Image improvement based on sub-pixel post-integration for a staring imaging system in geostationary orbit

Xiaoping Tao (陶小平)* and Feng Yan (闫 锋)

Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

**Corresponding author: TaoXP99@gmail.com*

Received December 11, 2013; accepted February 12, 2014; posted online April 4, 2014

This paper proposes the post-integration technology based on sub-pixel image registration and image fusion to improve the signal-to-noise ratio (SNR) of remote sensing images without motion degradation caused by satellite vibration. A two-dimensional vibration system is set up to simulate satellite disturbance. Image sequences with different exposure times are captured using a high-speed CMOS camera. The displacement plots are compared with the motion data measured by the grating linear encoder. These plots indicate that the accuracy of the registration algorithm is better than 0.1 pixels. The sub-pixel image fusion shows an improvement in image quality, thus indicating that this technology is powerful for staring imaging systems in geostationary orbit.

OCIS codes: 280.4788, 110.4153, 110.4155.

doi: 10.3788/COL201412.042802.

The geostationary orbit is a circular orbit that is approximately 36000 km above the equator and follows the rotation direction of Earth. A satellite in such an orbit has an orbital period equal to the rotational period of earth and thus appears motionless, at a fixed position to the ground, which denotes "staring imaging." Most commercial communication, broadcast, and weather satellites operate in geostationary orbit because of continuous observation in the same area. However, the height of 36000 km provides not only the advantage of a large field of view, but also the limitations of low ground sampling distance (GSD) for imaging satellites. The GSD of most geostationary satellites is greater than hundreds of meters[1]. However, the European Space Agency (ESA) has recently been preparing for the Geo-Oculus plan with a GSD of 10.5m(nadir)[2,3]. The Africa-Geosat1 is another project of ESA that has a GSD that is as good as 25 m. Geostationary imaging satellites are expected to serve a more important function in high-resolution remote sensing.

To support the missions of atmosphere, ocean, shoreline, water quality, and other environmental information monitoring, imaging systems usually require several multi-spectral focal planes, such as UV-blue, PAN, Red-NIR, SWIR, MWIR, and TIRchannels[2,3]. Each channel requires different exposure times to satisfy the requirement of signal-to-noise ratio (SNR). For channels with high SNR requiring long exposure time,image acquisition may suffer from perturbations that are mainly generated by moving reaction wheels or cryocoolers, solar array flexible excitation after slew maneuvers, and so on[4−11]. Maintaining a small integration is necessary to avoid pixel saturation and to relax the pointing stability of the satellite. Thus,we propose that several successive images should be captured with shorter integration time to suppress the image smear caused by point drift. Meanwhile, the image sequence is accumulated in the sub-pixel scale to compensate for the SNR loss, which is called "post-integration."

A two-dimensional vibration system described in following section is set up to simulate the oscillation environment on the satellite. A phase correlation method based on the centroid of "energy area" is then chosen for sub-pixel image registration[12−16]. The multi-frame image fusion shows an improvement in image quality.

The perturbations of the satellite imaging system can be divided by the drift along the optical axis and vertical to the optical axis. The lens and the imaging sensor of the space camera are generally fixed together, such that the drift along the optical axis is equal to the shift of the object. Compared with the object distance of 36000 km, the small drift can be ignored. Moreover, the drift vertical to the optical axis includes the translation and rotation of the image plane.In practice, the rotation angle is always small. Thus, the rotation can be approximate to translation in a small regional part of the image. Therefore, only translation is considered in the experiment.

In addition, the conjugate imaging theory suggests that the jitter of the satellite imaging system can be equivalent to the micro-vibration of the target on the ground. Thus,the oscillation of the camera is replaced by a two-dimensional object plane vibration system in the experiment.

