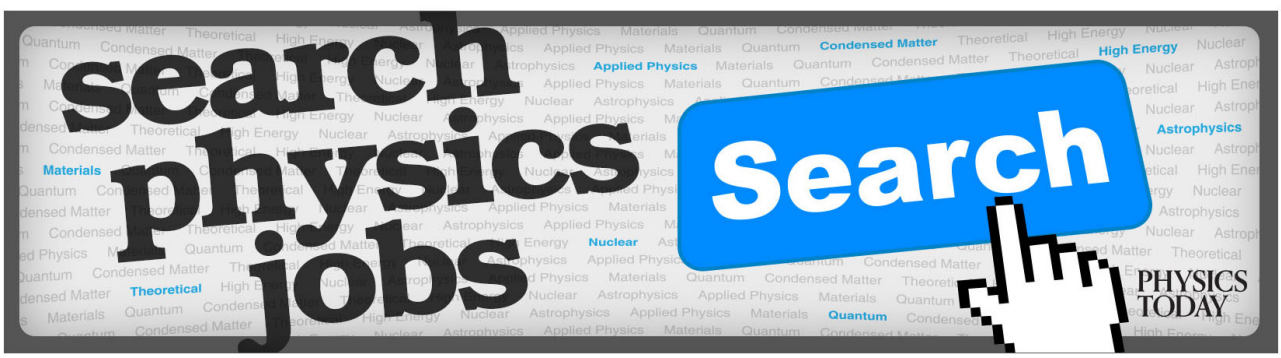
AIP | Review of Scientific Instruments

Extracting flat-field images from scene-based image sequences using phase correlation

James N. Caron, Marcos J. Montes, and Jerome L. Obermark

Citation: Review of Scientific Instruments **87**, 063710 (2016); doi: 10.1063/1.4954730
View online: http://dx.doi.org/10.1063/1.4954730
View Table of Contents: http://scitation.aip.org/content/aip/journal/rsi/87/6?ver=pdfcov
Published by the AIP Publishing



Extracting flat-field images from scene-based image sequences using phase correlation

James N. Caron,[1,a)] Marcos J. Montes,[2] and Jerome L. Obermark[3]
[1]*Research Support Instruments, 4325-B Forbes Boulevard, Lanham, Maryland 20706, USA*
[2]*Naval Research Laboratory, Code 7231, 4555 Overlook Avenue, SW, Washington, DC 20375, USA*
[3]*Naval Research Laboratory, Code 8231, 4555 Overlook Avenue, SW, Washington, DC 20375, USA*

(Received 23 December 2015; accepted 11 June 2016; published online 27 June 2016)

Flat-field image processing is an essential step in producing high-quality and radiometrically calibrated images. Flat-fielding corrects for variations in the gain of focal plane array electronics and unequal illumination from the system optics. Typically, a flat-field image is captured by imaging a radiometrically uniform surface. The flat-field image is normalized and removed from the images. There are circumstances, such as with remote sensing, where a flat-field image cannot be acquired in this manner. For these cases, we developed a phase-correlation method that allows the extraction of an effective flat-field image from a sequence of scene-based displaced images. The method uses sub-pixel phase correlation image registration to align the sequence to estimate the static scene. The scene is removed from sequence producing a sequence of misaligned flat-field images. An average flat-field image is derived from the realigned flat-field sequence. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4954730]

I. INTRODUCTION

Flat-field correction is a process that improves the quality of digital images. Image artifacts are created by pixel-by-pixel variations resulting from variations in gains and dark currents between pixels of a focal plane array. Dark currents are removed by imaging a scene with no illumination. This "dark frame" is then subtracted off the image of interest. The variation in gains is typically accounted for by imaging a uniformly white scene such as that created by a diffuser or an integrating sphere, producing the radiometrically uniform flat-field image. Subsequent images taken by the camera system are divided by the normalized flat-field, resulting in an improved representation of the scene.

