Deep Linear Array Pushbroom Image Restoration: A Degradation Pipeline and Jitter-Aware Restoration Network

Zida Chen[1*], Ziran Zhang[1,2*], Haoying Li[1], Menghao Li[1], Yueting Chen[1], Qi Li[1], Huajun Feng[1], Zhihai Xu[1], Shiqi Chen[1†]

[1]State Key Laboratory of Extreme Photonics and Instrumentation, Zhejiang University
[2]Shanghai Artificial Intelligence Laboratory
{zd_chen, naturezhanghn, lhaoying, limh, chenyt, liqi, fenghj, xuzh, chenshiqi}@zju.edu.cn

Abstract

Linear Array Pushbroom (LAP) imaging technology is widely used in the realm of remote sensing. However, images acquired through LAP always suffer from distortion and blur because of camera jitter. Traditional methods for restoring LAP images, such as algorithms estimating the point spread function (PSF), exhibit limited performance. To tackle this issue, we propose a Jitter-Aware Restoration Network (JARNet), to remove the distortion and blur in two stages. In the first stage, we formulate an Optical Flow Correction (OFC) block to refine the optical flow of the degraded LAP images, resulting in pre-corrected images where most of the distortions are alleviated. In the second stage, for further enhancement of the pre-corrected images, we integrate two jitter-aware techniques within the Spatial and Frequency Residual (SFRes) block: 1) introducing Coordinate Attention (CoA) to the SFRes block in order to capture the jitter state in orthogonal direction; 2) manipulating image features in both spatial and frequency domains to leverage local and global priors. Additionally, we develop a data synthesis pipeline, which applies Continue Dynamic Shooting Model (CDSM) to simulate realistic degradation in LAP images. Both the proposed JARNet and LAP image synthesis pipeline establish a foundation for addressing this intricate challenge. Extensive experiments demonstrate that the proposed two-stage method outperforms state-of-the-art image restoration models. Code is available at https://github.com/JHW2000/JARNet.

Introduction

Linear array cameras are widely employed in remote sensing for high-resolution optical imaging on the earth (Cui et al. 2023; Wang, Zhu, and Fan 2018). As shown in Fig. 1, a linear array camera contains an array of sensors arranged in a straight line. The linear array sensor captures images at ground scene while the whole camera moves along the pushbroom motion direction. This movement is similar to how a broom sweeps forward. Pixels imaged at different moments are stitched together to generate a LAP image. Due to adjustments in camera attitude and periodic movement (Iwasaki 2011), inevitable jitter arises in LAP imaging. This leads to pixel displacements of varying magnitudes, thereby inducing distortion and blur within LAP images. Low-frequency jitter causes image distortion effect, while high-frequency jitter leads to blur effect (Pan et al. 2020). In practice, roll jitter causes pixel offset in the cross-track direction, and pitch jitter causes pixel offset in the along-track direction. Jitter offsets in the along-track direction are much smaller than that in the cross-track direction, while jitter in the yaw direction is tiny enough to be neglected (Wang et al. 2017).

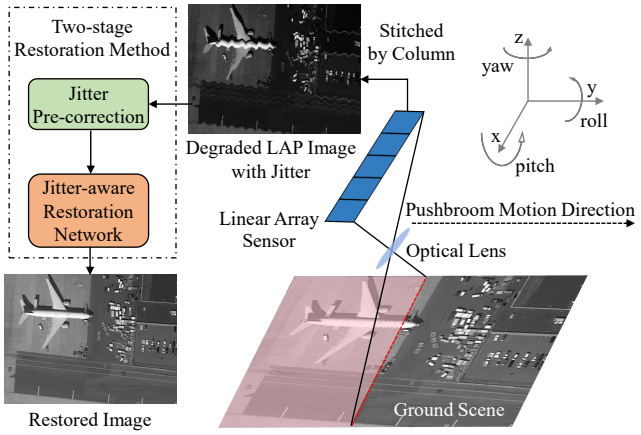

Figure 1: Illustration of our main idea: The stitched LAP image suffers from distortion and blur caused by camera jitter. To restore degraded LAP image, we propose a jitter pre-correction and enhancement method in a two-stage manner. We also design an image synthesis pipeline for training data acquisition. Finally, our proposed JARNet outperforms state-of-the-art methods on our LAP dataset.

Many efforts have been made to remove distortion and blur in LAP images. Their efforts are directed towards acquiring more precise jitter curves, either directly measuring from high-resolution equipment (Pan et al. 2020) or indirectly predicting from degraded images (Wang et al. 2021). However, the former is often hindered by the scarcity of high-precision equipment, while the latter often falls short in terms of predicting accuracy. Over the years, deep learning has demonstrated outstanding performance in many image restoration tasks due to its powerful data modeling and gen-

*These authors contributed equally.
†Corresponding author

eration capabilities. However, mainstream image restoration methods (Cho et al. 2021; Chen et al. 2021; Zamir et al. 2022; Wang et al. 2022) are not well-adopted in LAP image reconstruction due to the lack of jitter priors. As shown in Fig. 6, we test seven mainstream methods and they struggle to restore sufficient details of farmland and city buildings, showing their limitation in LAP image restoration task.

Another obstacle is the availability of data that comes with distortion and blur by jitter. As far as we are concerned, there is no customized LAP image dataset. Many studies (Zhang, Iwasaki, and Xu 2019; Wang et al. 2021) use several sinusoidal components to simulate jitter effect in LAP image (Wulich and Kopeika 1987). However, jitter in the real world is more complicated. So how to establish an efficient data synthesis pipeline remains a problem.

In this paper, we tackle the challenge of limited data availability for LAP images and propose a **novel restoration pipeline** for LAP image recovery. Inspired by Continue Dynamic Shooting Model (CDSM) which provides a finer-grained sampling strategy (Pan et al. 2020), we propose a **novel LAP image degradation pipeline** on CDSM-based jitter model, which simulates distortion and blur. We generate sufficient degraded LAP images from public dataset DOTA-v1.0 (Xia et al. 2018) for building our LAP dataset.