As shown in Fig. 1, the vibration simulation system consists of a two-channel signal generator, two power amplifiers, two modal vibration exciters, two elastic spring slices, two flat guides that are vertical to each other, an object plane, two optical grating linear encoders for real-time vibration detection, two air cylinders for counterweight, and two marble pedestals. The signal generator can export sine wave, triangular wave, square wave, random signal, and any other user-defined signal. The amplitude and frequency of signals are both adjustable. Two-channel independent signals are amplified and transferred to the executing unit to make the object plane vibrate in both the horizontal and vertical directions. The displacement of the vibration is captured by the grating linear encoder in realtime, which sets the criterion for checking the result calculated by the image registration algorithm.

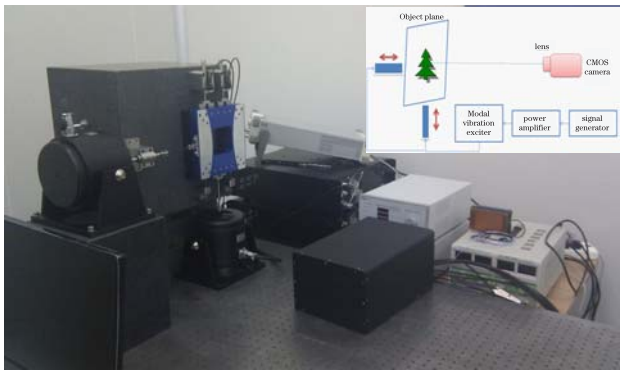

Fig. 1. Experimental setup.

The vibration simulation system is located approximately 0.8 m away from the imaging system, which consists of 2.2/50 mm fixed-focus lens (XENON-EMERALD Schneider, Germany) and high-speed CMOS sensor (MC1360, Mikrotron, Germany). The resolution of the CMOS sensor is 1280×1024, the pixel size is 14×14 (μm), and the maximum frame frequency is 110 fps. The focus of the lens is 50 mm, such that the imaging magnification factor is approximately 1/16. In this way, 0.1 pixels in the image space correspond to 22.4 μm in the object space. The accuracy of the grating linear encoder is better than 1 μm, which satisfiesthe detection requirement.

The phase correlation method based on the centroid of "energy area" for image registration has been proposed in previous research[15]. This method is suitable for images with low intensity and low SNR. The experiment results in Ref. [15] demonstrate that the accuracy of the registration algorithm is better than 0.1 pixels for images of different target scenes, such as city, shore, forest, desert, and so on. Although the SNR decreases from 55 db to 15 db,the accuracy remains the same. The brief procedure is shown in the following.

Assuming that the displacement between the original image $I_1(x, y)$ and the current image $I_2(x, y)$ along the X and Y axes is x_0 and y_0, respectively. The relationship between $I_1(x, y)$ and $I_2(x, y)$ is shown as

$$I_2(x, y) = I_1(x - x_0, y - y_0). \tag{1}$$

Based on Fourier shift theorem, the translation of an image corresponds to the change of phase in the Fourier frequency domain and is given by

$$F_2(u, v) = F_1(u, v) \exp[-\mathrm{i}(ux_0 + vy_0)], \tag{2}$$

where $F_1(u, v)$ and $F_2(u, v)$ are the Fourier transforms of images $I_1(x, y)$ and $I_2(x, y)$. Thus, the normalized cross-power spectrum is defined as

$$\mathrm{Corr}(u, v) = \frac{F_2(u, v)F_1^*(u, v)}{|F_2(u, v)F_1^*(u, v)|} = \exp[-\mathrm{i}(ux_0 + vy_0)], \tag{3}$$

where $F_1^*(u, v)$ denotes the conjugation of $F_1(u, v)$. In Eq. 3, the cross-power spectrum only depends on the phase term $\exp[-\mathrm{i}(ux_0 + vy_0)]$ but has no relation with the amplitude. In addition, the inverse Fourier transform of the phase is the Dirichlet pulse function, shown as

$$F\left(\exp\left(-\mathrm{i}\left(ux_0 + vy_0\right)\right)\right) = \delta\left(x - x_0, y - y_0\right). \tag{4}$$