There are situations where a flat-field image cannot be accurately captured, such as imaging via satellite. Once deployed in orbit, the camera electronics become susceptible to the radioactive environment and degrade over time, requiring periodic dark frames and flat field images to be taken. This was the case for the Hubble Space Telescope (HST) Faint Object Camera[1] and the Solar and Heliospheric Observatory (SOHO) Michelson Doppler Imager.[2] For both operations, the imaging system did not have an on-board calibration source to capture a flat-field image for use in processing. Instead, the researchers were able to extract a flat-field representation by knowing the precise movement of a scene in relation to the image coordinates. The HST team was able to sufficiently control the pointing of the telescope such that a sequence of images could be obtained where the translation differences between frames were known. The SOHO team had a fixed pointing position, but their target, the sun, rotated in a predictable manner. This approach was first described by Kuhn, Lin, and Loranz in 1991,[3] and has more recently been proposed for use on the Helioseismic and Magnetic Imager.[4] The techniques produced a sequence of images where the scene was translated in a known manner with respect to the fixed position of the pixels. With known translational differences, the groups were able to extract a suitable flat-field image from the sequences.

In this paper, we consider the case where the registration differences between frames are not known. By applying an image registration method based on phase-correlation, the translational differences are determined to sub-pixel accuracy, and the images are aligned. Once aligned, the flat-field image is extracted and applied to the original images. This extends the use of flat-field correction to systems where positioning cannot be accurately determined independently or there is significant jitter on the imaging platform.

The process, shown in Figure 1, will be described in Sec. II and tested using a simulated set of satellite imagery.

Techniques that provide scene-based non-uniform correction to images have been researched for several decades. These methods mostly differ in their method of registration. The techniques include gradient-based registration,[5] least-mean square approaches,[6] algebraic methods,[7–9] and the application of Kalman filters.[10,11]

To the best of our knowledge, use of phase correlation for scene-based non-uniform correction has not been previously reported. Phase correlation has the advantage of measuring translation differences that range from sub-pixel movements to hundreds of pixels efficiently in a single application. Since the alignment calculations are conducted in the spatial frequency domain, the phase correlation is less dependent on edges and point-like objects allowing application to images with varying levels of blur and noise.[12] With the additional use of a Log-Polar transformation, phase correlation can determine rotation and scale differences between images.[13]

[a)]Author to whom correspondence should be addressed. Electronic mail: Caron@RSImd.com. URL: www.RSImd.com.

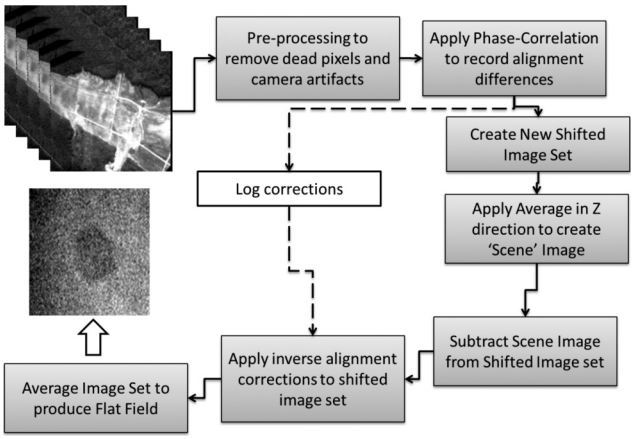


FIG. 1. The process for extracting a flat-field image from an image sequence.

II. APPROACH

The process requires a sequence of N images $I_z(m,n)$ with frame index z and pixel position (m,n) where each frame is displaced from the reference frame $I_c(m,n)$. We assume that the minimum translational difference between any two frames is greater than a pixel, but less than the dimensions of the image. We also assume that a dark frame has been acquired and removed from the image sequence, such that fixed pattern noise does not interfere with the registration process. The scene should be static and have sufficient illumination such that all pixels are above their noise level.

The goal of the registration is to align the images to a chosen reference frame such that a good representation of the scene content $I_{scene}(m,n)$ can be created and separated in each frame from the structure of the flat-field $\gamma(m,n)$. The first step in the process is to subtract off the dark frame from each image. Hot and dead pixels can also adversely affect the alignment process. These can be mitigated by applying a smoothing filter, or two-dimensional median filter. Improved image quality can be achieved using a selective mean filter such as the "Signal Dependent Rank Ordered Mean Filter."[14]

Next, the images are aligned to a reference frame which is often the middle frame of the set. The alignment method must account for all necessary degrees of freedom and be reversible. Using a combination of phase correlation, described in Sec. III, and four-point mapping,[15] we successfully applied this method to image sequences with translation, rotation, scale, and perspective differences.

