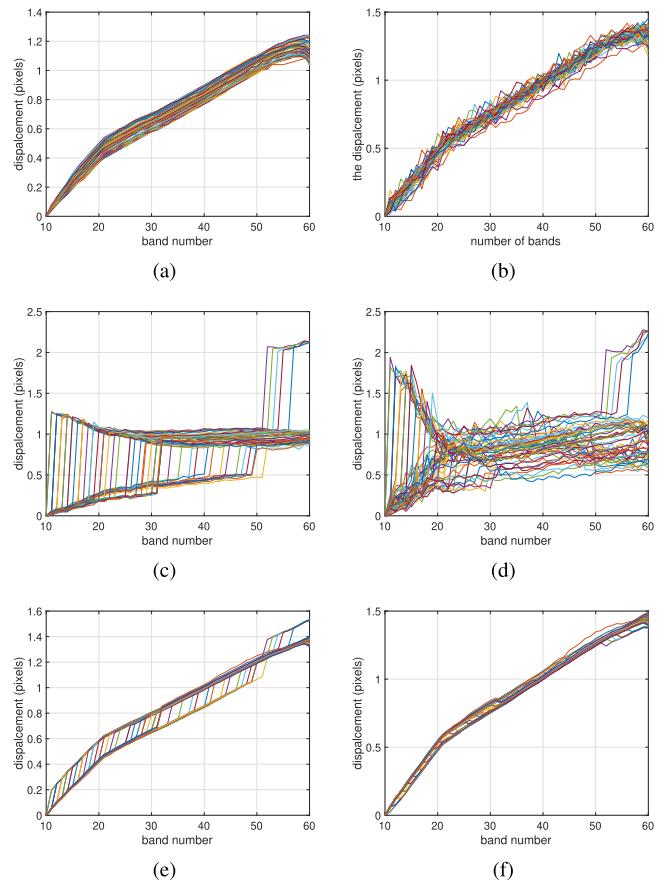


Fig. 12. Displacement curves. (a) IDFT-US. (b) SVD-RANSAC. (c) Stone. (d) CSM(3). (e) ANCPS(1). (f) ANCPS(3).

TABLE II
VARIANCE OF THE DISPLACEMENT CURVES FOR ALL METHODS

Methods	Variance
IDFT-US	0.2683
SVD-RANSAC	0.2493
Stone	1.9718
CSM(3)	1.8119
ANCPS(1)	0.3838
ANCPS(3)	0.1544

thus, for each method, 51 displacement curves are calculated. It is obvious that the closer the displacement curves of different bands are, the better the displacements are estimated.

In this experiment, the parameters of the methods are the same as those in the simulated images experiment. From the results shown in Fig. 12, we can find that the displacement curves of ANCPS(3) vary in a very small range, which can be further quantitatively demonstrated by the standard deviation of the curves shown in Table II. This indicates that the displacement estimated by ANCPS(3) is the most accurate one.

To intuitively show the performance of ANCPS(3), the displacements between bands in horizontal and vertical directions with band 10 selected as the reference image are plotted in Fig. 13. From the figure, it can be found that the horizontal displacement is close to zero, while the vertical displacement gradually increases with the band number. The matching

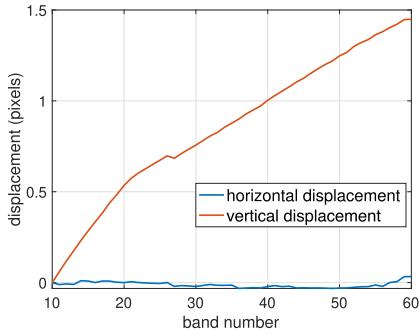


Fig. 13. Horizontal and vertical displacements between different bands for the farmland data (band 10 is selected as the reference band).