In order to utilize jitter prior and achieve better restoration performance, we propose JARNet, a **jitter-aware restoration network** based on the two-stage restoration strategy. In the first stage, we design an Optical Flow Correction (OFC) block to refine optical flow from jitter prior. Then we warp the degraded LAP image by the refined optical flow to make precise pre-correction, which removes most of the distortion. In the second stage, we design a Spatial and Frequency Residual (SFRes) block, which integrates two jitter-aware techniques of coordinate attention (Hou, Zhou, and Feng 2021) (CoA) block and frequency branch (Mao et al. 2023). The CoA block in spatial branch captures jitter state in orthogonal direction. The frequency branch parallel to the spatial branch extracts both low and high-frequency jitter, guiding the overall removal of distortion and blur. Compared with state-of-the-art methods, our proposed JARNet achieves competitive performance in LAP image restoration task. In summary, our contributions are outlined as follows:

- We develop a novel LAP image synthesis pipeline with CDSM integration, which achieves a finer-grained level of simulation fidelity and boosts restoration performance.
- We propose the first jitter-aware restoration network, employing optical flow correction and two jitter-aware techniques to utilize both spatial and frequency information.
- Extensive experiments show that our method achieves superior results (+ 1.28dB in PSNR) compared with state-of-the-art methods on our LAP dataset.

Related Work

LAP Image Restoration for Remote Sensing

Many efforts have been made to remove distortion and blur effects in LAP images. For distortion caused by low-frequency jitter, many researchers improve the accuracy of jitter detection for image resampling. While high-precision attitude sensors (Tang et al. 2014) directly measure the attitude of linear array cameras, the feasibility is constrained by its economic viability. Some works focus on leveraging parallax imaging systems across distinct spectral bands within multispectral images for indirect prediction of jitter state (Hu, Zhang, and Liu 2018). Nevertheless, such strategies are often rendered inapplicable within the context of panchromatic LAP images. Furthermore, solutions based on deep learning have also been explored to predict jitter curves (Zhang, Iwasaki, and Xu 2019). However, these approaches only work within a limited range of jitter amplitude.

For blur effects caused by high-frequency jitter, traditional techniques are divided into blind and non-blind restoration methods. The former targets scenarios where blur kernel remains unknown, while the latter undertakes algorithms based on the point spread function (PSF) (Vimal 2019; Chen et al. 2013). Most studies usually apply the same PSF to the entire image, because jitter is regarded as uniform within a short imaging interval (Pan et al. 2020). However, this assumption is invalid for high-frequency jitter.

As the real LAP image data is rare, researchers often (Zhang, Iwasaki, and Xu 2019; Wang et al. 2021) simulate jitter effects by using multiple sinusoidal components (Wulich and Kopeika 1987). However, jitter in the real world is more intricate. Consequently, establishing an effective data synthesis pipeline remains a challenge. In this paper, we establish a CDSM-based LAP image degradation pipeline to acquire sufficient LAP data. We utilize deep learning technology and propose JARNet, to process distortion and blur in a two-stage restoration manner. We remove most of the distortions with jitter prior processed by CDSM in the first stage. Then we deal with the rest of distortion and blur effect by restoration network in the second stage.

Learning-based Single Image Restoration

Deep learning has emerged as a powerful tool for learning data-driven models end-to-end, especially in low-level vision tasks (e.g. image restoration). Many methods have achieved remarkable performance on public degraded datasets (Nah, Hyun Kim, and Mu Lee 2017; Abdelhamed, Lin, and Brown 2018; Li et al. 2023), proving their potential in natural image restoration. These methods often employ multi-scale architectures. For example, MIMO-UNet (Cho et al. 2021) utilizes multi-scale inputs and outputs, HINet (Chen et al. 2021) applies a two-stage UNet, and NAFNet (Chen et al. 2022) leverages a simplified UNet backbone. Some transformer-based approaches utilize self-attention mechanism and reduce time complexity of vanilla vision transformer (Dosovitskiy et al. 2021), such as Uformer (Wang et al. 2022), Restormer (Zamir et al. 2022) and Stripformer (Tsai et al. 2022). However, such mainstream image restoration methods do not work well in LAP image restoration task due to the domain gap between LAP and natural imaging scenes. With the refined optical flow from OFC block and two jitter-aware techniques of CoA block and frequency branch, our JARNet can capture jitter state in orthogonal direction and extract jitter at different frequencies, which is well-adopt in LAP image restoration.

