Restoration and Reconstruction from Overlapping Images for Multi-Image Fusion

Stephen E. Reichenbach and Jing Li

Abstract—**This paper describes a technique for restoring and reconstructing a scene from overlapping images. In situations where there are multiple, overlapping images of the same scene, it may be desirable to create a single image that most closely approximates the scene, based on the data in all of the available images. For example, successive swaths acquired by NASA's moderate imaging spectrometer (MODIS) will overlap, particularly at wide scan angles, creating a severe visual artifact in the output image. Resampling the overlapping swaths to produce a more accurate image on a uniform grid requires restoration and reconstruction. The one-pass restoration and reconstruction technique developed in this paper yields mean-square optimal resampling, based on a comprehensive end-to-end system model that accounts for image overlap and is subject to user-defined and data-availability constraints on the spatial support of the filter.**

I. INTRODUCTION

RESAMPLING is required in many imaging applications and is particularly important in remote sensing to correct for geometric distortion and to register, rescale, or otherwise remap. Resampling requires *reconstruction*, the process of determining values at arbitrary spatial locations. Determining the best value requires *restoration*, the process of correcting for degradations that are introduced during the imaging process in order to obtain more accurate estimates of the scene radiance field. With multiple images, a resampling technique should use all available data in restoring and reconstructing the scene.

This paper develops a technique for multi-image fusion that, in one-pass through overlapping input images, restores and reconstructs the scene radiance field. The technique is effective because it maximizes fidelity based on a comprehensive end-to-end system model that accounts for scene statistics, acquisition blurring, sampling, and noise. The technique is efficient because the filter is derived subject to user-defined and data-availability constraints on spatial support for spatial-domain processing.

Our approach is similar in some respects to the interpolation method described by Moreno and Melia [1]. Both techniques account for nonuniform sampling and overlap related to off-nadir satellite geometries. Both methods use mean-square difference metrics, but our technique minimizes the mean-square difference between the scene and the resampled image while the interpolation method in [1] minimizes the mean-square difference between the original system response and the postresampling response. Our method accounts for the scene statistics and noise to perform restoration during resampling.

Section II presents the end-to-end imaging system model that is the basis for the derivation and extends the imaging system model to account for overlapping images. Section III describes the derivation of the optimal constrained filter. It defines the performance criteria for the derivation, presents the derivation of the optimal unconstrained filter for overlapping images, and then imposes user-defined and spatial-availability constraints on the derivation. It also describes performance evaluation of the optimal constrained filters. Section IV gives the experimental results. It describes the experimental simulation of a wide-angle, scanning imaging system like NASA's moderate imaging spectrometer (MODIS) presents the expected experimental performance of the optimal unconstrained filter, gives the actual experimental results of the optimal constrained filters, and compares these results with the results of the nearest-neighbor reconstruction. Section V looks briefly at the more general problem of multi-image fusion.

II. DIGITAL IMAGING SYSTEM MODEL WITH OVERLAPPING IMAGES

A. Digital Imaging System Model

This section presents a model of the basic components of the digital image acquisition process. The model is a simplification of the more complex interactions in real imaging systems, but it captures the most fundamental effects of the acquisition process. Even a well designed imaging system introduces degradations that make the output image an imperfect representation of the scene. Fig. 1 illustrates a comprehensive, end-to-end, continuous-discrete-continuous (CDC) model of the normal operation of a digital imaging system with digital image processing. The optics, scanner, and photodetector cause blurring. The combined response is characterized by the acquisition point-spread function (PSF) h. Spatial sampling with a uniform, rectangular lattice causes aliasing. During acquisition, noise is introduced including shot noise, circuit noise, and quantization error.

Mathematically, we model the image acquisition PSF as a linear shift-invariant (LSI) process. The LSI operation is a convolution process. Here, we assume dimensional separability for simplicity of presentation, but the approach does not depend upon this assumption. The combined PSF of the optics, scanner (motion during temporal integration), and detector (spatial integration) is

$$h(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_o(x-x_0)h_s(x_0-x_1)h_d(x_1) \, dx_0 \, dx_1 \quad (1)$$

The authors are with the Computer Science and Engineering Department, University of Nebraska, Lincoln, NE 68588-0115 USA (e-mail: reich@cse.unl.edu; jli@cse.unl.edu).
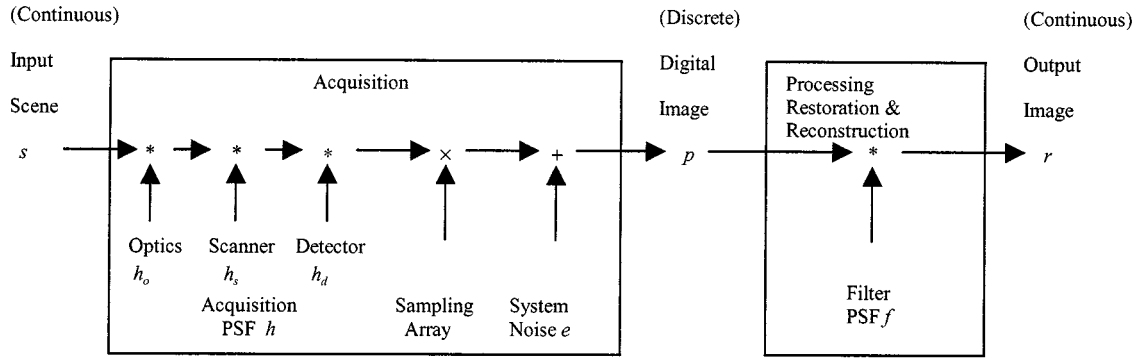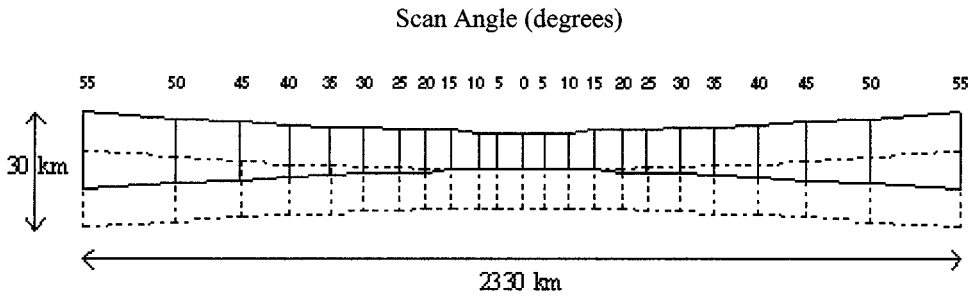
Fig. 1. Digital imaging system model.

Fig. 2. Scan overlap due to GIFOV as a function of scan angle.

Fig. 3. Landsat TM scene.

where

h_o optical system PSF;
h_s scanner PSF;
h_d detector PSF.