Equation (4) shows that the location of the maximum of the pulse function is the translation x_0 and y_0. The accuracy of the matching position is calculated through integer pixel. However, the vibration of the satellite platform is usually in the sub-pixel scale. Thus, a centroid method based on "energy area" is proposed to improve the accuracy of the registration algorithm[15]. The peak of the Dirichlet pulse function is defined as a seed in the center. The surrounding region next to the seed is described as the "energy area," in which the energy of each pixel is higher than 0.1% of the maximum. This result is better than four or eight connected domains in our experiments, especially when the peak area is not centrosymmetric. The centroid of the area can then be calculated as

$$x_0 = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} x_i I(x_i, y_j)}{\sum_{i=1}^{M}\sum_{j=1}^{N} I(x_i, y_j)},$$
$$y_0 = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} y_j I(x_i, y_j)}{\sum_{i=1}^{M}\sum_{j=1}^{N} I(x_i, y_j)}, \tag{5}$$

where M and N are the width and height of the energy window, respectively.

Numerous image frames will be observed in the practical experiment. But only the most distinct frame should be chosen to be the reference image. The assessment algorithms of image quality,such as mean-squared error, gray mean gradient, and largest entropy method,are helpful in the selection the proper image.

In the experiment, the data recorded by the grating linear encoder are the displacements of the object plane of the imaging system, whereas the calculation result of the image registration algorithm is the displacement of the imaging plane. Therefore, the so-called "magnification calibration" is applied to transform the two kinds of data into the same scale.

Calibration is performed by imaging a point-pair target fixed on the object plane, as shown in Fig. 2. Five pairs of points are marked as A1–B1, A2–B2, A3–B3, A4–B4, and A5–B5. The distance between the two points of each pair is 20, 30, 40, 50, and 60 mm, respectively. After taking the image of the point pairs, the coordinate of each point is calculated by using the centroid algorithm, as illustrated in Table 1. Accordingly, the distance in pixels between the two points of each pair in the image is alsocomputed. The magnification factor β is defined as the image distance divided by the object distance. Thus, the magnification of the imaging system is $1/15.691 \approx 0.064$ by taking the average of the five β values.

The initialization of the experiment consists of two steps. Firstly, two channels of the signal generator are

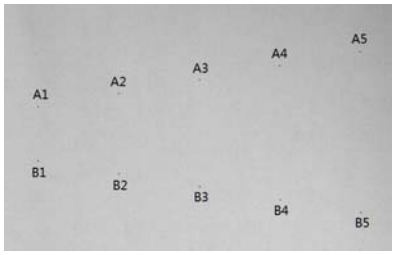

Fig. 2. Point-pair test target for magnification calibration.

Table 1. Calibration Table of the Magnification of the Imaging System

Distance in Object Space (mm)	A	B	Distance in Image Space (pixel)	β
20	(154.99,58.96)	(246.00,58.93)	91.01	1/15.697
30	(133.04,197.02)	(269.01,197.03)	135.97	1/15.759
40	(109.95,333.97)	(291.97,334.01)	182.02	1/15.697
50	(87.95,470.94)	(315.97,471.04)	228.02	1/15.663
60	(65.94,609.02)	(339.99,606.95)	274.06	1/15.638

separately set with different models. Secondly, the power amplifiers are adjusted to excite the object plane, and the optical grating linear encoders simultaneously start to record the displacement of the object plane along the horizontal and vertical directions.

For example, the signal generator is set at 5-Hz sine vibration for both channels, but with different amplitudes. The data detected by the grating linear encoder are shown in Fig. 3, in which the blue solid line illustrates the displacement in the horizontal direction, whereas the red dashed line denotes that in the vertical direction. The amplitudes of the vibration along the two dimensions are 0.75 and 0.51 mm. The vibration in the object space is transformed into the image space by multiplying with the magnification imaging factor, shown as

$$\frac{(D_{\max}-D_{\min})\times\beta}{p}=\begin{cases}\dfrac{748.5\ \mu\text{m}\times 0.064}{14\ \mu\text{m}}\approx 3.4\ \text{pixel}\\ \dfrac{506.5\ \mu\text{m}\times 0.064}{14\ \mu\text{m}}\approx 2.3\ \text{pixel}\end{cases}, \tag{6}$$

where $D_{\max}$ and $D_{\min}$ are the maximum and minimum displacements, and p is the pixel size of the CMOS sensor.