A single frame in the sequence can be represented as

$$I_z(m,n) = \gamma(m,n) R_z[I_{scene}(\hat{m},\hat{n})], \tag{1}$$

where $R_z[\cdots]$ represents the application of a set of operations, such as translations, rotations, and scale, that displaces each image with respect to the control image $I_c(m,n)$. For affine transformations, the operator $R_z[\cdots]$ has an identity, an inverse, and is associative.[16] The aligned image set is represented as

$$\begin{aligned} K_z(m,n) &= R_z^{-1}[\gamma(\hat{m},\hat{n}) R_z[I_{scene}(m,n)]] \\ &\approx R_z^{-1}[\gamma(\hat{m},\hat{n})] I_{scene}(m,n), \end{aligned} \tag{2}$$

where $R_z^{-1}[\cdots]$ represents the set of operations that align each target image to the control image $I_c(m,n)$.

Provided that the registration process is accurate to the sub-pixel level, each frame in the K_z sequence will display a static scene while the flat-field structure varies from frame-to-frame. With enough frames

$$I_{scene}(m,n) \approx K_{ave}(m,n), \tag{3}$$

based on the premise that

$$\frac{1}{N_p} \sum_{N_p} R_z^{-1}[\gamma(\hat{m},\hat{n})] \approx 1, \tag{4}$$

for all pixel values N_p that provide viable information. As a result of motion, pixels near the edges of the image often do not contain N viable data points. A threshold is applied such that only N_p viable data points are included in the average per pixel location.

The accuracy of the approximation improves with the number of frames. An optimum approximation is achieved when the features in the flat field structure have been averaged over many pixel locations such that the flat field features are literally lost in the image noise. This produces the condition

$$\frac{\sigma_{ff}}{\sqrt{N}} < \frac{\sigma_I}{\langle I(m,n) \rangle}, \tag{5}$$

where σ_{ff} is the standard deviation of the flat field, $\langle I(m,n) \rangle$ is the image average over all pixels, and σ_I is the standard deviation of the image representing the noise level. Thus, if the flat field has a standard deviation of 0.2, and the image noise is 5 per cent of the image mean, sixteen images would be required to fully flat-field the image. As discussed, some cases will require more images since the motion may not produce distinct pixel values for all pixel locations.[5]

Following the image alignment, the average of the set $K_z(m,n)$ is taken in the $\hat{z}$ dimension, producing $K_{ave}(m,n)$, the representation of the static scene. Alternatively, a median filter can be applied if significant outliers are present in the data set. The scene is removed from the image set using

$$\begin{aligned} L_z(m,n) &= K_z(m,n)/K_{ave}(m,n) \\ &\approx \frac{R_z^{-1}[\gamma(\hat{m},\hat{n}) R_z[I_{scene}(m,n)]]}{I_{scene}(m,n)} \\ &\approx R_z^{-1}[\gamma(\hat{m},\hat{n})]. \end{aligned} \tag{6}$$

In the absence of noise, each frame in $L_z(m,n)$ reproduces the flat-field structure with known registration differences. Having saved the alignment corrections, an inverse alignment is applied to the sequence such that the pixel variations reside in the location of the pixel that produced them. Taking the average for each pixel location in the sequence produces the extracted flat-field image represented by

$$F(m,n) = \frac{1}{N_p} \sum_{z=0}^{N_P-1} R_z\left[R_z^{-1}[\gamma(m,n)]\right], \tag{7}$$

where N_p is the number of viable pixels values for the specific pixel location.

III. IMAGE REGISTRATION

The strength of this method relies on the accuracy of the registration algorithm. Phase correlation is a well established method that determines the translational differences between images. The earliest reference to phase correlation is from a paper by Kuglin and Hines published in 1975.[17] An article by Zitová and Flusser in 2003[18] provides a more recent survey of image registration methods.