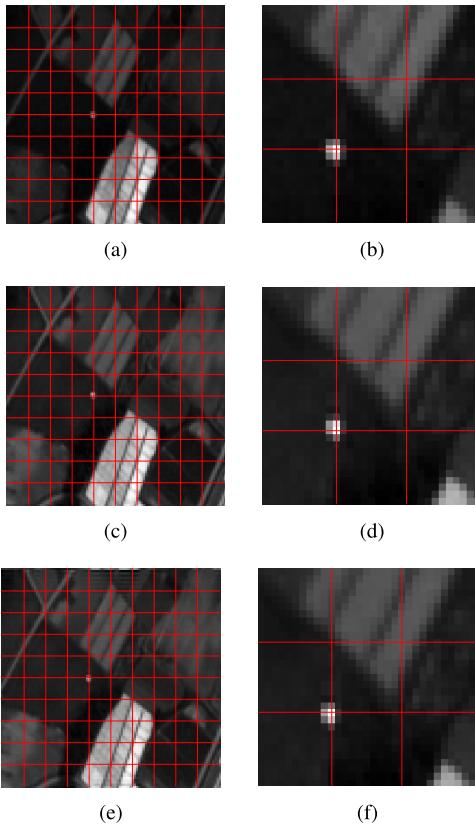


Fig. 14. the matching result between band 10 and band 30 of part of the farmland data. (a) Band 10 (reference image). (b) Local magnification of band 10. (c) Band 30. (d) Local magnification of band 30. (e) Calibrated band 30. (f) Local magnification of calibrated band 30.

result is consistent with the working principle of the PHI spectrometer, which moves perpendicular to the linear array of the sensor while acquiring the data.

For better visualization, a small square area locating at the left top of the farmland data is cut out, and the matching result between band 10 and band 30 is shown in Fig. 14. Comparing with Fig. 14(a) and (c) or the local magnifications of them, it can be found that Fig. 14(c) has a little offset from Fig. 14(a) in the vertical direction. The displacements estimated by ANCPS(3) are 0.612 (vertical) and -0.016 (horizontal) pixels, which is roughly consistent with a visual inspection. The calibrated image of Fig. 14(c) is shown in Fig. 14(e), which is well-aligned with Fig. 14(a) in both horizontal and vertical directions.

IV. CONCLUSION

Traditional NCPS-based image matching methods suffer from noise. Especially, when the images are heavily corrupted by noise, it is difficult for all existing NCPS-based image matching methods to achieve satisfactory results. In this article, the concept of ANCPS is proposed to alleviate the influence of the noise in NCPS, and the experiments show that ANCPS has a powerful antinoise performance. In addition, a new displacement estimation algorithm, which can avoid phase extraction, is proposed to estimate the displacements from ANCPS. In order to eliminate the influence of the unmatched areas, the CSM is also integrated into our method. Moreover, besides the translation matching problem, ANCPS could also be used to improve the performance of the phase-based rotation and affine matching methods.

APPENDIX

A. Basic Knowledge

In this section, some preliminary knowledge required to prove the lemmas and theorem in the Method section is discussed.

Property 1: After DFT, the zero-mean Gaussian noise added to the image will be turned into complex Gaussian noise, whose magnitude is a Rayleigh distributed random variable and whose phase is uniformly distributed from $-\pi$ to π .

Proof: Let us take a 1-D time series as an example for simplicity. Assume that $\{x_n\}$ is a sequence that contains N independent zero-mean Gaussian distributed random variables with variance of σ^2 . Then, DFT of the sequence is defined as

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi nk/N}. \quad (21)$$

Focusing, for now, on the real part, we have

$$\text{Real}(X_k) = \sum_{n=0}^{N-1} x_n \cos(2\pi nk/N). \quad (22)$$

It is obvious that the real part of X_k is the sum of a sequence of independent Gaussian random variables. Therefore, the distribution of the real part of X_k is still Gaussian. Similarly, we have that the imaginary part of X_k

$$\text{Imag}(X_k) = \sum_{n=0}^{N-1} x_n \sin(2\pi nk/N) \quad (23)$$

is also a Gaussian random variable and has the same variance as the real part

$$\begin{aligned} \text{Var}[\text{Real}(X_k)] &= E[(\sum_{n=0}^{N-1} x_n \cos(2\pi nk/N))^2] \\ &= E[\sum_{n=0}^{N-1} x_n^2 \cos^2(2\pi nk/N)] \\ &= \sigma^2 \sum_{n=0}^{N-1} \cos^2(2\pi nk/N) \\ &= \sigma^2 N/2 \end{aligned}$$

$$= \sigma^2 \sum_{n=0}^{N-1} \sin^2(2\pi nk/N) \\ = \text{Var}[\text{Imag}(X_k)]. \quad (24)$$

In addition, we also have

$$\begin{aligned} & \text{Cov}[\text{Real}(X_k), \text{Imag}(X_k)] \\ &= E \left[\sum_{n=0}^{N-1} x_n \cos(2\pi nk/N) \sum_{n=0}^{N-1} x_n \sin(2\pi nk/N) \right] \\ &= \sigma^2 \sum_{n=0}^{N-1} \cos(2\pi nk/N) \sin(2\pi nk/N) = 0. \end{aligned} \quad (25)$$