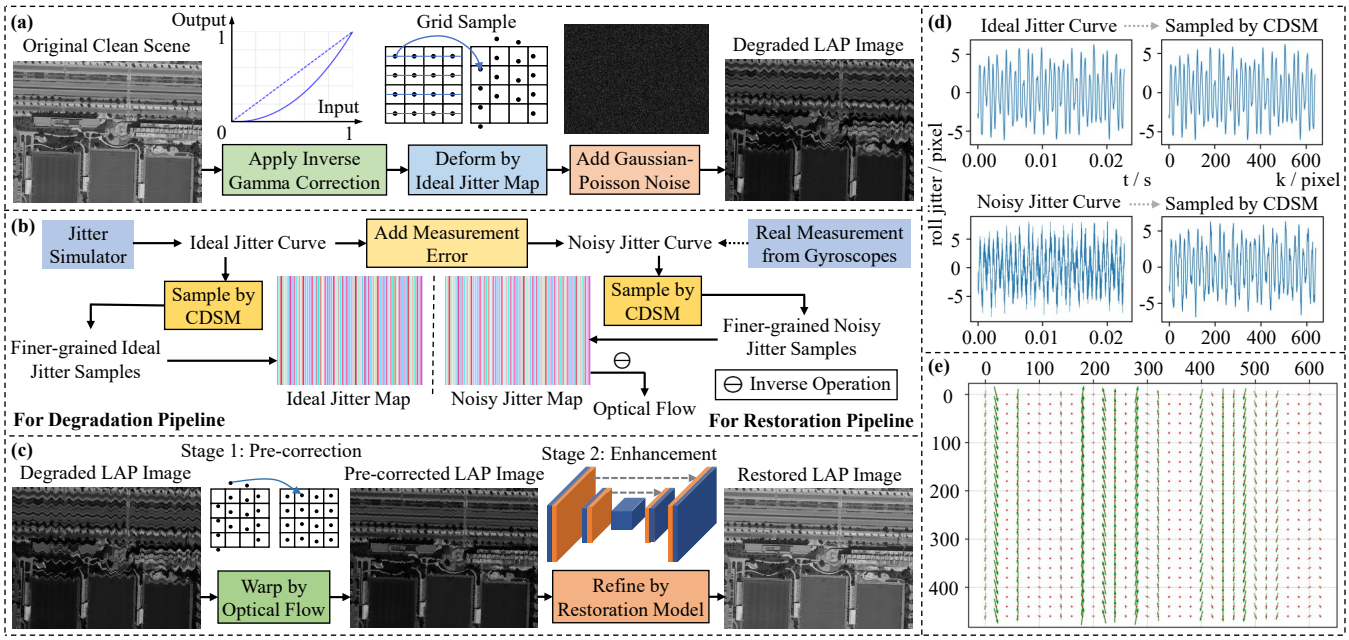

Figure 2: Overall degradation and restoration pipeline for LAP images. (a) Proposed LAP image degradation pipeline. (b) Proposed CDSM-based jitter map generating procedure. (c) Proposed two-stage restoration pipeline for LAP images. (d) Visualization of jitter curve in (b). From upper left to lower right: ideal jitter curve, the average of finer-grained ideal jitter samples from CDSM, noisy jitter curve, and the average of finer-grained noisy jitter samples from CDSM. For simplicity, we only display jitter curve in the roll direction. (e) Visualization of optical flow from noisy jitter map.

Proposed Method

We first introduce the principle of LAP jitter with the CDSM-based dense sampling strategy. On this basis, we propose an image synthesis pipeline for LAP image degradation. Finally, we present the jitter-aware image restoration network, JARNet, in a two-stage restoration manner.

CDSM-Based Jitter Model

For linear array sensors in the remote sensing field, the main factor of degradation is the image deformation and blur effect caused by the camera jitter. The jitter can be described by a series of time-varying sinusoidal functions (Wulich and Kopeika 1987) in Eq. 1:

$$\phi_d(t) = \sum_{i}^{N} A_i \sin\left(2\pi f_i t + \varphi_i\right), \tag{1}$$

where ϕ_d is a time-varying jitter angle curve in a certain direction, A, f and φ are jitter amplitude, jitter frequency, and initial random phase, respectively. N means the number of sinusoidal components. Due to tiny ϕ_d, the pixel offset on the sensor caused by jitter can be approximated by $J_d = \phi_d \frac{f}{\mu}$, where J_d is the number of pixel offset in a certain direction, f is focal length of optical system, and u is pixel size of linear array sensor. Proportional to ϕ_d, J_d consists of various sinusoidal components. As shown in Fig. 2 (b), (d), we obtain ideal jitter curve J_d for LAP image degradation pipeline by jitter simulator in Eq. 1. The real jitter curve measured by gyroscopes is noisy that contains much measurement error. We denote it as noisy jitter curve. However, it is difficult to obtain sufficient noisy jitter curves in the real world. Therefore, we simulate noisy jitter curve for LAP two-stage restoration pipeline by adding measurement error to the ideal jitter curve.

It is worth noting that ideal jitter curve from jitter simulator simulates image distortion with only tiny blur effect. And correcting deformed LAP image directly by noisy jitter curve is imprecise. So we introduce CDSM into our jitter model to address these issues, which will be discussed in Section and Section . CDSM provides a denser sampling strategy (Pan et al. 2020). Without CDSM, we sample jitter curve at the time interval of imaging τ. In other words, in Eq. 1, $t \subseteq k\tau$, where k is column pixel index in LAP image. With the application of CDSM, we utilize a shorter time interval for finer-grained sampling. Derived from Eq. 1, The CDSM-based jitter model is depicted by Eq. 2:

$$J_d^{sub}(t,m) = \sum_{i}^{N} A_i' \sin\left(2\pi f_i (t + \frac{m}{M}\tau) + \varphi_i\right), \tag{2}$$

where J_d^{sub} is the subdivision jitter curve, $A_i^{'} = A_i \frac{f}{\mu}$, M is subdivision number, m is subdivision index, τ is time interval of imaging. With CDSM, the sample time interval becomes $\frac{\tau}{M}$, achieving finer-grained sampling. In the upper right and lower right of Fig. 2 (d), we visualize the effect of CDSM by averaging finer-grained jitter curves in Eq. 3:

$$J_d^{CDSM}(k) = \frac{1}{M}\sum_{m}^{M} J_d^{sub}(k\tau, m), \tag{3}$$

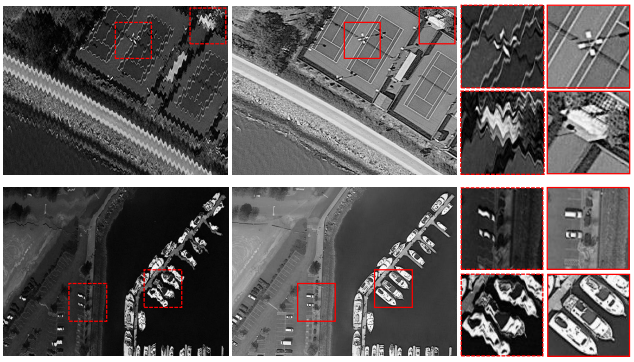

Figure 3: Examples of our LAP image dataset. The 1^{st} and 3^{rd} columns are simulated LAP images from our proposed degradation pipeline. The 2^{nd} and 4^{th} columns are the corresponding original clean scene.

where J_d^{CDSM} is jitter curve processed by CDSM. Compared to their original state, the ideal jitter curve shows only minor changes, while the noisy jitter curve is smoothed.