Given two images, $I_c(m,n)$ and $I_t(m,n)$, that differ only by a non-integer translation $(\Delta x, \Delta y)$,

$$I_t(m,n) = I_c(m - \Delta x, n - \Delta y), \tag{8}$$

the Fourier transforms of those images are related by

$$F_t(u,v) = e^{-i(u\Delta x + v\Delta y)} F_c(u,v), \tag{9}$$

where (u,v) are the coordinates in frequency space. Thus, translations in image space appear as phase changes in frequency space. The translation errors can be derived from the phase difference between the transformed images. The phase difference is equivalent to

$$e^{(u\Delta x + v\Delta y)} = \frac{F_c(u,v)F_t^*(u,v)}{|F_c(u,v)F_t^*(u,v)|}, \tag{10}$$

where $*$ represents the complex conjugate. For translational differences, the phase difference appears as a sinusoidal pattern. Application of a Fourier transform to the phase difference ideally contains a delta function centered at the displacement coordinates. The $\hat{x}$ and $\hat{y}$ distances from this point to the center of the image are equal to the translation differences between the two images.

An effective method for determining the sub-pixel location of the displacement coordinates is to calculate the center of the peak by applying the centroid formula. This is a computationally efficient method that avoids up-sampling the data.[19] If the peak in the x direction is identified at m_p, then the translation difference is

$$\Delta x = \frac{\sum_{m=m_p-a}^{m_p+a} m\, I_{cp}(m,n)}{\sum_{m=m_p-a}^{m_p+a} I_{cp}(m,n)}, \tag{11}$$

where a is the local search region, usually set to $a = 2$. Similarly,

$$\Delta y = \frac{\sum_{n=n_p-a}^{n_p+a} n\, I_{cp}(m,n)}{\sum_{n=y_p-a}^{n_p+a} I_{cp}(m,n)}, \tag{12}$$

for the y direction.

The accuracy of the approach depends on the noise level in the correlation image and image structure, but accuracy greater than 1/10th of a pixel can typically be achieved. More sophisticated methods have been developed that achieve accuracy to greater than 1/100th of a pixel.[19,20]

Phase correlation can be extended to measure rotation and scale differences by transforming Cartesian image coordinates to log-polar coordinates, a method first described by Reddy and Chatterji.[13] As with translation differences, this method is responsive to comparatively large differences in rotation and scale.[21,22] For unknown perspective differences, phase correlation can be applied separately to image quadrants. The relative difference between each quadrant can be used to correct for the perspective difference using 4-point mapping techniques.

IV. IMAGE SEQUENCE SIMULATION

As mentioned, this flat-field extraction process was developed for sequences of images that had perspective, rotation, scale, and translation differences. The method is applied here to a simulated sequence that possesses constant translational changes and random jitter. The use of a simulated set allows errors to be quantified.

The simulation process starts with a 1k × 1k image with good signal-to-noise. A sequence of translated images is created by repeatedly cropping out 512 × 512 sections of the image, displaced in the $\hat{x}$ and $\hat{y}$ directions according to

$$\begin{aligned} \Delta x_z &= dx * z + j_t * \varrho_{xz}, \\ \Delta y_z &= dy * z + j_t * \varrho_{yz}, \end{aligned} \tag{13}$$

where dx and dy are non-integers, z is the frame index, j_t is the average amount of jitter, and ϱ_{xz} and ϱ_{yz} are random numbers between −0.5 and 0.5.

Figure 2 (top) shows a LandSat image of Gettysburg, Pennsylvania, USA that served as the base image. Our chosen scene has a city center that is traversed by roads and surrounded by farms. The motivation for choosing this image is the scene possesses a fair amount of structure that will enable accurate registration. All image pixels have values above the sensor's noise level ensuring that each pixel of the flat-field will have signal. The original LandSat image was cropped and converted to grayscale for use in this simulation.

Also shown in Figure 2 is a flat-field image (middle) created by combining several real flat-field images, containing different types of artifacts, such as uneven illumination, particles on the focal plane array, and banding in the vertical direction. After the translation provided by Equation (13) is applied to the image, the flat-field is applied and random noise is added. The third frame in the simulated sequence is shown in Figure 2.