Therefore, the real part and imaginary part of X_k are uncorrelated. According to [37], for such a complex Gaussian random variable, its magnitude is a Rayleigh distributed random variable, and its phase is uniformly distributed from $-\pi$ to π . \square

Property 2: Let $ce^{j\theta_c} = ae^{j\theta_a} + be^{j\theta_b}$, where θ_b is a uniformly distributed random variable from $-\pi$ to π , and a, b , and θ_a are already known. Then, the probability density function (pdf) of θ_c is symmetric about θ_a .

Proof: For simplicity, θ_c is written as a continuous function of θ_b

$$\theta_c = h(\theta_b). \quad (26)$$

From Fig. 15, we can find that function $h(\cdot)$ satisfies

$$2\theta_a - \theta_c = 2\theta_a - h(\theta_b) = h(2\theta_a - \theta_b). \quad (27)$$

Because θ_b is a uniformly distributed random variable from $-\pi$ to π , the pdf of θ_b has a constant value $1/2\pi$. Assume that the cumulative distribution function (CDF) of θ_c is $F_{\theta_c}(\theta)$, which can be calculated by

$$F_{\theta_c}(\theta) = \int_{h(\theta_b) \leq \theta} \frac{1}{2\pi} d\theta_b. \quad (28)$$

Substitute θ_b with $2\theta_a - \theta_c$, and then, we have

$$\begin{aligned} F_{\theta_c}(\theta) &= - \int_{h(2\theta_a - \theta_c) \leq \theta} \frac{1}{2\pi} d(2\theta_a - \theta_c) \\ &= \int_{h(\theta_b) \geq 2\theta_a - \theta} \frac{1}{2\pi} d\theta_b \\ &= 1 - F_{\theta_c}(2\theta_a - \theta). \end{aligned} \quad (29)$$

Noting that pdf is the differential of CDF, therefore, the pdf of θ_c satisfies

$$p_{\theta_c}(\theta) = F'_{\theta_c}(\theta) = -F'_{\theta_c}(2\theta_a - \theta) = p_{\theta_c}(2\theta_a - \theta) \quad (30)$$

which means that the distribution of θ_c is symmetric about θ_a . \square

Property 3: Let $\{X_n\}$ be a series that contains N independent random variables, and assume that the expectation $E[X_n] = \mu_n$ and the variance $\text{Var}[X_n] = \sigma_n^2$ exist and are finite. Also, let $s_N^2 = \sum_{n=1}^N \sigma_n^2$. If the series of independent random variables satisfies Lindeberg's condition

$$\lim_{N \rightarrow \infty} \frac{1}{s_N^2} \sum_{n=1}^N E[(X_n - \mu_n)^2 \cdot \mathbf{1}_{\{|X_n - \mu_n| > \epsilon s_N\}}] = 0 \quad (31)$$

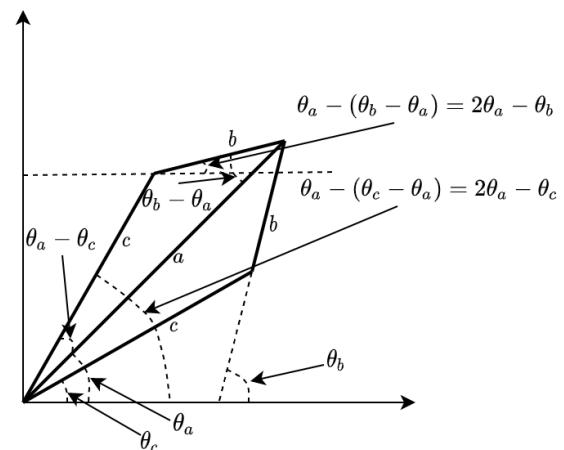


Fig. 15. Illustration of the relationship between θ_a , θ_b , and θ_c .

for all $\epsilon > 0$, where $\mathbf{1}_{\{\cdot\}}$ is the indicator function, then the central limit theorem holds, i.e., the random variable

$$\frac{\sum_{n=1}^N (X_n - \mu_n)}{s_N} \quad (32)$$

converges in distribution to a standard normal random variable as $N \rightarrow \infty$ [38].