Using the system model illustrated in Fig. 1, the digital image p is

$$p(x_n) = \int_{-\infty}^{\infty} h(x_n - x)s(x)\,dx + e(x_n) \qquad (2)$$

where the continuous scene radiance field s is convolved with the presampling acquisition PSF h, sampled at spatial location x_n and degraded by noise e. In reality, some noise is caused by spatially continuous processes, but for any such processes, one can define a discrete noise process that has statistically identical effects on the image. Spatial coordinates x are normalized to the ground sampling interval (GSI), and pixels have integer indices n that determine spatial location x_n within the field of view (FOV). This is a fairly modest acquisition model, but it is adequate to demonstrate the radiometric issues in resampling and to develop an improved restoration and resampling technique.

The filter f restores and reconstructs a continuous image that then can be resampled. The resulting (continuous) image r is

$$r(x) = \sum_{n=-\infty}^{\infty} f(x - x_n)p(x_n). \qquad (3)$$

As a practical matter, only a finite portion of the image p is available. To address the limited FOV and to facilitate Fourier analysis, it is common to assume the scene and hence the image are periodic with period equal to the FOV and to constrain the filter support to the extent of the image (or smaller). The effect of this assumption in the filter derivation is negligible because the image FOV is large relative to the scene mean spatial detail and system PSF.

B. Overlapping Swaths in MODIS Images

In a wide-angle scanning imager such as MODIS, pixel size is a function of scan angle. As the scan angle from nadir increases, the cross-track pixel size increases at a faster rate than the along-track pixel size. For example, a detector with 1 km square ground-projected instantaneous FOV (GIFOV) at nadir (bands 8–36) has a GIFOV of 4.83 km cross-track by 2.01 km

TABLE I
IMAGE SIZE, GSI, AND GSI RATIOS FOR THE SIMULATED MODIS IMAGE TILES

View angle (Radians)	Image size (Pixels)	GSI (Meters)	wfac (GSI ratio)	hfac (GSI ratio)
0.0826	465x467	251x250	1.0032	0.9990
0.2426	437x466	267x257	1.0676	1.0289
0.3890	390x465	299x272	1.1977	1.0863
0.5170	334x463	349x292	1.3962	1.1671
0.6254	280x461	417x317	1.6670	1.2668
0.7156	232x459	504x345	2.0163	1.3812
0.7900	190x458	613x377	2.4519	1.5071
0.8511	156x456	746x410	2.9837	1.6416
0.9014	129x455	907x446	3.6264	1.7830
0.9428	106x454	1099x482	4.3964	1.9295

TABLE II
SIMULATION PARAMETERS FOR THE ACQUISITION TRANSFER FUNCTION

View angle (Radians)	Cross-Track radius of $\hat{h}_o$ (Cycles/km)	Along-Track radius of $\hat{h}_o$ (Cycles/km)	Cross-Track radius of $\hat{h}_d$ (Cycles/km)	Along-Track radius of $\hat{h}_d$ (Cycles/km)	Cross-Track radius of $\hat{h}_s$ (Cycles/km)	Along-Track radius of $\hat{h}_s$ (Cycles/km)
0.0826	1.2846	1.2900	3.9873	4.0042	3.9873	3.5×10^9
0.2426	1.2071	1.2525	3.7468	3.8877	3.7468	3.5×10^9
0.3890	1.0760	1.1864	3.3398	3.6823	3.3398	3.5×10^9
0.5170	0.9230	1.1042	2.8650	3.4273	2.8650	3.5×10^9
0.6254	0.7731	1.0173	2.3995	3.1577	2.3995	3.5×10^9
0.7156	0.6392	0.9330	1.9839	2.8960	1.9839	3.5×10^9
0.7900	0.5256	0.8551	1.6314	2.6540	1.6314	3.5×10^9
0.8511	0.4319	0.7850	1.3406	2.4366	1.3406	3.5×10^9
0.9014	0.3554	0.7228	1.1030	2.2434	1.1030	3.5×10^9
0.9428	0.2931	0.6679	0.9099	2.0731	0.9099	3.5×10^9

along-track at a scan angle of 55° [2]–[4]. The GSI is equal to the GIFOV and increases in the same manner as the GIFOV.

The along-track dimension of the MODIS scan swath is 10 km at nadir (0°) but reaches approximately 20 km at the maximum scan angle (55°) at the end of each swath. Because of its geometric shape, narrow at the center and broad at the left and right, this is referred to as the bow-tie effect [2]–[4]. The along-track ground speed is 10 km per swath, so successive scans overlap off-nadir. At wide scan angles, the overlap is as much as 50% with the swath above and 50% and the swath below. The overlap in coverage for two consecutive MODIS scan swaths is depicted in Fig. 2. The two scan swaths are shown with solid and dashed lines.

In overlapping images generally, the pixels may or may not be uniformly spaced. In MODIS, pixel size and GSI is a function of scan angle, but the change is so gradual that we can assume locally uniform sampling. Consider two overlapping sets of uniformly spaced pixels, such as would be produced by successive swaths of MODIS. Referencing the coordinate system to one of the sample sets, the pixel locations are

$$x_n = \begin{cases} \Delta x \dfrac{n}{2}, & \text{if } n \bmod 2 = 0 \\ \Delta x \dfrac{n}{2} + \varepsilon, & \text{if } n \bmod 2 = 1 \end{cases} \quad (4)$$

where Δx is the interpixel spacing in both sample sets, and ε is the offset between the sampling sets. The even indices reference locations in one sample set and the odd indices reference locations in the other sample set.

Aliasing introduced by sampling can be analyzed more directly in the Fourier spatial-frequency domain. The Fourier transform or spatial-frequency spectrum $\hat{p}$ of a single (periodic) image with uniform sampling is

$$\hat{p}(v_n) = \sum_{k=-\infty}^{\infty} \hat{h}(v_n - k)\hat{s}(v_n - k) + \hat{e}(v_n) \quad (5)$$

where

v_n spatial frequency in cycles per image;
$\hat{h}$ modulation transfer function (MTF);
$\hat{s}$ spatial-frequency spectrum of the scene;
$\hat{e}$ spatial-frequency spectrum of the noise.

Assuming a periodic scene and image, the spectra coefficients v_n are discrete, uniformly spaced, and have integer indices n. Sampling causes folding of the spatial-frequency components of the attenuated scene spectrum (hence, the summation with index k). The noise spectrum is not folded because noise is modeled as a discrete process (which can model equivalently the effect of any continuous noise process).