A sequence of $N = 16$ frames was created using $dx = 4.5$ pixels, $dy = 3.6$ pixels, and $j_t = 2.9$ pixels. The flat-field image was scaled such that the values varied from 0.74 to 1.2 and had a mean value of one. Incoherent noise was added by creating an image with random values that fell between −0.5 and 0.5. The noise image was scaled by multiplying the pixel values by 0.15 times the maximum value of the target image, before adding the noise image to the target image. The signal-to-noise ratio of 15 was high enough to produce visible noise in each frame without obscuring most objects. In comparison, the signal-to-noise ratio of the original image was about 57.

V. ACCURACY OF REGISTRATION

The graph in Figure 3 shows the input values for the simulation using the Gettysburg image. Translations range from 35.91 to −32.19 pixels in the $\hat{x}$ direction and 27.89 to

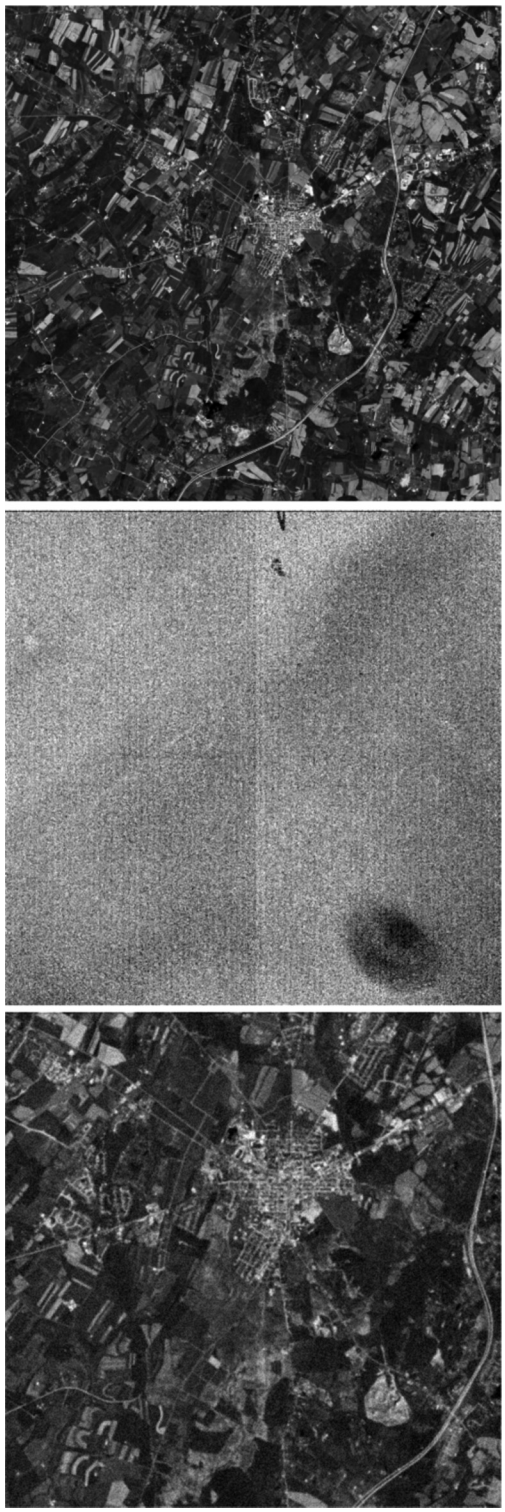


FIG. 2. Images used for the simulation. The base image for the simulation is an aerial image of Gettysburg taken by LandSat 8. The image was translated, cropped, and converted to grayscale for each frame in the simulation. A simulated flat-field image (middle) was created from combining several real flat-field images. The simulated frame (bottom) was produced by applying the flat-field image and random noise to the translated images.

−26.08 pixels in the $\hat{y}$ direction over 16 frames. After applying the Phase Correlation image registration, the alignment differences were calculated. These are also shown in the graph but as a result of the accuracy, they are overlapped by the input