LAP Image Degradation Pipeline

Fig. 2 (a) illustrates the pipeline of generating degraded LAP images from original clean scene. Firstly, we apply inverse gamma correction for original clean image to obtain the energy representation of the scene, as the sensor noise follows Gaussian-Poisson distribution in this domain. Next, we deform the image by ideal jitter map. Finally, Gaussian-Poisson noise is added to the deformed image.

Ideal jitter map determines the amount of jitter offset for each pixel in LAP image. Given a LAP image with shape of $(H, W, 1)$, where the width direction is the push-broom direction, we establish a sampling time array $t \subseteq \{\tau, 2\tau, ..., W\tau\}$. According to Eq. 2, we calculate a finer-grained ideal jitter sample in a certain direction from CDSM with a shape of $(W, 1)$, denoted as $\widetilde{J}_d^{sub}$. The height direction represents the direction of the linear array sensor. Since all the pixels are imaged at the same time, their jitter states remain consistent. So we simply duplicate the jitter sample along the height direction to a shape of $(H, W, 1)$. Subsequently, we concatenate jitter from roll and pitch direction, resulting in an ideal jitter map with a shape of $(H, W, 2)$. Eq. 4 shows the process above.

$$J^{sub}(m) = \boldsymbol{C}\big(\boldsymbol{D}(\widetilde{J}^{sub}_{roll}(m), h), \boldsymbol{D}(\widetilde{J}^{sub}_{pitch}(m), h)\big), \quad (4)$$

where J^{sub} is the m^{th} subdivided ideal jitter map. $\boldsymbol{D}$ is duplicate operation. h represents the height direction. $\boldsymbol{C}$ is concatenate operation. Finally, we make a grid sample to deform LAP image by several subdivision ideal jitter maps. Since the pixel offset may not be an integer, we apply bilinear interpolation during resampling. The LAP image degradation pipeline is depicted by Eq. 5:

$$I^{lq} = \frac{1}{M}\sum_{m}^{M} \boldsymbol{G}\big((I^{gt})^{\gamma}, J^{sub}(m)\big) + n, \quad (5)$$

where I^{lq} is the degraded LAP image, I^{gt} is the original image, γ is inverse gamma correction coefficient, $\boldsymbol{G}$ denotes grid sample operation, and n is Gaussian-Poisson noise. Through averaging deformed LAP images from ideal jitter maps with different subdivision indexes, we successfully introduce blur effect into degraded LAP image thanks to CDSM-based jitter model.

Finally, we establish our LAP dataset through our proposed degradation pipeline, which contains degraded-clean LAP image pairs. We demonstrate two LAP image pairs and their zoom-in details in Fig. 3.

Jitter-Aware Restoration Network

To restore the LAP images from the pixel displacement and blur caused by camera jitter, we propose a jitter-aware restoration network, JARNet, which restores LAP images in two stages, as shown in Fig. 2 (c). The first stage is called the pre-correction stage, where most of the distortion in degraded LAP image is removed by optical flow. Noticeably, we introduce Optical Flow Correction (OFC) block in JARNet to enable more precise warping and effective distortion removal. The second stage is called the enhancement stage, where we employ a Spatial and Frequency Residual (SFRes) block in the U-shaped network to further enhance the pre-corrected LAP image. The two-stage restoration strategy offers us an effective approach to enhance the performance of state-of-the-art methods in LAP image restoration task. We will introduce the two stages, the proposed jitter-aware techniques, and losses in the following paragraphs.

Pre-correction Stage and OFC Block Since jitter state provides degradation information, we attempt to use it as prior to warp the distorted LAP images in the pre-correction stage. Specifically, we simulate several subdivided noisy jitter maps from noisy jitter curve as shown in Fig. 2 (b), whose process is similar to ideal jitter map. Then we average all the subdivided noisy jitter maps, assisting in smoothing the noise thanks to CDSM-based jitter model. Subsequently, the degraded LAP image is warped by optical flow in Eq. 6, which is approximated as the inverse of noisy jitter map:

$$\begin{aligned} I^{warp} &= \boldsymbol{G}(I^{lq}, \omega) \approx \boldsymbol{G}(I^{lq}, -J_{noisy}) \\ &= \boldsymbol{G}\big(I^{lq}, -\frac{1}{M}\sum_{m}^{M} J^{sub}_{noisy}(m)\big), \end{aligned} \quad (6)$$

where I^{warp} is the pre-corrected image, most of whose distortions are removed. ω denotes optical flow, and J_{noisy} denotes noisy jitter map. Fig. 2 (e) shows the optical flow map with shape of $(H, W, 2)$. The vector at each pixel represents direction and relative magnitude of pixel offset.

Instead of warping the degraded image by original optical flow, we refine the optical flow by our proposed Optical Flow Correction (OFC) block. As shown in Fig. 5, the OFC block has a shallow attention-based convolution architecture and employs the SCA module (Chen et al. 2022) to reweight feature map of optical flow between roll and pitch directions, which assists better embedding of jitter prior. Then we utilize the refined optical flow to warp the degraded image.