B. Detailed Proofs

In this section, the detailed proofs of the lemmas and theorem in the method section are demonstrated.

Proof: The expectation of $\hat{S}(u, v)$ has the same phase as $S(u, v)$, and the expectation of $\hat{S}(u, v)\hat{S}^*(u - \mu, v - v)$ has the same phase as $S(\mu, v)$.

Proof: According to Property 1, the DFT of the image $f(m, n)$ added by zero-mean Gaussian noise can be written as

$$\hat{F}(u, v) = F(u, v) + A(u, v)e^{jN(u, v)} \quad (33)$$

where $A(u, v)$ is a Rayleigh distributed random variable and $N(u, v)$ uniformly distributes from $-\pi$ to π . To obtain the pdf of the phase noise in $\hat{F}(u, v)$, we first consider the random variable $N(u, v)$, and assume that $A(u, v)$ is a constant value. According to Property 2, the conditional pdf of the phase of $\hat{F}(u, v)$ is symmetric about the phase of $F(u, v)$; in other words, the conditional pdf of $N_f(u, v)$ (the phase of the multiplicative noise in $\hat{F}(u, v)$) is symmetric about zero. Let $p_{N_f(u, v)|A(u, v)}(\theta|a)$ be the pdf of $N_f(u, v)$ on the condition of $A(u, v)$, and it is obvious that, for any a , $p_{N_f(u, v)|A(u, v)}(\theta|a)$ is always symmetric about zero. Let $p_{A(u, v)}(a)$ be the pdf of $A(u, v)$, and then, we have that the pdf of $N_f(u, v)$ satisfies

$$\begin{aligned} p_{N_f(u, v)}(\theta) &= \int p_{N_f(u, v)|A(u, v)}(\theta|a) p_{A(u, v)}(a) da \\ &= \int p_{N_f(u, v)|A(u, v)}(-\theta|a) p_{A(u, v)}(a) da \\ &= p_{N_f(u, v)}(-\theta). \end{aligned} \quad (34)$$

Thus, the pdf of $N_f(u, v)$ is symmetric about zero. Similarly, we have that the pdf of $N_g(u, v)$ [the phase of the multiplicative noise in $\hat{G}(u, v)$] is also symmetric about zero. Because $N_s(u, v) = N_f(u, v) - N_g(u, v)$, we can simply infer

that the pdf of $N_s(u, v)$ is symmetric about zero. Therefore, the expectation of $\hat{S}(u, v)$ is

$$\begin{aligned} E[\hat{S}(u, v)] &= E[S(u, v)e^{jN_s(u, v)}] \\ &= S(u, v)E[\cos N_s(u, v) + j \sin N_s(u, v)] \\ &= S(u, v)E[\cos N_s(u, v)]. \end{aligned} \quad (35)$$

Because $E[\cos N_s(u, v)]$ is a real number, the phase of $\hat{S}(u, v)$ is the same as that of $S(u, v)$.

In addition, the noise in NCPS is assumed to be independent; thus, we have

$$\begin{aligned} E[\hat{S}(u, v)\hat{S}^*(u - \mu, v - \nu)] &= E[\hat{S}(u, v)]E[\hat{S}^*(u - \mu, v - \nu)] \\ &= S(u, v)S^*(u - \mu, v - \nu) \\ &\quad \cdot E[\cos N_s(u, v)]E[\cos N_s(u - \mu, v - \nu)] \\ &= S(\mu, \nu)E[\cos N_s(u, v)]E[\cos N_s(u - \mu, v - \nu)]. \end{aligned} \quad (36)$$

Therefore, the expectation of $\hat{S}(u, v)\hat{S}^*(u - \mu, v - \nu)$ has the same phase as $S(\mu, \nu)$. \square

Theorem 1: The expectation of $R(\mu, \nu)$ has the same phase as $S(\mu, \nu)$.