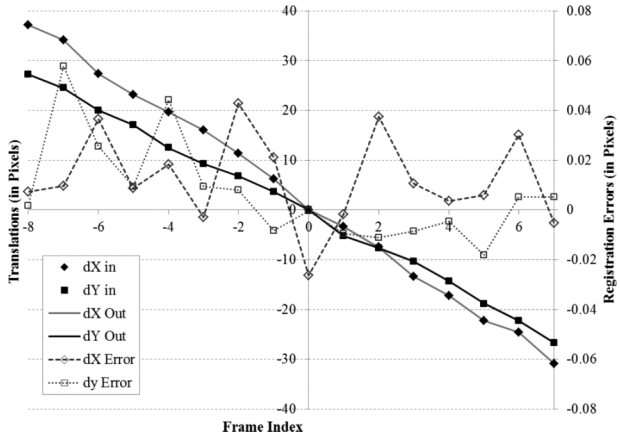


FIG. 3. A plot of the translations used in the simulation. The input (dX in and dY in) and output values (dX Out and dY Out) displayed using the y-axis on the left overlap on this scale, as the errors are two orders of magnitude smaller. The difference between the input and output translation values is displayed with the y-axis on the right.

values. The error, expressed in pixels on the second $\hat{y}$ axis, is shown, ranging from −0.0578 to 0.0263 pixels. The standard deviation of the error is 0.0193 pixels for the $\hat{x}$ direction and 0.0204 pixels for the $\hat{y}$ direction. The maximum error was 0.0578 pixels, demonstrating that the phase correlation is accurate for this combination of scene, noise, and movement to at least 1/17th of a pixel.

VI. RESULTS

The extracted flat-field image is shown in Figure 4 with the same image contrast as Figure 2 (middle). The image exhibits the small scale artifacts and vertical banding. The large-scale features are discernable, but not as pronounced. This is demonstrated by the profile plots shown in Figure 5,

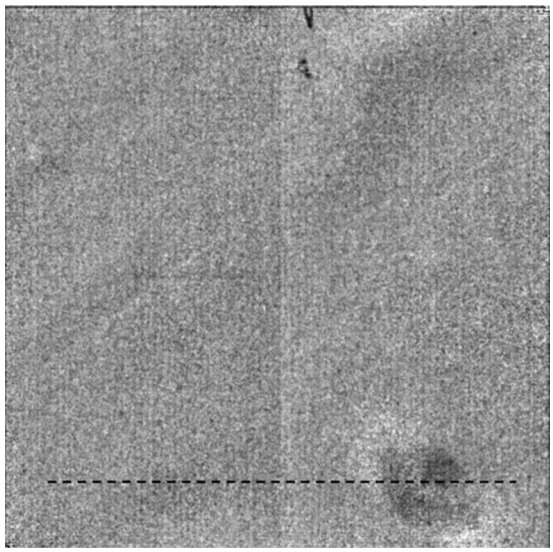


FIG. 4. The extracted flat-field image displays much of the small artifacts from the original flat-field image. Large scale structures are not as pronounced. The dashed line indicates the data for profile plots shown in Figure 5.

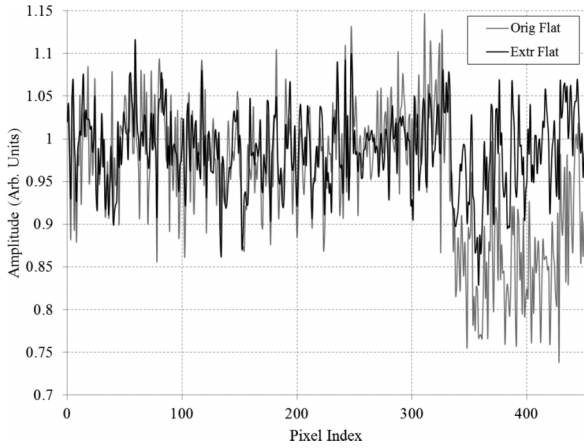


FIG. 5. Profile plots taken from the original and extracted flat-field images taken from the line shown in Figure 4. The variations in the extracted flat-field image closely match the variations of the original aside from the large artifact in the lower-right corner.

as derived from the images across the dashed line shown in Figure 4.

The variations in the extracted image closely resemble those in the original flat-field image. The two lines diverge for the larger artifact in the lower right corner of the image. This behavior is to be expected since each pixel in the extracted flat-field image is derived from a set of neighboring pixels. At most, a single pixel in our simulation is dependent on the pixels in the neighboring area of about 3600 square pixels, where as the artifact covers an area of 7600 square pixels. Thus, recovery of large-scale artifacts is not as successful as smaller ones.