Enhancement Stage and SFRes Block Since optical flow is calculated approximately, there still remains little

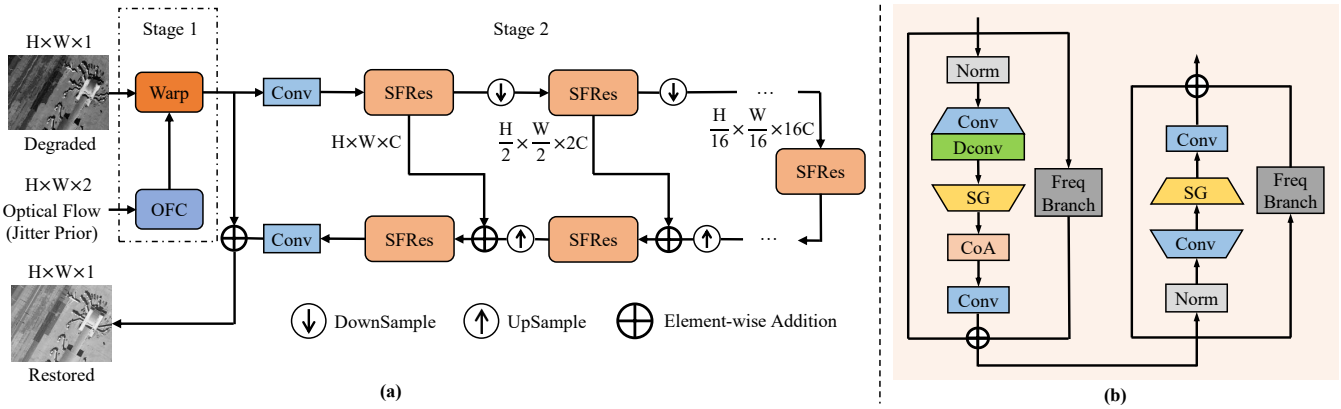

Figure 4: Architecture of JARNet for LAP image restoration. (a) Overview of our two-stage JARNet. (b) Details of SFRes block. 'Norm' means layer normalization. 'Dconv' means depthwise convolution. 'SG' means SimpleGate (Chen et al. 2022). 'CoA' means coordinate attention (Hou, Zhou, and Feng 2021). 'Freq Branch' means frequency branch (Mao et al. 2023).

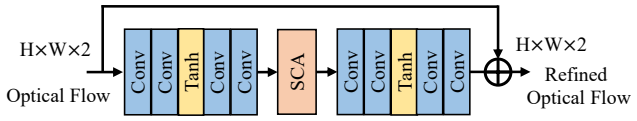

Figure 5: Details of our proposed Optical Flow Correction (OFC) block. 'Tanh' means hyperbolic tangent activation function. 'SCA' means Simplified Channel Attention (Chen et al. 2022).

distortion and blur in pre-corrected LAP image. In the enhancement stage, we further improve the LAP image quality by a U-shaped network with Spatial and Frequency Residual (SFRes) blocks as key blocks, as illustrated in Eq. 7:

$$I^R = \boldsymbol{N}(I^{warp}; \Theta), \tag{7}$$

where I^R is restored image. $\boldsymbol{N}$ is restoration network. Θ is parameter set in image restoration network. As shown in Fig. 4, the enhancement stage adopts a U-shaped network with skip connections between the encoder and the decoder. After the initial convolution layer, each level of encoder and decoder uses a series of SFRes blocks, inspired by the design of NAFNet block (Chen et al. 2022). The SFRes block includes a spatial branch, a frequency branch, and a residual path, taking advantage of both local and global information. In both of the two kinds of branches, we adopt jitter-aware techniques to further improve the LAP image quality.

The spatial branch of SFRes block employs coordinate attention (CoA) (Hou, Zhou, and Feng 2021) block to reweight feature map in the vertical and horizontal directions of LAP image. With the input feature shape of (H, W, C), CoA outputs two attention maps with shapes of $(H, 1, C)$ and $(1, W, C)$, respectively. Then element-wise multiplication is applied between the input feature map and attention maps. CoA block helps capture the jitter state in orthogonal directions for better LAP image restoration.

The frequency branch (Mao et al. 2023) in the SFRes applies fast Fourier transform algorithm to convert the feature map to the frequency domain. Each pixel of feature map in frequency branch contains global information of LAP image, which not only extracts both low and high-frequency jitter but also assists in restoring high-frequency details.

Losses To supervise the training of JARNet, we apply a restoration loss $\mathcal{L}_{res}$ and an optical flow loss $\mathcal{L}_{flow}$ in the total loss function $\mathcal{L}_{total}$, as illustrated in Eq. 8:

$$\mathcal{L}_{total} = \mathcal{L}_{res} + \lambda_1 \cdot \mathcal{L}_{flow}, \tag{8}$$

where $\lambda_1 = 0.1$. The refined optical flow ω^R is supervised by noise-free optical flow ω^{gt}, where $\omega^{gt} \approx -\frac{1}{M}\sum_m^M J^{sub}(m)$. We define the flow loss in Eq. 9:

$$\mathcal{L}_{flow} = \mathcal{L}_1(\omega^R, \omega^{gt}), \tag{9}$$

where $\mathcal{L}_1$ is mean absolute error loss function. $\mathcal{L}_{flow}$ is utilized to constrain the optimization of the optical flow. We train OFC block as part of JARNet in a joint end-to-end manner. We apply a restoration loss to supervise the training of JARNet between I^R and I^{gt} in Eq. 10:

$$\mathcal{L}_{res} = \mathcal{L}_1 + \lambda_2 \cdot \mathcal{L}_{percep} + \lambda_3 \cdot \mathcal{L}_{fft}, \tag{10}$$

where $\lambda_2 = 10^{-4}$ and $\lambda_3 = 0.1$. $\mathcal{L}_{res}$ contains 3 components, mean absolute error loss, perceptual loss (Johnson, Alahi, and Li 2016), and FFT loss (Cho et al. 2021).