Proof: According to Lemma 1, the phase of the expectation of $R(\mu, \nu; u, v)$ is the same as that of $S(\mu, \nu)$

$$\begin{aligned} E[R(\mu, \nu; u, v)] &= S(\mu, \nu)E[\cos N_s(u, v)]E[\cos N_s(u - \mu, v - \nu)]. \end{aligned} \quad (37)$$

Therefore, the expectation of $R(\mu, \nu)$ is

$$\begin{aligned} E[R(\mu, \nu)] &= E\left[\frac{1}{k(\mu, \nu)}\sum_{u,v} R(\mu, \nu; u, v)\right] \\ &= \frac{1}{k(\mu, \nu)}\sum_{u,v} E[R(\mu, \nu; u, v)] \\ &= S(\mu, \nu)\frac{\sum_{u,v} E[\cos N_s(u, v)]E[\cos N_s(u - \mu, v - \nu)]}{k(\mu, \nu)}. \end{aligned} \quad (38)$$

Thus, the phase of $E[R(\mu, \nu)]$ is the same as that of $S(\mu, \nu)$. \square

Lemma 2: If c and σ^2 are assumed to be the mean and variance of the multiplicative noise in NCPS, separately, then the variance of the multiplicative noise in ANCPS is $(\sigma^4 + 2c^2\sigma^2)/k(\mu, \nu)$.

Proof: According to Lemma 1, the mean of the multiplicative noise in NCPS c is a real number

$$E[e^{jN_s(u, v)}] = E[\cos N_s(u, v)] = c. \quad (39)$$

In addition, the multiplicative noise in NCPS is assumed to be independent and have a constant variance

$$\text{Var}[e^{jN_s(u, v)}] = \sigma^2. \quad (40)$$

Then, the variance of $e^{jN_s(u, v) - jN_s(u - \mu, v - \nu)}$ can be easily obtained

$$\begin{aligned} \text{Var}[e^{jN_s(u, v) - jN_s(u - \mu, v - \nu)}] &= \text{Var}[e^{jN_s(u, v)}e^{-jN_s(u - \mu, v - \nu)}] \end{aligned}$$

$$\begin{aligned} &= E[(e^{jN_s(u, v)}e^{-jN_s(u - \mu, v - \nu)} - c^2)^2] \\ &= E[(e^{jN_s(u, v)}e^{-jN_s(u - \mu, v - \nu)})^2] - c^4 \\ &= (\text{Var}[e^{jN_s(u, v)}] + c^2)(\text{Var}[e^{jN_s(u - \mu, v - \nu)}] + c^2) - c^4 \\ &= \sigma^4 + 2c^2\sigma^2. \end{aligned} \quad (41)$$

Therefore, assume that the noise in $R(\mu, \nu; u, v)$ with different u 's and v 's is independent, and the variance of the multiplicative noise in ANCPS is

$$\begin{aligned} &\text{Var}\left[\frac{\sum_{u,v} e^{jN_s(u, v) - jN_s(u - \mu, v - \nu)}}{k(\mu, \nu)}\right] \\ &= \frac{1}{k(\mu, \nu)^2} \sum_{u,v} \text{Var}[e^{jN_s(u, v) - jN_s(u - \mu, v - \nu)}] \\ &= \frac{\sum_{u,v} (\sigma^4 + 2c^2\sigma^2)}{k^2(\mu, \nu)} \\ &= \frac{\sigma^4 + 2c^2\sigma^2}{k(\mu, \nu)}. \end{aligned} \quad (42)$$

\square

Lemma 3: $W(\mu, \nu)$ tends to be a Gaussian random variable as the images grow larger.

Proof: For simplicity, we denote $N_r(\mu, \nu; \mu, \nu) = N_s(\mu, \nu) - N_s(\mu - \nu, \nu - \nu)$. Noting that additive noise $W(\mu, \nu)$ is the sum of a series of random variables

$$\begin{aligned} W(\mu, \nu) &= S(\mu, \nu)\left[\frac{\sum_{u,v} e^{jN_r(\mu, \nu; \mu, \nu)}}{k(\mu, \nu)} - 1\right] \\ &= S(\mu, \nu)\frac{\sum_{u,v} [e^{jN_r(\mu, \nu; \mu, \nu)} - 1]}{k(\mu, \nu)}. \end{aligned} \quad (43)$$