The extraction process took 22.94 seconds on a standard office computer using non-optimized Fast Fourier Transform routine. Optimizing the FFT should result in an order of magnitude reduction in the processing time.

VII. CONCLUSIONS

We have described a method of extracting a flat-field image from a sequence of translated non-uniform images, where the alignment differences are not previously known. Although a flat-field image should be measured empirically whenever possible, this method can greatly improve the quality of images taken by systems that have degraded over time, and a flat-field image cannot be produced by conventional means. The technique requires a sequence of partially overlapping scenes. Phase correlation image registration has been utilized to achieve sub-pixel alignment of the image sequence and accurate inverse alignment of the extracted flat-field images.

As demonstrated by our simulation, the extracted flat-field effectively reproduces the high spatial frequency of the original flat-field, but is not as effective in recovering low spatial frequency. The success of recovering low-frequency spatial content depends on the size of the translations and the number of frames in the sequence.

[1]P. Greenfield, in *Astronomical Data Analysis Software and Systems III*, edited by D. R. Crabtree, R. J. Hanisch, and J. Barnes, ASP Conference Series Vol. 61 (Astronomical Society of the Pacific, 1994), pp. 276–279.
[2]H. E. Potts and D. A. Diver, Astron. Astrophys. **492**, 863 (2008).
[3]J. R. Kuhn, H. Lin, and D. Loranz, Publ. Astron. Soc. Pac. **103**, 1097 (1991).
[4]R. Wachter and J. Schou, Sol. Phys. **258**, 331 (2009).
[5]R. C. Hardie, M. M. Hayat, E. Armstrong, and B. Yasuda, Appl. Opt. **39**, 1241 (2000).
[6]D. A. Scribner, K. A. Sarkady, J. T. Caulfield, M. R. Kruer, G. Katz, and C. J. Gridley, in *Infrared Detectors and Focal Plane Arrays*, edited by E. L. Dereniak and R. E. Sampson (SPIE, 1990), pp. 224–233.
[7]B. M. Ratliff, M. M. Hayat, and J. S. Tyo, J. Opt. Soc. Am. A **22**, 239 (2005).
[8]Y.-J. Liu, H. Zhu, and Y.-G. Zhao, Appl. Opt. **48**, 2364 (2009).
[9]B. M. Ratliff, M. M. Hayat, and J. S. Tyo, J. Opt. Soc. Am. A **20**, 1890 (2003).
[10]S. N. Torres, J. E. Pezoa, and M. M. Hayat, Appl. Opt. **42**, 5872 (2003).
[11]A. Averbuch, G. Liron, and B. Z. Bobrovsky, Image Vision Comput. **25**, 833 (2007).
[12]H. Foroosh, J. B. Zerubia, and M. Berthod, IEEE Trans. Image Process. **11**, 188 (2002).
[13]B. S. Reddy and B. N. Chatterji, IEEE Trans. Image Process. **5**, 1266 (1996).
[14]E. Abreu, M. Lightstoneand, S. K. Mitra, and K. Arakawa, IEEE Trans. Image Process. **5**, 1012 (1996).
[15]B. Jähne, *Image Processing for Scientific Applications* (CRC Press, Boca Raton, 1997).
[16]Y. Hel-Or and P. C. Teo, in *Proceedings of the 1998 IEEE International Symposium on Circuits and Systems, Monterey, CA, USA, May 31-June 3, 1998* (IEEE, 1998), pp. 337–340.
[17]C. D. Kuglin and D. C. Hines, in *Proceedings of the International Conference on Cybernetics and Society, San Franciso, CA, USA, September 23-25, 1975* (IEEE, 1975), pp. 163–165.
[18]B. Zitová and J. Flusser, Image Vision Comput. **21**, 977 (2003).
[19]M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, Opt. Lett. **33**, 156 (2008).
[20]K. Takita, Y. Sasaki, T. Higuchi, and K. Kobayashi, IEICE Trans. Fundam. Electron., Commun., Comput. Sci. **86**, 1925 (2003).
[21]Y. Keller, A. Averbuch, and M. Israeli, IEEE Trans. Image Process. **14**, 12 (2005).
[22]J. N. Sarvaiya, S. Patnaik, and K. Kothari, J. Pattern Recognit. Res. **7**, 90 (2012).