With two overlapping images, the image transform $\hat{p}$ is

$$\hat{p}(v_n) = \sum_{k=-\infty}^{\infty} \hat{h}(v_n - k)\hat{s}(v_n - k)(1 + \exp(-i2\pi\varepsilon k)) + \hat{e}_0(v_n) + \hat{e}_\varepsilon(v_n) \quad (6)$$

where $(1 + \exp(-i2\pi\varepsilon k))$ is related to the offset between the two overlapping images, and $\hat{e}_0$ and $\hat{e}_\varepsilon$ are the spatial-frequency spectrums of the noise in the two overlapping images. The offset
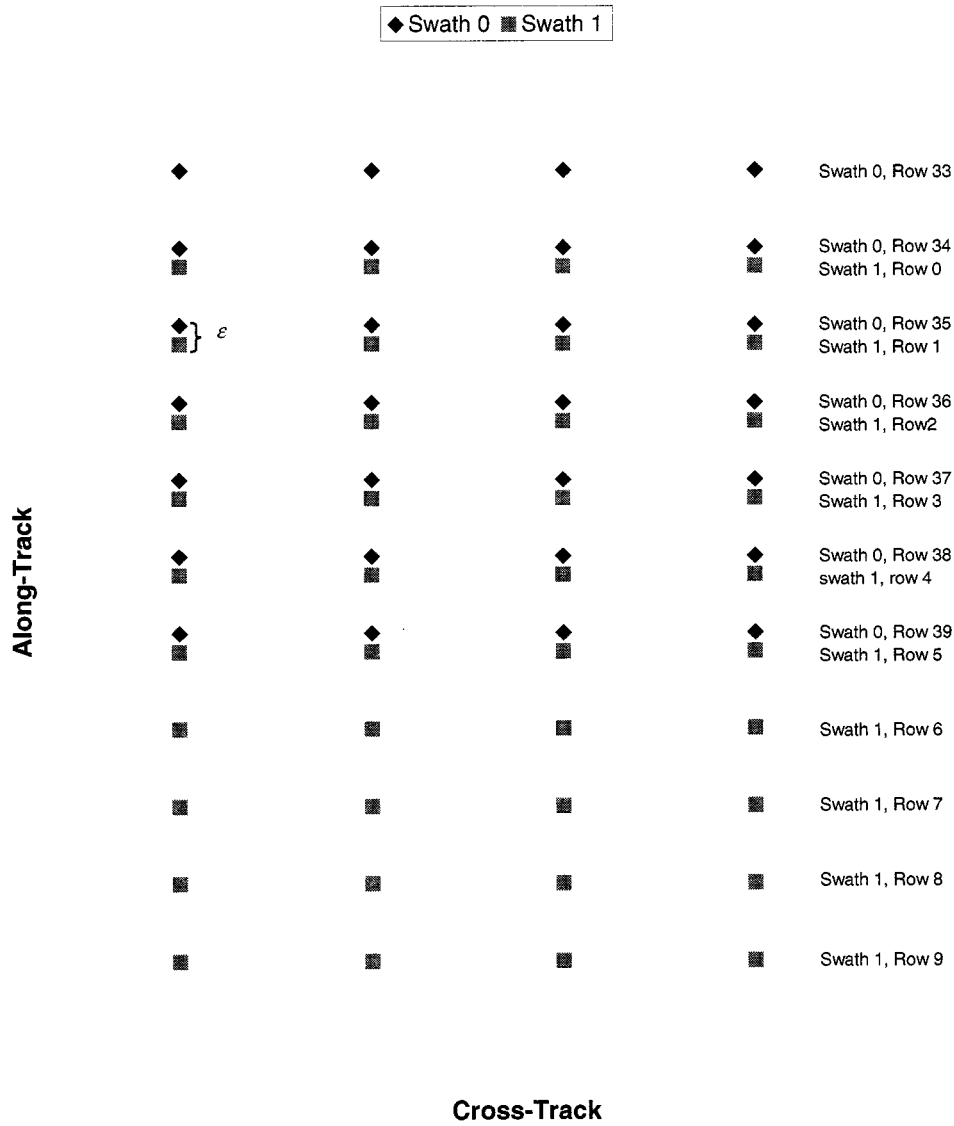
◆ Swath 0 ▦ Swath 1

Fig. 4.   Two sets of pixels from two overlapping swaths.

Fig. 5.   Simulated MODIS image (half-swath, nadir to swath edge).

(or shift) ε in the second sample set does not change the stochastic nature of the noise process. The period of the image spectrum $\hat{p}$ in (6) will be at least as large as the period of the spectrum of a single image in (5).

The frequency-domain equation for restoration and reconstruction filtering [corresponding to (3)] is

$$\hat{r}(v_n) = \hat{f}(v_n)\hat{p}(v_n). \tag{7}$$

The problem of deriving the filter is the subject of the next section.

III. DERIVATION OF OPTIMAL AND CONSTRAINED FILTERS

A. Digital Imaging System Performance

System performance is measured by how closely the output image r matches the scene s. Linfoot [5] used the expected mean-square error (MSE) of an imaging system (with stochastic scene and noise)

$$S^2 = E\left\{\int_{-\infty}^{\infty} |s(x) - r(x)|^2 \, dx\right\}$$
$$= E\left\{\int_{-\infty}^{\infty} |\hat{s}(v) - \hat{r}(v)|^2 \, dv\right\} \tag{8}$$

to define image fidelity as

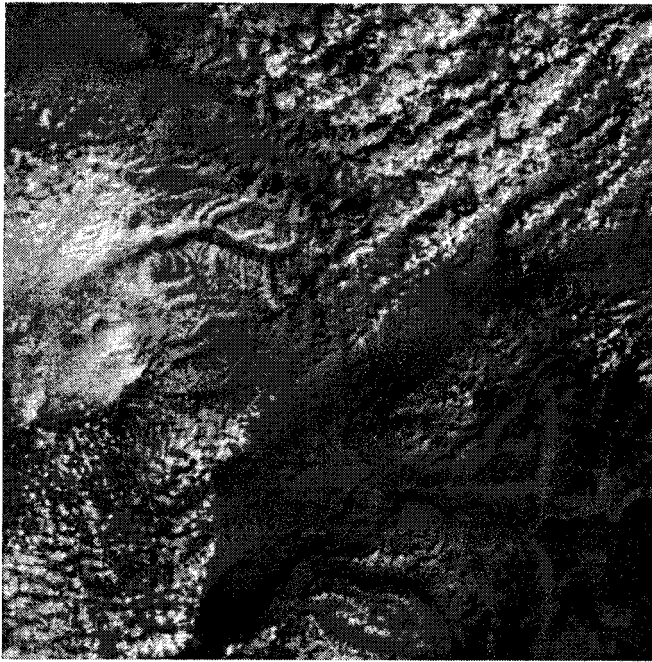$$F = 1 - \frac{S^2}{\sigma_s^2} \tag{9}$$

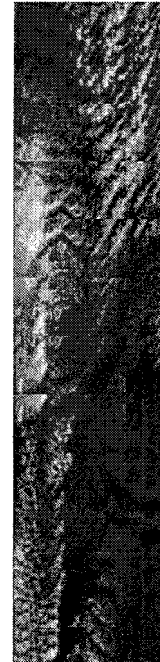Fig. 6. Simulated MODIS image tile with scan angle 0.0826.