Let $W(\mu, \nu; u, v) = e^{jN_r(\mu, \nu; \mu, \nu)} - E[e^{jN_r(\mu, \nu; \mu, \nu)}]$, and it is obvious that $W(\mu, \nu; u, v)$ is a zero-mean random variable. In addition, $|W(\mu, \nu; u, v)| \leq 2$; thus, the variance of $W(\mu, \nu; u, v)$ is limited. Noting that the Gaussian noise in the spatial domain has a small influence on the low-frequency domain and a large influence on the high-frequency domain, therefore, $\sigma^2(\mu, \nu; u, v)$, the variance of $W(\mu, \nu; u, v)$, tends to be a positive constant value as u, v grow larger (the zero frequency is at the point $u = 0, v = 0$). However, if the sum of the series exists, the series must converge to zero. Therefore, the sum of the variance series tends to be infinity

$$\lim_{M, N \rightarrow \infty} z^2(\mu, \nu) = \sum_{u,v} \sigma^2(\mu, \nu; u, v) = \infty \quad (44)$$

which means that, for any $\epsilon > 0$, the possibility of $|W(\mu, \nu; u, v)|$ to be greater than $\epsilon z(\mu, \nu)$ tends to be zero

$$\lim_{M, N \rightarrow \infty} P[|W(x, y; u, v)| > \epsilon z(x, y)] = 0. \quad (45)$$

Then, we have Lindeberg's condition

$$\lim_{M, N \rightarrow \infty} \frac{\sum_{u,v} E[W^2(\mu, \nu; u, v) \cdot \mathbf{1}_{\{|W(\mu, \nu; u, v)| > \epsilon z(\mu, \nu)\}}]}{z^2(\mu, \nu)} = 0. \quad (46)$$

Finally, if the components in the series of $W(\mu, \nu; u, v)$ are assumed to be independent, $\sum_{u,v} W(\mu, \nu; u, v)$ tends to be a zero-mean Gaussian random variable according to Property 3. Therefore, $W(\mu, \nu)$ tends to be a Gaussian random variable. \square