The exposure time of the imaging system is set at 16 ms, 30 successive frames are captured with a 60-fps frame frequency, and the displacement of the image sequence is calculated by image registration. The registration result of the vibration image sequence is shown in Fig.4, in which the blue line with "-+" represents the displacement of the horizontal direction,whereas the red line with "-x" represents that in the vertical direction. The amplitudes of the two dimensions are 3.4 and 2.3 pixels, which are consistent with the data recorded by the grating linear encoder.This result proves that the accuracy of the algorithm is better than 0.1 pixels[15,16].

When the exposure time is shortened from 15 to 2 ms, the illuminance of the imaging sensor, as well as the intensity and the SNR of the image, sequentially decreases. Figures 5(a) to 5(d) show the images captured with 4, 8, 12, and 16-ms exposure time. The contrast of the images is enhanced for observation with a linearly contrast stretch. The highest intensity of the original image is set to 255 in the new image, and the other intensity values are linearly increased. In this way, the SNR will be unchanged.More serious noises are observed in the images with shorter exposure time, as illustrated in Figures. 5(e) to 5(h). All images are normalized by 1, and several "flat areas" (which indicate the regions without structure information) are selected to evaluate the noise level of the images. The variance of noise provided in column 4 of Table 2 is employed for the SNR assessment.

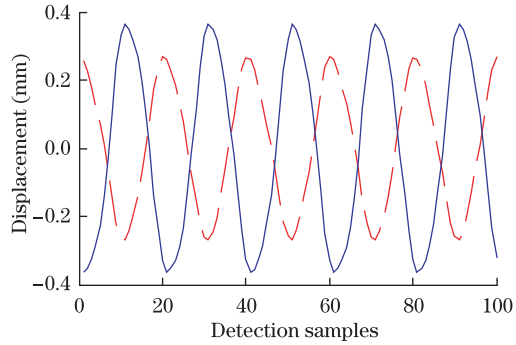

Fig. 3. Vibration detection plot of the grating linear encoder.

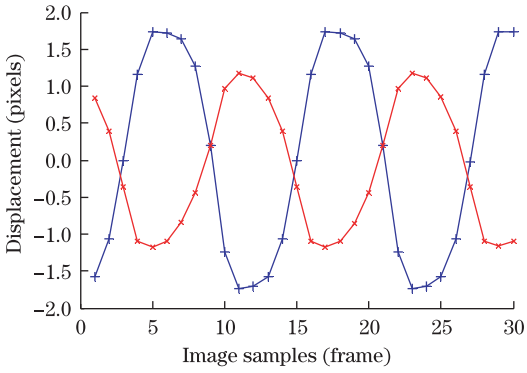

Fig. 4. Image registration plot of the vibration image sequence.

Simultaneously, the vibration during the image capturing period causes the sharpness of the image to change with exposure time. A longer integration time results in more motion blur in the image. Figures. 5(e) to 5(f) are sharper than Figs. 5(g) to 5(h), although with lower SNR. Thus, appropriately decreasing the exposure time and compensating the SNR loss by using sub-pixel multiframe image fusion is a feasible means to improve image quality.

The image sequences with different exposure times are captured for registration in the same manner. The calculated displacement coincides with the result obtained by 16-ms exposure-time images, which validates the robustness of the registration algorithm for low SNR images[15].