Experiments

Dataset

We utilize DOTA-v1.0 dataset (Xia et al. 2018) as our source of clean data, which is originally designed for object detection in aerial images. We apply our proposed LAP image degradation pipeline on training set of DOTA-v1.0, which contains 2,806 images, to create our simulated LAP dataset. We crop each image into a size of 640×480. For degradation parameter, $f \in \{1000, 2000, 3000, 4000\}$ in Hz, $A_{roll} \in \{4, 1.5, 1.0, 0.5\}$ in pixel number, $A_{pitch} \in \{1, 0.5, 0.3, 0.2\}$ (Teshima and Iwasaki 2007; Zhu et al. 2018), $\tau = 3.54 \times 10^{-5}s$, $M = 6$. We simulate four

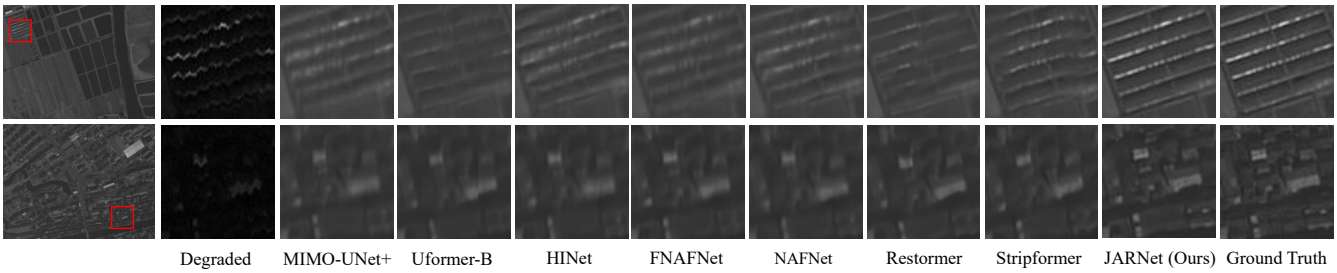

Figure 6: Visual comparison of different restoration methods on our LAP dataset.

main sinusoidal components. To promise a larger degradation space, for each cropped image we multiply amplitude and frequency respectively by a vibration factor, where amplitude vibration factor satisfies Gaussian distribution with $\mu = 1$, $\sigma = 0.1$ and frequency vibration factor satisfies Gaussian distribution with $\mu = 1$, $\sigma = 0.01$. The Gaussian-Poisson noise is applied with $\sigma_{gauss} = 0.01$, $\lambda_{poisson} = 10^{-4}$. The maximum jitter measurement error is 20% of current jitter offset. Overall, we obtain 20,614 train image pairs and 2,577 test image pairs in our synthetic LAP dataset.

Implementation Details

Our experiments are all trained and evaluated on our LAP image dataset. We set our batch size as 4 and training patch size as 128×128. We train our JARNet from scratch for 450k steps on a single NVIDIA GeForce RTX 4090 GPU with 24GB of memory, which takes approximately 26 hours. We apply AdamW (Loshchilov and Hutter 2019) optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 10^{-3}). Cosine learning rate strategy is applied from 3×10^{-4} to 10^{-7}. We compare our proposed method with other state-of-the-art approaches on our LAP image dataset. For a fair comparison, other methods are all trained from scratch on our LAP dataset. Except for batch size, patch size, and training steps, we follow the training protocols of each method if not specified. In testing, we use PSNR, SSIM, and gradient magnitude similarity deviation (GMSD) (Xue et al. 2013) as our evaluation metrics with the full image size of 640×480.

Comparisons with State-of-the-Art Methods

We compare the proposed method on our LAP dataset with the following RGB-based methods: MIMO-UNet+ (Cho et al. 2021), HINet (Chen et al. 2021), Uformer-B (Wang et al. 2022), NAFNet (Chen et al. 2022), Restormer (Zamir et al. 2022), Stripformer (Tsai et al. 2022), and FNAFNet (Mao et al. 2023). Some spectral-based methods, such as MST (Cai et al. 2022b), MST++ (Cai et al. 2022c), CST (Cai et al. 2022a), DAUHST (Cai et al. 2022d), and HDNet (Hu et al. 2022), are also evaluated. Quantitative evaluation results in Tab. 1 demonstrate that our proposed JARNet outperforms all other state-of-the-art methods in terms of PSNR, SSIM, and GMSD. Compared to the existing best single image restoration method, JARNet achieves 39.02dB in PSNR and 0.9493 in SSIM, which corresponds to 1.28dB improvement in PSNR, 0.0104 improvements in SSIM and 0.0087 improvements in GMSD. Fig. 6 presents visual comparison of an interior details of farmland and city buildings. Our proposed JARNet not only effectively removes jitter in LAP images but also successfully recovers more details in the contour of farmland and buildings. While other existing methods can effectively remove deformation caused by jitter, most of them fail to remove blur effect and recover sufficient details. In contrast, JARNet utilizes the prior knowledge of jitter to make pre-correction in the first stage and reduces the learning difficulty for subsequent restoration network so as to pay more attention to image details.

Models	↑PSNR(dB)	↑SSIM	↓GMSD	↓Params(M)
MIMO-UNet+	34.94	0.9127	0.0447	16.10
Uformer-B	36.08	0.9213	0.0439	50.88
HINet	36.18	0.9208	0.0376	88.66
FNAFNet	36.94	0.9311	0.0370	68.02
NAFNet	36.95	0.9291	0.0366	67.89
Restormer	37.51	0.9366	0.0344	26.12
Stripformer	<u>37.74</u>	<u>0.9389</u>	<u>0.0326</u>	19.71
MST-L	35.67	0.9155	0.0464	2.03
MST++	35.65	0.9154	0.0456	1.33
CST-L*	35.16	0.9083	0.0499	3.00
DAUHST-9stg	36.08	0.9214	0.0435	6.15
HDNet	34.43	0.8958	0.0595	2.37
JARNet(Ours)	**39.02**	**0.9493**	**0.0239**	13.46

Table 1: Quantitative results of different RGB-based and spectral-based restoration methods and JARNet. 'Params' denotes the number of parameters.

Ablation Studies

Effectiveness of Pre-Correction

To validate the effectiveness of pre-correction, we utilize jitter prior to enhance other methods in a two-stage restoration manner as shown in Fig. 2 (c). Tab. 2 shows that performances of all methods enhanced by jitter prior are improved, compared with Tab. 1. Notably, Stripformer* achieves a slightly higher SSIM compared with our method, while JARNet maintains the best PSNR, GMSD metric, and relatively acceptable amount of GMACs among all methods. The pre-correction in the first stage effectively reduces learning difficulty of restoration network in the second stage, leading to enhanced performance for mainstream methods.