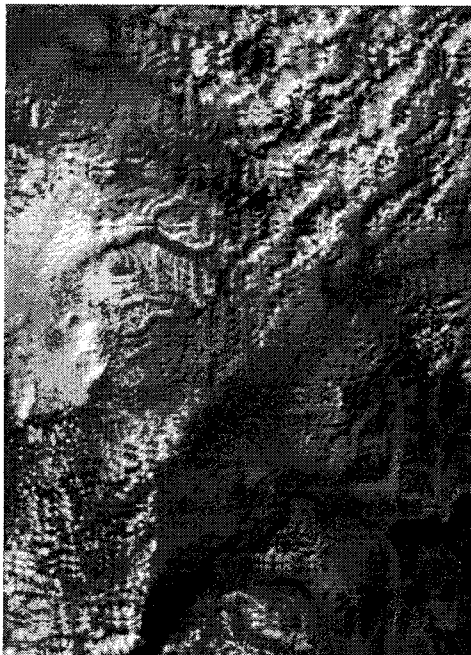Fig. 8. Simulated MODIS image tile with scan angle 0.9428.

Fig. 7. Simulated MODIS image tile with scan angle 0.5170.

where σ_s^2 is the expected (ensemble average) variance of the scene radiance field

$$\sigma_s^2 = E\left\{\int_{-\infty}^{\infty} |s(x)|^2 \, dx\right\} = E\left\{\int_{-\infty}^{\infty} |\hat{s}(v)|^2 \, dv\right\}. \quad (10)$$

For notational simplicity, equations in this paper assume the scene radiance field is a zero-mean process. In practice, the mean can be accounted for during filtering. Fidelity is bounded by 1 with $F = 1$, if and only if the output image is identical to the scene radiance field. MSE metrics such as fidelity facilitate mathematical analyses but do not correspond directly to subjective visual quality. However, a more definitive objective measure of image quality has proven elusive.

B. Derivation of Optimal Unconstrained Filter

The derivation of the CDC Wiener filter is conditioned on the following assumptions.

- The scene power spectra $\hat{\Phi}_s$ and noise power spectra $\hat{\Phi}_e$ are known.
- The sideband components of the scene spectrum that alias to the same frequency are uncorrelated.
- The noise is random and the scene and noise stochastic processes are uncorrelated.

Mathematically, these assumptions are

$$E\{|\hat{e}(v_n)|^2\} = \hat{\Phi}_e(v_n) \quad (11)$$

$$E\{\hat{s}(v_n)\hat{s}^*(v_n - k)\} = \begin{cases} \hat{\Phi}_s(v_n), & \text{if } k = 0 \\ 0, & \text{if } k = \pm 1, \pm 2 \cdots \end{cases} \quad (12)$$

$$E\{\hat{s}(v_n - k)\hat{e}^*(v_n)\} = E\{\hat{s}^*(v_n - k)\hat{e}(v_n)\} = 0 \quad (13)$$

where the "$*$" superscript denotes complex conjugation.

The filter design criterion is to minimize MSE (or equivalently to maximize fidelity) of the system. From (6)–(8) and (11)–(13), it can be shown that the expected MSE after filtering (8) is

$$S^2 = \sigma_s^2 - 2\sigma_{s,r} + \sigma_r^2 \quad (14)$$

where σ_r^2 is the expected variance of the output image, and $\sigma_{s,r}$ is the expected covariance of the scene radiance field and image. These terms can be expressed as

$$\sigma_s^2 = \sum_{n=-\infty}^{\infty} \hat{\Phi}_s(v_n) \quad (15)$$

TABLE III
EXPECTED RMSE AND FIDELITY FOR THE CDC WIENER FILTER

| View angle (Radians)	Expected RMS Error	Expected Fidelity	Expected Blurring RMS Error	Expected Sampling RMS Error	Expected Noise RMS Error
0.0826	19.4092	0.9001	18.1504	5.4680	4.1691
0.2426	19.5163	0.8990	18.3776	5.7764	3.1281
0.3890	20.3424	0.8903	19.1794	6.0085	3.1404
0.5170	21.2818	0.8800	20.1100	6.1272	3.3114
0.6254	22.3762	0.8673	21.1918	6.1669	3.6840
0.7156	24.8060	0.8369	23.5004	7.2984	3.1315
0.7900	25.6401	0.8257	24.3151	6.5424	4.8361
0.8511	28.2908	0.7879	26.9125	7.9023	3.6927
0.9014	30.5687	0.7523	29.1246	8.5198	3.6898
0.9428	33.1746	0.7083	31.6567	9.3416	3.3379
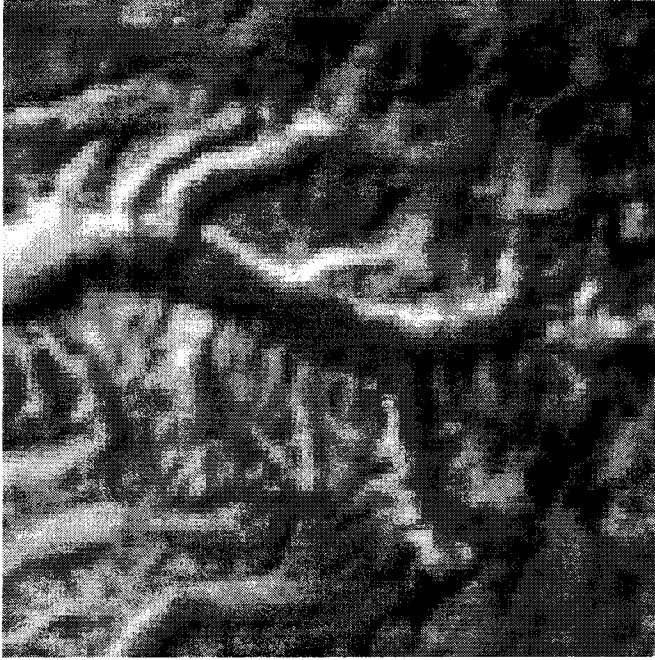
Fig. 9.   Nearest-neighbor reconstruction: resampled MODIS image tile with scan angle 0.0826.
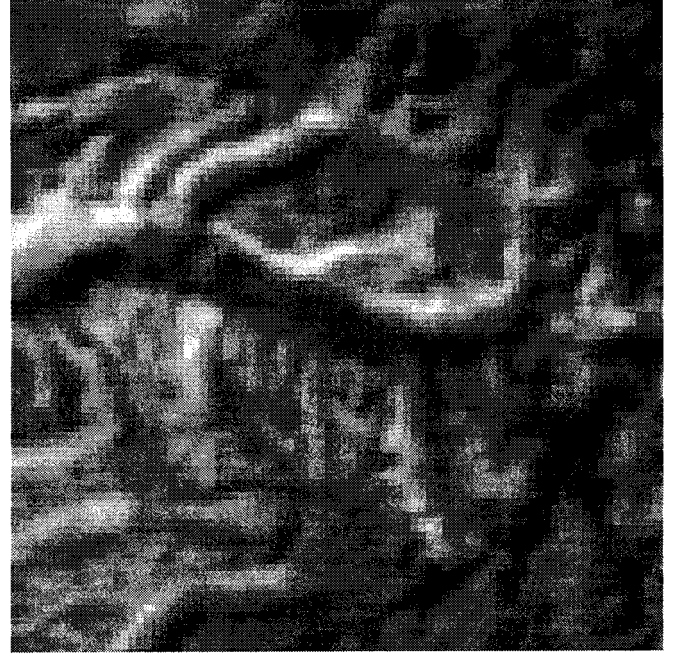