Image fusion is based on the sub-pixel displacement obtained by image registration in the previous section. Firstly, the sub-pixel drift between each frame is normalized by 0.2 pixels. Then,each frame is zoomed five times using the cubic interpolation method. Thereafter, the integral pixel drift is compensated in the large-scale image. Finally, the accumulation of all large-scale images is down-sampled to the original scale, in which every 5×5

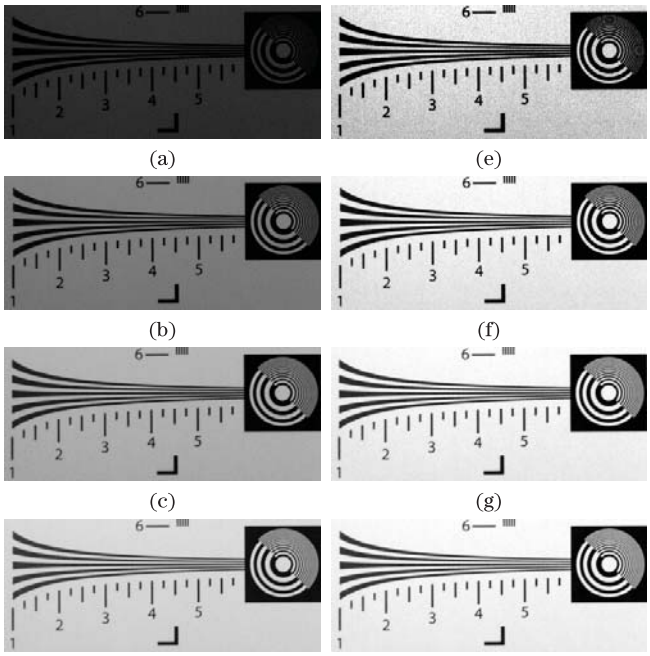

Fig. 5. Images before and after contrast enhancement. (a)–(d) Images captured with 2, 4, 12, and 16-ms exposure time, respectively. (e)–(h) Images when the contrast is stretched for Figs. (a)–(d).

Table 2. Noise Deviation Comparison of Single-image and Multi-image Fusion

Exposure Time (ms)	$I_{\max}$	I_{mean}	S_{Single} (10^{-2})	S_{Fusion} (10^{-3})	$S_{\text{Fusion}}/S_{\text{Single}}$
2	18	1.57	1.74	3.14	5.56
3	37	18.17	3.27	5.27	6.21
4	61	35.82	3.58	5.57	6.43
5	84	52.69	3.47	5.68	6.11
6	105	68.53	3.70	5.98	6.18
7	122	83.89	3.72	6.58	5.66
8	141	97.98	3.57	6.53	5.48
9	152	110.24	3.30	6.32	5.22
10	167	122.26	3.35	6.36	5.26
11	180	133.01	3.47	6.36	5.46
12	190	143.70	3.44	6.52	5.28
13	201	153.10	3.14	6.34	4.96
14	211	162.38	2.97	6.18	4.81
15	220	170.88	2.73	5.48	4.97
16	228	179.57	2.67	5.60	4.76

window is summed to be one pixel. Given that random noises, which are located in different positions, are found in every image, when the image sequences after displacement compensation are added together, the useful signals are integrated in the same position, whereas the noises remain distributed in different locations. Next, when the accumulation is divided by the fame number N, the energy of useful signals remains the same, whereas that of the noise in the original position is decreased by the average. Thus, the SNR of the image is improved.

Each group of images with different exposure times is fused in this manner. The fusion of the images with 4, 8, 12, and 16-ms exposure time is displayed in Fig. 6. As shown in the figure, the noise of a multi-frame fused image is weakened.The noise deviation of the fused image is presented in column 5 of Table 2. Compared with a single frame, the SNR of the fused image is increased nearly five times to six times, as shown in column 6. Thus, the SNR improvement of sub-pixel post-integration with N frames is approximated to the square root of N ($\sqrt{30} \approx 5.5$).

Notably, the mass data introduced by multi-frame capturing is a challenge to satellite storage and transmission. However, only several interesting regions exist in the full frame. These regions include the areas suffering from flood, forest fire, windstorm, algal bloom, oil slick pollution, and so on. These areas can be separately identified for image registration and fusion. For example, if the resolution of the imaging sensor is 10×10 (k), the data of one 256×256 region is only 1/1600 of the full frame, which is acceptable for state-of-the-art satellites.