Models	↑PSNR(dB)	↑SSIM	↓GMSD	↓GMACs
MIMO-UNet+*	36.07	0.9402	0.0293	154.3
Uformer-B*	37.93	0.9431	0.0286	85.75
HINet*	37.97	0.9425	0.0275	170.3
FNAFNet*	37.70	0.9404	0.0296	63.34
NAFNet*	38.01	0.9428	0.0279	63.18
Restormer*	38.46	0.9472	0.0263	140.8
Stripformer*	<u>38.72</u>	**0.9500**	<u>0.0247</u>	170.4
JARNet(Ours)	**39.02**	<u>0.9493</u>	**0.0239**	85.71

Table 2: Quantitative results of different enhanced RGB-based restoration methods. All methods denoted '$*$' are enhanced by pre-correction in the first stage. We evaluate GMACs with input tensor shape of $(1, 1, 256, 256)$.

No.	Freq	CoA	OFC	↑PSNR	↑SSIM	↓Params(M)
1				38.17	0.9444	13.80
2	✓			37.98	0.9428	13.94
3		✓		38.50	0.9478	12.26
4			✓	38.56	0.9463	14.87
5	✓	✓		<u>38.69</u>	<u>0.9492</u>	12.39
6	✓		✓	38.44	0.9456	15.01
7		✓	✓	<u>38.69</u>	0.9471	13.33
8(Ours)	✓	✓	✓	**39.02**	**0.9493**	13.46

Table 3: Ablation of frequency branch, CoA block, and OFC block in JARNet. 'Freq' denotes frequency branch.

Effectiveness of Components in JARNet

In this section, we verify the effectiveness of frequency branch, CoA block, and OFC block in JARNet by conducting 7 extra experiments. As shown in Tab. 3, when we add CoA block alone in No.3, we observe a significant performance gain of 0.33dB in PSNR compared with No.1 baseline. This improvement indicates that the CoA block is effective in extracting orthogonal jitter state. Additionally, the number of parameters reduces by 11.2% approximately, indicating the efficiency of incorporating the CoA block in our JARNet. When we add frequency branch in No.2, the performance has a decline of 0.19dB in PSNR, compared with No.1 baseline. Nevertheless, by incorporating the global information extracted by frequency branch and jitter state extracted by CoA block, No.5 achieves better performance compared with No.3. So we apply frequency branch and CoA block simultaneously. Similarly, when we add OFC block in No.4, we achieve a performance gain of 0.39dB in PSNR, compared with No.1 baseline. This improvement indicates that refined optical flow effectively pre-corrects degraded LAP image. For the best restoration performance, we apply all three components in No.8 JARNet.

We further analyze visual results of different combinations of components. As shown in Fig. 7, when the frequency branch and CoA block are combined, the vertical details in the window are partly restored, which indicates that the CoA block, along with the frequency domain branch, has capabilities to capture jitter state and high-frequency features. OFC block helps reconstruct zebra crossing correctly on the edge of the degraded image. In situations where pre-corrected images have black edges, architectures without OFC block fail to handle the edge details correctly. However, the application of the OFC block helps alleviate the influence of these black edges, resulting in better restoration of edge details. Our JARNet combines advantages of all three components and achieves high-quality restoration results.

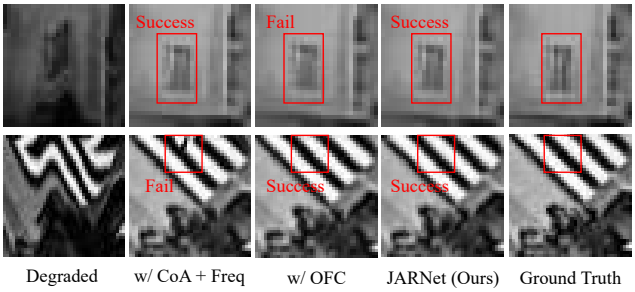

Figure 7: Visual comparison of different components.

	NAFNet*	Restormer*	Stripformer*	JARNet(Ours)
PSNR	36.64	37.06	37.52	38.50
SSIM	0.9270	0.9338	0.9403	0.9465

Table 4: Results on LAP dataset without CDSM.

Effectiveness of CDSM

In order to verify CDSM enhances the restoration effect of the LAP dataset. We remove CDSM and generate a new CDSM-free LAP dataset on which we compare different enhanced state-of-the-art methods. As shown in Tab. 4, other enhanced methods suffer from a sharp performance decline compared with results in Tab. 2, even worse than non-enhanced versions in Tab. 1. Without CDSM, the noisy jitter curve is not smoothed. Therefore, pre-correction in the first stage does not work. However, JARNet only shows a slight decrease in performance, because the OFC block in the first stage compensates part of smoothing effect from CDSM. This fully demonstrates the importance of CDSM in LAP image restoration pipeline. CDSM helps smooth the noisy jitter curve and obtain better pre-correction results.

Conclusion

In order to restore degraded LAP image caused by camera jitter and overcome the obstacle of data availability, we proposed a CDSM-based LAP image degradation pipeline and created a LAP dataset. Then we presented the first jitter-aware restoration network in a two-stage restoration manner. In the first stage, we utilized optical flow refined by OFC block to warp the degraded LAP image. In the second stage, we incorporated CoA block and frequency branch in SFRes block to realize jitter-aware character. CoA block captures jitter state in orthogonal direction, while frequency branch extracts both low and high-frequency jitter. Extensive experiments demonstrate that our approach performs favorably against state-of-the-art methods qualitatively and quantitatively on our LAP dataset.

Acknowledgments

This project is supported by National Natural Science Foundation of China (No. 62275229). We thank Meijuan Bian and Weige Lyu from the facility platform of optical engineering of Zhejiang University for instrument support.