Fig. 10.   Nearest-neighbor reconstruction: resampled MODIS image tile with scan angle 0.5170.

$$\sigma_{s,r} = \sum_{n=-\infty}^{\infty} \hat{\Phi}_{s,p}(v_n)\hat{f}^*(v_n) \tag{16}$$

$$\sigma_r^2 = \sum_{n=-\infty}^{\infty} \hat{\Phi}_p(v_n)|\hat{f}(v_n)|^2 \tag{17}$$

where

$\hat{\Phi}_s$ power spectrum of the scene;
$\hat{\Phi}_{s,p}$ cross-power spectrum of the scene and image;
$\hat{\Phi}_p$ power spectrum of the image.

$$\hat{\Phi}_s(v_n) = E\{|\hat{s}(v_n)|^2\} \tag{18}$$

$$\hat{\Phi}_{s,p}(v_n) = E\{\hat{s}(v_n)\hat{p}^*(v_n)\} = 2\hat{\Phi}_s(v_n)\hat{h}^*(v_n) \tag{19}$$

$$\hat{\Phi}_p(v_n) = E\{|\hat{p}(v_n)|^2\}$$
$$= 2\sum_{k=-\infty}^{\infty} \hat{\Phi}_s(v_n - k)|\hat{h}(v_n - k)|^2(1 + \cos(2\pi\varepsilon k))$$
$$+ 2\hat{\Phi}_e(v_n). \tag{20}$$

The coefficients "2" in (19) and (20) reflect the overlap of two images. Because all of the spatial functions are real, the frequency-domain transforms are Hermitian and the power spectra $\hat{\Phi}_s$, $\hat{\Phi}_p$, and $\hat{\Phi}_e$ are real, nonnegative, and even.

The expression for expected MSE can be regrouped to make clear the tradeoff between blurring and aliasing in system design

$$S^2 = \sum_{n=-\infty}^{\infty} \hat{\Phi}_s(v_n)|1 - 2\hat{h}(v_n)\hat{f}(v_n)|^2 + 2\sum_{n=-\infty}^{\infty}$$
$$\cdot \left(\sum_{k\neq 0} \hat{\Phi}_s(v_n - k)|\hat{h}(v_n - k)|^2(1 + \cos(2\pi\varepsilon k))\right)$$
$$\cdot |\hat{f}(v_n)|^2 + 2\sum_{n=-\infty}^{\infty} \hat{\Phi}_e(v_n)|\hat{f}(v_n)|^2. \tag{21}$$

The first term represents the error associated with blurring the image by both the acquisition PSF h and the reconstruction PSF f. To minimize this term, $\hat{f}$ should be equal to one-half the
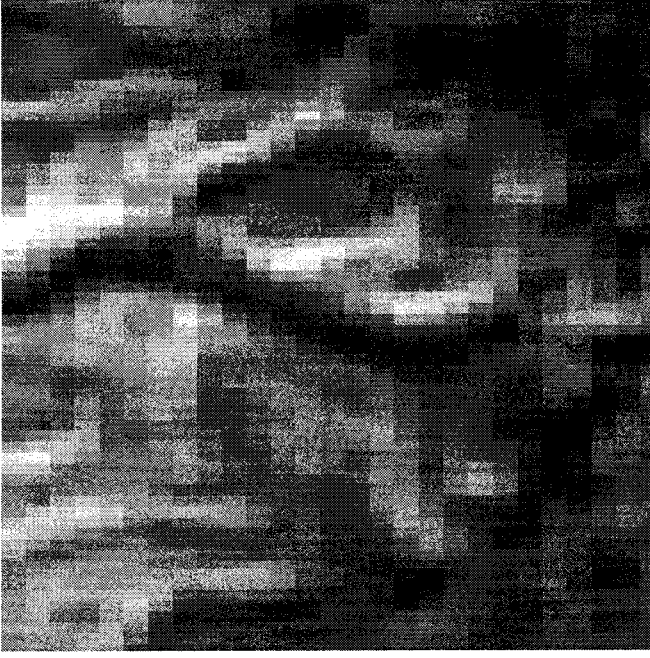
Fig. 11. Nearest-neighbor reconstruction: resampled MODIS image tile with scan angle 0.9428.

reciprocal of $\hat{h}$ at all frequencies. The second term represents the error associated with the aliasing. To minimize this term, $\hat{f}$ should be equal to zero at all frequencies. The final term is associated with system noise. To minimize this term, $\hat{f}$ should be equal to zero at all frequencies. So there is clearly a tradeoff between blurring, aliasing, and noise.

With no restrictions on the spatial support of the filter, the MSE is minimized when the filter transfer function is [6]

$$\hat{f}_w = \frac{\hat{\Phi}_{s,p}(v_n)}{\hat{\Phi}_p(v_n)}. \tag{22}$$

We refer to this filter as the CDC Wiener filter because it is based on the CDC model. Acquisition converts from a continuous scene to a digital image, and the filter converts from a digital image to a continuous result.

The CDC Wiener filter cannot be implemented practically via spatial convolution because it is continuous and its support is the extent of the FOV. Moreover, it has been assumed to this point that both sample sets are fully populated, but as with the overlapping swaths of the MODIS instrument, the overlap may occur in only part of the image. To produce a filter that is both more efficient and weights only available sample values, we impose support constraints on the filter, limiting the filter to weight a subset of the image values during convolution prior to the derivation.