In conclusion, to support the missions of atmosphere, ocean, and disaster monitoring, imaging systems in geostationary orbit usually require multi-channel focal planes for information analysis, which require different exposures to satisfy SNR demands. To suppress the motion degradation caused by the attitude drift and jitter of the satellite, image acquisition with long exposure time is suggested to be split into several successive images with short integration time. The SNR loss is compensated by post-integration technology based on sub-pixel image registration and fusion. A two-dimensional vibration system is set up to simulate satellite disturbance. Image sequences with different exposure times are captured for sub-pixel registration, and the multi-frame image fusion shows the improvement in image quality. The experiments validate the feasibility of post-integration technology in prospective remote sensing.

This work was supported by the National Natural Science Foundation of China (No.61036015). We wish to thank the reviewers for their comments and suggestions that have improved the content of this letter.

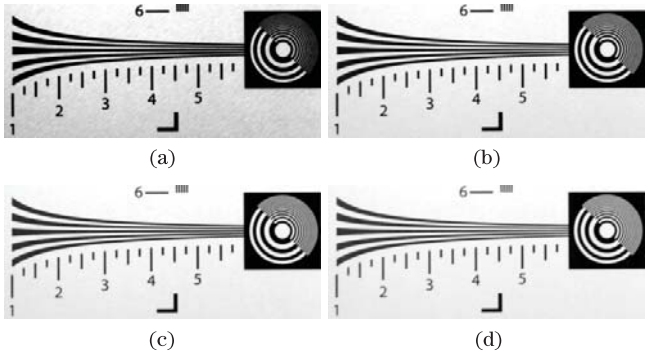

Fig. 6. Comparison of images with different exposure time, which are fused with 30 frames, and with enhanced contrast: fusion of images with (a) 4-ms exposure time, (b) 8-ms exposure time, (c) 12-ms exposure time, and (d) 16-ms exposure time.

References

1. L. Guo, Z. Deng, J. Tao, X. Sun, B. Wang, S. Pei, W. Liang, and W. Gu, Spacecraft Recovery Remote Sensing **6**, 23 (2010).
2. L. Vaillon, U. Schull, T. Knigge, and C. Bevillon, in *Proceedings of International Conference on Space Optics* (2010).
3. Astrium GmbH, "GeoOculus: A Mission for Real-Time Monitoring through High-Resolution Imaging from Geostationary Orbit", Final Report (2009).
4. M. Mesrine, E. Thomas, S. Garin, P. Blanc, C. Alis, F. Cassaing, and D. Laubier, in *Proceedings of 6th Internat. Conf. on Space Optics* (2006).
5. P. Blanc and G. Monroig, in *Proceedings of 6th Internat. Conf. on Space Optics* (2006).
6. M. Toyoshima, Y. Takayama, and H. Kunimori, Opt. Eng. **49**, 083604 (2010).
7. M. Toyoshima and K. Araki, Opt. Eng. **40**, 827 (2001).
8. G. E. Mosier, M.Femiano, K. Ha, P. Y. Bely, R. Burg, D. C. Redding, A. Kissil, J. M. Rakoczy, and L. D. Craig, in *Proceedings of SPIE Conference on Space Telescopes and Instruments* 1070 (1998).
9. S. Y. Wang, "The analyse of imaging quality of the optical system in geostationary orbit", PhD. Thesis (Harbin Institute of Technology, 2009).
10. T. Iwata, Proc. SPIE **5659**, 34 (2005).
11. K. Janschek, V. Tchernykh, and S. Dyblenko, Control Engineering Practice **15**, 333 (2007).
12. M. Guizar-Sicairos, S. Y. Thurman, and J. R. Fienup, Opt. Lett. **33**, 156 (2008).
13. B. Pan, H. Xie, B. Xu, and F. Dai, Meas. Sci. Technol. **17**, 1615 (2006).
14. H. Foroosh, J. B. Zerubia, and M. Berthod, IEEE Transactions on Image Processing **11**, 188 (2002).
15. X. Tao, D. Xue, F. Li, and F. Yan, Acta Photon. Sin. **41**, 1359 (2012).
16. X. Tao, X. Luo, and D. Xue , Opt. Precision Engineering **21**, 2169 (2013).