References

Abdelhamed, A.; Lin, S.; and Brown, M. S. 2018. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1692–1700.

Cai, Y.; Lin, J.; Hu, X.; Wang, H.; Yuan, X.; Zhang, Y.; Timofte, R.; and Van Gool, L. 2022a. Coarse-to-Fine Sparse Transformer for Hyperspectral Image Reconstruction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 686–704.

Cai, Y.; Lin, J.; Hu, X.; Wang, H.; Yuan, X.; Zhang, Y.; Timofte, R.; and Van Gool, L. 2022b. Mask-Guided Spectral-Wise Transformer for Efficient Hyperspectral Image Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17502–17511.

Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; Pfister, H.; Timofte, R.; and Van Gool, L. 2022c. MST++: Multi-Stage Spectral-Wise Transformer for Efficient Spectral Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 745–755.

Cai, Y.; Lin, J.; Wang, H.; Yuan, X.; Ding, H.; Zhang, Y.; Timofte, R.; and Gool, L. V. 2022d. Degradation-Aware Unfolding Half-Shuffle Transformer for Spectral Compressive Imaging. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, 37749–37761.

Chen, H.; Cao, M.; Wang, H.; Yan, Y.; and Ma, L. 2013. Estimating the point spread function of motion-blurred images of the Ochotona Curzoniae. In *2013 6th International Congress on Image and Signal Processing (CISP)*, volume 1, 369–373. IEEE.

Chen, L.; Chu, X.; Zhang, X.; and Sun, J. 2022. Simple baselines for image restoration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 17–33.

Chen, L.; Lu, X.; Zhang, J.; Chu, X.; and Chen, C. 2021. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 182–192.

Cho, S.-J.; Ji, S.-W.; Hong, J.-P.; Jung, S.-W.; and Ko, S.-J. 2021. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4641–4650.

Cui, Y.; Liu, C.; Liu, S.; Xu, M.; and Xie, P. 2023. Optical design and precision analysis of a single-line-array pendulum sweep high-resolution mapping camera. *Applied Optics*, 62(5): 1183–1192.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations (ICLR)*.

Hou, Q.; Zhou, D.; and Feng, J. 2021. Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13713–13722.

Hu, K.; Zhang, Y.; and Liu, W. 2018. High-frequency jitter detection by registration error curve of high-resolution multi-spectral satellite image. In *2018 26th International Conference on Geoinformatics*, 1–6. IEEE.

Hu, X.; Cai, Y.; Lin, J.; Wang, H.; Yuan, X.; Zhang, Y.; Timofte, R.; and Van Gool, L. 2022. HDNet: High-Resolution Dual-Domain Learning for Spectral Compressive Imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17542–17551.

Iwasaki, A. 2011. Detection and estimation satellite attitude jitter using remote sensing imagery. *Advances in Spacecraft Technologies*, 13: 257–272.

Johnson, J.; Alahi, A.; and Li, F.-F. 2016. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 694–711.

Li, H.; Zhang, Z.; Jiang, T.; Luo, P.; Feng, H.; and Xu, Z. 2023. Real-world deep local motion deblurring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 1314–1322.

Loshchilov, I.; and Hutter, F. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations (ICLR)*.

Mao, X.; Liu, Y.; Liu, F.; Li, Q.; Shen, W.; and Wang, Y. 2023. Intriguing findings of frequency selection for image deblurring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 1905–1913.

Nah, S.; Hyun Kim, T.; and Mu Lee, K. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3883–3891.

Pan, J.; Ye, G.; Zhu, Y.; Song, X.; Hu, F.; Zhang, C.; and Wang, M. 2020. Jitter detection and image restoration based on continue dynamic shooting model for high-resolution TDI CCD satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6): 4915–4933.

Tang, X.; Xie, J.; Wang, X.; and Jiang, W. 2014. High-precision attitude post-processing and initial verification for the ZY-3 satellite. *Remote Sensing*, 7(1): 111–134.

Teshima, Y.; and Iwasaki, A. 2007. Correction of attitude fluctuation of Terra spacecraft using ASTER/SWIR imagery with parallax observation. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1): 222–227.

Tsai, F.-J.; Peng, Y.-T.; Lin, Y.-Y.; Tsai, C.-C.; and Lin, C.-W. 2022. Stripformer: Strip transformer for fast image deblurring. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 146–162.

Vimal, V. 2019. Mixture of Gaussian Blur Kernel Representation for Blind Image Restoration. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 10(1): 589–595.

Wang, M.; Fan, C.; Pan, J.; Jin, S.; and Chang, X. 2017. Image jitter detection and compensation using a high-frequency angular displacement method for Yaogan-26 remote sensing satellite. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130: 32–43.

Wang, M.; Zhu, Y.; and Fan, C. 2018. Development of platform jitter geometric analysis and processing for high-resolution optical satellite imagery. *Geomatics and Information Science of Wuhan University*, 43(12): 1899–1908.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17683–17693.

Wang, Z.; Zhang, Z.; Dong, L.; and Xu, G. 2021. Jitter detection and image restoration based on generative adversarial networks in satellite images. *Sensors*, 21(14): 4693.

Wulich, D.; and Kopeika, N. 1987. Image resolution limits resulting from mechanical vibrations. *Optical engineering*, 26(6): 529–533.

Xia, G.-S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; and Zhang, L. 2018. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Xue, W.; Zhang, L.; Mou, X.; and Bovik, A. C. 2013. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE transactions on image processing*, 23(2): 684–695.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5728–5739.

Zhang, Z.; Iwasaki, A.; and Xu, G. 2019. Attitude jitter compensation for remote sensing images using convolutional neural network. *IEEE Geoscience and Remote Sensing Letters*, 16(9): 1358–1362.

Zhu, Y.; Wang, M.; Cheng, Y.; He, L.; and Xue, L. 2018. An improved jitter detection method based on parallax observation of multispectral sensors for Gaofen-1 02/03/04 satellites. *Remote Sensing*, 11(1): 16.