C. Derivation of Optimal Constrained Filter

The derivation of the optimal, constrained filter f_c is conditioned on constraints imposed on its spatial support. The support of the kernel is a nonempty set of spatial locations C, for which filter values can be nonzero. Except for locations in the support set, the filter value is 0

$$f_c(x) = 0 \qquad \text{if } x \notin C. \tag{23}$$

The optimal, spatially constrained kernel is derived by minimizing the MSE S^2 with respect to the elements in the support set. Minimization of S^2 with respect to the kernel elements requires that

$$\frac{\partial S^2}{\partial f_c(x)} = 0 \qquad x \in C. \tag{24}$$

The transfer function of the spatially constrained kernel is

$$\hat{f}_c(v_n) = \sum_{x \in C} f_c(x) \exp(-i2\pi v_n x). \tag{25}$$

After substituting this expression for $\hat{f}$ into (14) [with (15)–(20)], it can be shown [7] that minimization of S^2 with respect to the kernel elements requires

$$\sum_{x' \in C} \Phi_p(x - x') f_c(x') = \Phi_{s,p}(x) \qquad x \in C \tag{26}$$

where Φ_p is the auto-covariance of the image, and $\Phi_{s,p}$ is the cross-covariance of the scene and image (again, assuming zero-mean processes). The number of unknowns in (26) is equal to the number of elements in the support set of the kernel. There are $|C|$ equations in $|C|$ unknowns. Solving for f_c yields the optimal constrained filter.

D. Filter Performance

Equation (14), with the expressions in (15)–(20), can be used to compute the expected MSE from $\hat{\Phi}_s$, $\hat{\Phi}_e$, $\hat{h}$, and $\hat{f}$. For the CDC Wiener filter $\hat{f}_w$ in (22)

$$S_w^2 = \sum_{n=-\infty}^{\infty} \hat{\Phi}_s(v_n) - \frac{|\hat{\Phi}_{s,p}(v_n)|^2}{\hat{\Phi}_p(v_n)}. \tag{27}$$

Therefore, the fidelity (9) for the CDC Wiener filter is

$$F_w = \frac{1}{\sigma_s^2} \sum_{n=-\infty}^{\infty} \frac{|\hat{\Phi}_{s,p}(v_n)|^2}{\hat{\Phi}_p(v_n)}. \tag{28}$$

No filter can restore with higher fidelity (smaller MSE) than F_w in (28). However, for typical imaging systems, a small filter with a few centrally located elements can perform nearly as well [7]. For any filter f, including the constrained filter f_c, the expected MSE is

$$S^2 = S_w^2 + \sum_{n=-\infty}^{\infty} \hat{\Phi}_p(v_n) |\hat{f}(v_n) - \hat{f}_w(v_n)|^2. \tag{29}$$

As can be seen in (29), $S^2 \geq S_w^2$ and F_w is the upper bound on fidelity.

IV. EXPERIMENTAL RESULTS

A. Imaging System Simulation

To illustrate the bow-tie effect described in Section II-B and to demonstrate resampling and restoration, we have constructed a simulation of a wide FOV scanning imaging system similar to MODIS. The simulation uses as its input scene a Landsat TM image (Band 5) with GSI of 28.5 m, a portion of which is pictured in Fig. 3. Following the imaging system model illustrated in Fig. 1, the Landsat TM scene is blurred, subsampled, and cor-

(-1.3086, 0.6829)	(-0.3086, 0.6829)	(0.6914, 0.6829)	(1.6914, 0.6829)
(-1.3086, 0.4390)	(-0.3086, 0.4390)	(0.6914, 0.4390)	(1.6914, 0.4390)
(-1.3086, -0.3171)	(-0.3086, -0.3171)	(0.6914, -0.3171)	(1.6914, -0.3171)
(-1.3086, -0.5610)	(-0.3086, -0.5610)	(0.6914, -0.5610)	(1.6914, -0.5610)

0.1842	0.5788	0.3500	0.0830
-0.3510	-0.1662	-0.3542	-0.1417
-0.1927	0.7250	0.1080	-0.0490
0.0290	0.1673	0.0497	0.0208

(a)

(-1.3086, 0.3171)	(-0.3086, 0.3171)	(0.6914, 0.3171)	(1.6914, 0.3171)
(-1.3086, 0.0732)	(-0.3086, 0.0732)	(0.6914, 0.0732)	(1.6914, 0.0732)
(-1.3086, -0.9268)	(-0.3086, -0.9268)	(0.6914, -0.9268)	(1.6914, -0.9268)
(-1.3086, -1.9268)	(-0.3086, -1.9268)	(0.6914, -1.9268)	(1.6914, -1.9268)

0.0254	-0.4846	-0.2364	0.0511
-0.3228	1.7120	0.3895	-0.1152
0.0024	-0.0375	-0.0185	-0.0015
0.0159	-0.0087	0.0065	0.0027

(b)

(-1.3086, 1.0244)	(-0.3086, 1.0244)	(0.6914, 1.0244)	(1.6914, 1.0244)
(-1.3086, 0.0244)	(-0.3086, 0.0244)	(0.6914, 0.0244)	(1.6914, 0.0244)
(-1.3086, -0.9756)	(-0.3086, -0.9756)	(0.6914, -0.9756)	(1.6914, -0.9756)
(-1.3086, -1.9756)	(-0.3086, -1.9756)	(0.6914, -1.9756)	(1.6914, -1.9756)

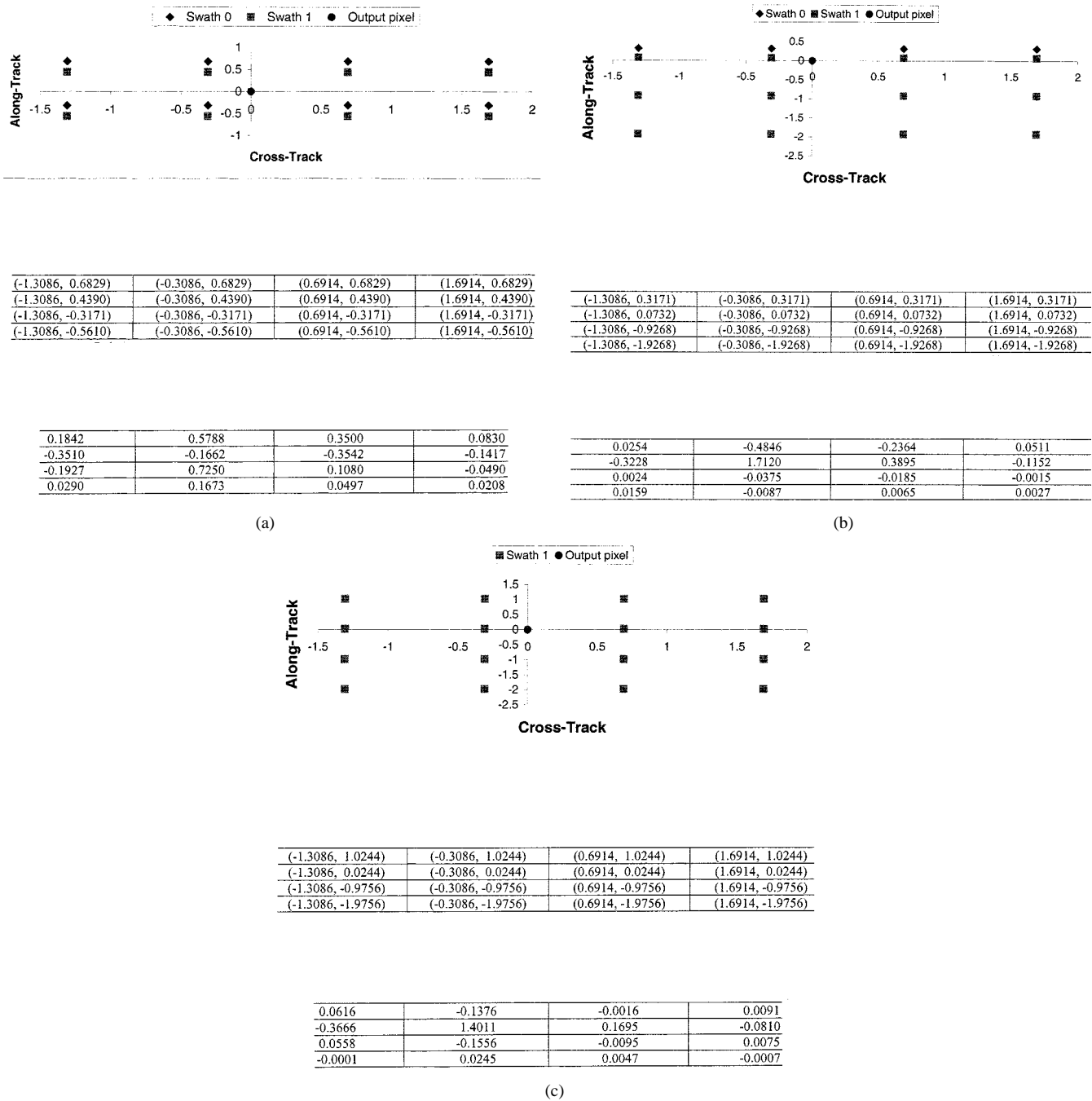| 0.0616	-0.1376	-0.0016	0.0091
-0.3666	1.4011	0.1695	-0.0810
0.0558	-0.1556	-0.0095	0.0075
-0.0001	0.0245	0.0047	-0.0007