REFERENCES

- [1] Z. Yi, C. Zhiguo, and X. Yang, "Multi-spectral remote image registration based on SIFT," *Electron. Lett.*, vol. 44, no. 2, pp. 107–108, Jan. 2008.
- [2] Z. Hossein-Nejad and M. Nasri, "RKEM: Redundant keypoint elimination method in image registration," *IET Image Process.*, vol. 11, no. 5, pp. 273–284, Apr. 2017.
- [3] K. A. J. Eppenhof, M. W. Lafarge, P. Moeskops, M. Veta, and J. P. W. Pluim, "Deformable image registration using convolutional neural networks," in *Proc. Med. Imag., Image Process.*, vol. 10574, 2018, pp. 192–197.
- [4] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Isgum, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. New York, NY, USA: Springer, 2017, pp. 204–212.
- [5] G. Tzimiropoulos, V. Argyriou, S. Zafeiriou, and T. Stathaki, "Robust FFT-based scale-invariant image registration with image gradients," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1899–1906, Oct. 2010.
- [6] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, pp. 95–113, Oct. 2007.
- [7] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, Aug. 2018.
- [8] 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.
- [9] K. Punithakumar, P. Boulanger, and M. Noga, "A GPU-accelerated deformable image registration algorithm with applications to right ventricular segmentation," *IEEE Access*, vol. 5, pp. 20374–20382, 2017.
- [10] M. A. Viergever, J. B. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. W. Pluim, "A survey of medical image registration—Under review," *Med. Image Anal.*, vol. 33, pp. 140–144, Oct. 2016.
- [11] B. Ayhan, M. Dao, C. Kwan, H.-M. Chen, J. F. Bell, and R. Kidd, "A novel utilization of image registration techniques to process mastcam images in mars rover with applications to image fusion, pixel clustering, and anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4553–4564, Oct. 2017.
- [12] A. Alba, R. M. Aguilar-Ponce, J. F. Vigueras-Gómez, and E. Arce-Santana, "Phase correlation based image alignment with subpixel accuracy," in *Proc. Adv. Artif. Intell.*, 2013, pp. 171–182.
- [13] Y.-H. Jiang, G. Zhang, X.-M. Tang, D. Li, W.-C. Huang, and H.-B. Pan, "Geometric calibration and accuracy assessment of ZiYuan-3 multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4161–4172, Jul. 2014.
- [14] J. Luce *et al.*, "Medical image registration using the Fourier transform," *Int. J. Med. Phys., Clin. Eng. Radiat. Oncol.*, vol. 3, pp. 49–55, Jan. 2014.
- [15] S. Feng, L. Deng, G. Shu, F. Wang, H. Deng, and K. Ji, "A subpixel registration algorithm for low PSNR images," in *Proc. IEEE 5th Int. Conf. Adv. Comput. Intell. (ICACI)*, Oct. 2012, pp. 626–630.
- [16] J. G. Liu and H. Yan, "Phase correlation pixel-to-pixel image co-registration based on optical flow and median shift propagation," *Int. J. Remote Sens.*, vol. 29, no. 20, pp. 5943–5956, Oct. 2008.
- [17] H. Shekarforoush, M. Berthod, and J. Zerubia, "Subpixel image registration by estimating the polyphase decomposition of cross power spectrum," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1996, pp. 532–537.
- [18] H. S. Stone, M. T. Orchard, E.-C. Chang, and S. A. Martucci, "A fast direct Fourier-based algorithm for subpixel registration of images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 10, pp. 2235–2243, Oct. 2001.
- [19] M. Miura, S. Sakai, S. Aoyama, J. Ishii, K. Ito, and T. Aoki, "High-accuracy image matching using phase-only correlation and its application," in *Proc. SICE Annu. Conf. (SICE)*, 2012, pp. 307–312.
- [20] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 188–200, Mar. 2002.
- [21] X. Tong *et al.*, "A novel subpixel phase correlation method using singular value decomposition and unified random sample consensus," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4143–4156, Aug. 2015.
- [22] I. E. Abdou, "Practical approach to the registration of multiple frames of video images," in *Proc. Vis. Commun. Image Process.*, vol. 3653, Dec. 1998, pp. 371–382.
- [23] V. Argyriou and T. Vlachos, "A study of sub-pixel motion estimation using phase correlation," in *Proc. BMVC*, Jan. 2006, pp. 387–396.
- [24] J. Xie, F. Mo, C. Yang, P. Li, and S. Tian, "A novel sub-pixel matching algorithm based on phase correlation using peak calculation," *ISPRS-Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLI-B1, pp. 253–257, Jun. 2016.
- [25] J. Ren, J. Jiang, and T. Vlachos, "High-accuracy sub-pixel motion estimation from noisy images in Fourier domain," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1379–1384, May 2010.
- [26] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Opt. Lett.*, vol. 33, no. 2, pp. 156–158, Jan. 2008.
- [27] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E86-A, pp. 1925–1934, Aug. 2003.
- [28] Q. Tian and M. N. Huhns, "Algorithms for subpixel registration," *Comput. Vis., Graph., Image Process.*, vol. 35, no. 2, pp. 220–233, Aug. 1986.
- [29] J. G. Liu and H. Yan, "Robust phase correlation methods for sub-pixel feature matching," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 37, pp. 1751–1756, 2006.
- [30] S. Leprinse, S. Barbot, F. Ayoub, and J.-P. Avouac, "Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1529–1558, Jun. 2007.
- [31] X. Geng and W. Yang, "Cyclic shift matrix—A new tool for the translation matching problem," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8904–8913, 2019.
- [32] W. S. Hoge, "A subspace identification extension to the phase correlation method," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 277–280, Feb. 2003.
- [33] C. Ling, X. Geng, W. Yang, and Y. Zhao, "An improved subpixel image registration algorithm based on singular value decomposition," *J. Univ. Chin. Acad. Sci.*, vol. 36, pp. 101–108, Jan. 2019.
- [34] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm, "USAC: A universal framework for random sample consensus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 2022–2038, Aug. 2013.
- [35] H. Wang and D. Suter, "MDPE: A very robust estimator for model fitting and range image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 139–166, Sep. 2004.
- [36] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Signal Process.*, vol. 87, no. 10, pp. 2283–2302, Oct. 2007.
- [37] M. A. Richards. (2013). *The Discrete-Time Fourier Transform and Discrete Fourier Transform of Windowed Stationary White Noise*. [Online]. Available: <http://users.ece.gatech.edu/mrichard/DFT%20of%20Noise.pdf>
- [38] J. W. Lindeberg, "Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung," *Mathematische Zeitschrift*, vol. 15, no. 1, pp. 211–225, Dec. 1922.



Liangliang Zhu received the B.S. degree from the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, in 2018. He is pursuing the Ph.D. degree with the Key Laboratory of Technology in Geo-spatial Information Process and Application Systems, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing.

His research interests include hyperspectral target detection, image registration, and machine learning.



Xiurui Geng received the B.S. and M.S. degrees in applied mathematics from Beihang University, Beijing, China, in 1999 and 2002, respectively, and the Ph.D. degree in hyperspectral remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, in 2005.

He is a Professor with the Key Laboratory of Technology in Geo-spatial Information Process and Application Systems, Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include hyperspectral image processing, machine learning, and the basic theory of matrices.