(c)

Fig. 12. Restoration kernels and spatial locations in the corresponding support set.

rupted by noise. In this simulation, the GSI is 250 m at nadir with 40 pixel swaths (as in MODIS Bands 1–2).

In order to facilitate comparison of the effects at different scan angles and because the ground projection of the MODIS scan is much wider than the Landsat TM scene, the Landsat TM image is tiled repeatedly edge-to-edge to produce a scene as wide as the half-width of a MODIS scan (i.e., nadir to one end of the scan). The 4096 × 4096 Landsat TM image represents a 116.736 km × 116.736 km ground area. The half-width of a MODIS scan is 10 km along-track at nadir and 1165 km cross-track [2]–[4]. Thus, the size of the ten-tile TM image 116.736 km × 1167.36 km is approximately 12 swaths of a half-width MODIS scan.

To simplify processing, we use a constant scan angle for each tile in the simulation. The left-most tile has a scan angle of 0.0826 radians (set according to the center of the tile). This tile is used to approximately simulate the MODIS image near nadir. The right-most tile with scan angle of 0.9428 radians is used to approximate MODIS at the largest scan angle. At a larger scan angle, the GSI of MODIS is larger. Table I lists the scan angle and GSI (cross-track × along-track) for each tile. The ratios of the off-nadir GSI to at-nadir and cross-track GSI also are listed in Table I, where $wfac$ and $hfac$ denote the cross-track ratio and along-track ratio, respectively.

In the first step of the simulation, the Landsat TM scene tiles are blurred. It is more efficient to compute the blurred image

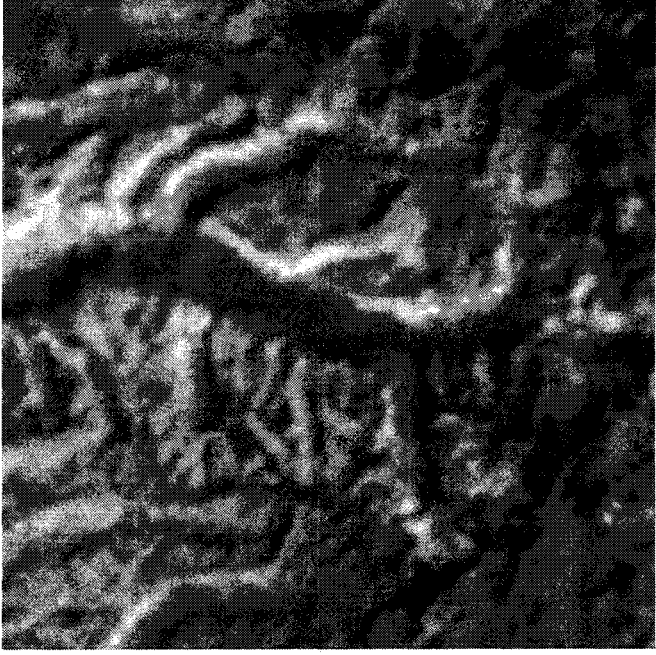Fig. 13. Restoration and reconstruction: resampled MODIS image (half-swath, nadir to swath edge).

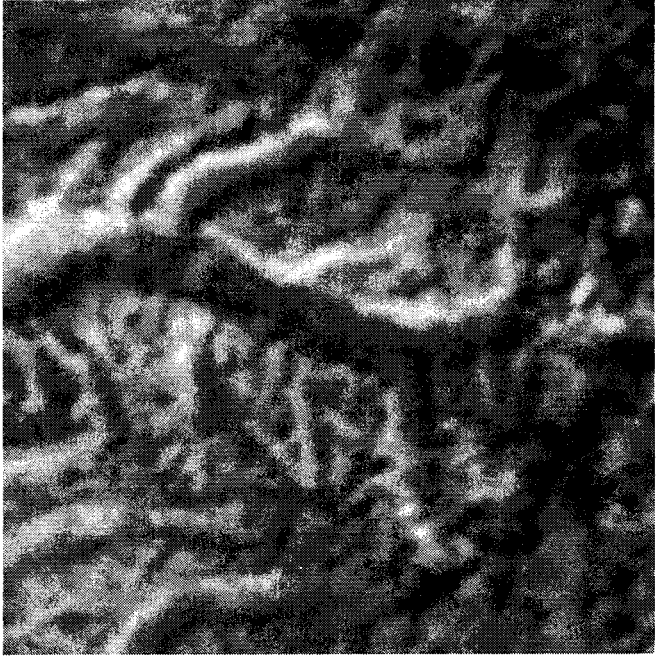Fig. 14. Restoration and reconstruction: resampled MODIS image tile with scan angle 0.0826.

Fig. 15. Restoration and reconstruction: resampled MODIS image tile with scan angle 0.5170.

in the spatial-frequency domain, where the scene spectrum $\hat{s}$ is multiplied by $\hat{h}_o$, $\hat{h}_s$, and $\hat{h}_d$. The Gaussian function is a convenient, commonly used model for the optics transfer function $\hat{h}_o$. Sinc functions are used to simulate the scanner and the detector

transfer functions $\hat{h}_s$ and $\hat{h}_d$. At larger scan angles, the GIFOV of the detector is larger, and the scene is more blurred during acquisition. In order to simulate this increasing blur, different acquisition transfer functions are applied to the scene tiles at different scan angles. The equation of the Gaussian function is

$$\hat{h}_o(v_n) = \exp\left(\frac{-v_n^2}{2\sigma^2}\right) \qquad (30)$$

where the radius σ is the standard deviation of the Gaussian MTF and is related to GSI as

$$\sigma = \frac{1}{2\pi \times 0.494 \times \text{GSI}}. \qquad (31)$$

The value 0.494 is selected to give a moderate response of 0.3 at the Nyquist limit. In the spatial-frequency domain, the radius of the Gaussian MTF $\hat{h}_o$ decreases as a function of $wfac$ and $hfac$ as the scan angle increases. Table II lists the radius of the Gaussian MTF in cycles per kilometer for each tile. The equation of the sinc function is

$$\hat{h}_d(v_n) = \frac{\sin(\text{GSI} \times \pi \times v_n)}{\text{GSI} \times \pi \times v_n}. \qquad (32)$$

The radius of the first zero of the detector MTF $\hat{h}_d$ also decreases as the scan angle increases in both cross-track and along-track directions. The scanner MTF $\hat{h}_s$ is the same as the detector in the cross-track dimension. In the along-track dimension, the radius of the first zero of $\hat{h}_s$ is very large in all tiles because the scanning due to orbital motion is small during the temporal integration of a pixel. Table II also lists the radius of the first zero for the sinc function.

In the second step of the simulation, the blurred scene tiles are scanned and subsampled. Twelve swaths are extracted from the blurred scene to simulate twelve scans. The along-track centers of the swaths are the same for different tiles, but the height of each swath increases as scan angle increases. Off-nadir, successive swaths overlap. Note that, at larger scan angles, the GSI is larger and so fewer MODIS pixels are sampled in the same ground area. The size of the subsampled tile at 0.5170 radians is 334×463 (cross-track $\times$ along-track). The size of the subsampled tile at the largest scan angle is 106×454. Because 4096 scene pixels do not yield exactly twelve swaths, the along-track size of the subsampled image is slightly different from tile to tile. Table I lists the sizes of subsampled tiles at different scan angles.

Fig. 4 illustrates two sets of subsampled pixels from the two overlapping swaths. The offset between the two sampling sets ε is a constant value inside the same tile (due to the assumption that scan angle is constant within a tile). Between different tiles, ε varies.

In the third and final step of the simulation, random noise (zero-mean, Gaussian noise with variance = 8.5549) is added to the subsampled image in order to achieve SNR of 21 (in the typical range for many imaging systems). SNR is defined as

$$\text{SNR} = \frac{\sigma_s}{\sigma_e}. \qquad (33)$$

The simulated image produced by this process is illustrated in Fig. 5. The left edge of the image is at nadir and right edge
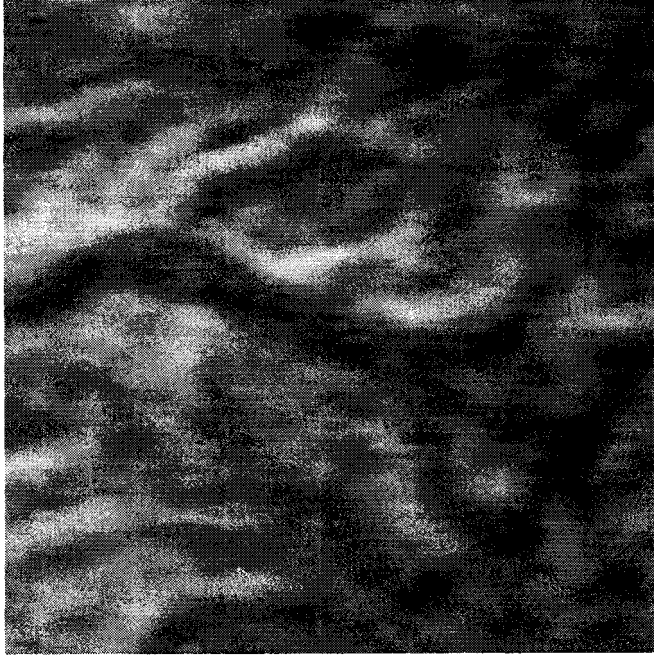
Fig. 16. Restoration and reconstruction: resampled MODIS image tile with scan angle 0.9428.

is at the extreme scan angle. The left-most tile, with its center at a scan-angle of 0.0826 radians, is enlarged in Fig. 6. A tile from the middle of the half-scan, with its center at a scan-angle of 0.5170 radians, is enlarged in Fig. 7. The right-most tile, with its center at scan angle of 0.9428 radians, is enlarged in Fig. 8. The bow-tie effect is a dominant artifact in the portion of the tile pictured in Fig. 8 and is clearly visible in Fig. 7. In these images, the delineation between swaths is clear and overlap makes it difficult to discern spatial structures in the scene.

B. Scene Model and Expected Performance for CDC Wiener Filter

In the derivation in Section III-B, the scene power spectra $\hat{\hat{\Phi}}_s$ must be known. In this paper, we use a 2-D Markov random field scene model. We assume the mean spatial detail (in pixel units) denoted by μ is the same along-track and cross-track and the power spectrum of the stochastic scene is [8]

$$\hat{\Phi}_s(u_n, v_n) = \frac{2\pi \dfrac{\mu^2}{wfac \times hfac} \sigma_s^2}{\left(1 + 4\pi^2\mu^2\left(\left(\dfrac{u_n}{wfac}\right)^2 + \left(\dfrac{v_n}{hfac}\right)^2\right)\right)^{3/2}} \tag{34}$$

where $wfac$ and $hfac$ are pixel size ratios along-track and cross-track, respectively, as described in Table I.

Table III gives the expected RMS error (RMSE) and fidelity for the CDC Wiener filter (27) and (28) based on this scene model with $\mu = 3$. The value of μ is chosen to roughly match the characteristics of the actual Landsat TM scene used in the simulation. Table III also gives the expected RMSE from blurring, sampling, and noise after filtering (21) as described in Section III-B.

C. Experimental Results for Optimal Constrained Filters

Resampling the overlapping scans on a uniform rectangular raster eliminates the troublesome bow-tie effect. The MODIS land data storage approach, Level 2 Grid (L2G), stores multiple observations and allows Level 3 gridded products based on multiple observations from different orbits and resulting from the bowtie effect [4]. The MODIS L2G structure supports restoration and resampling.

One of the simplest algorithms for resampling is to take the value of the nearest pixel to the resampling point. This reconstruction method is called *nearest-neighbor interpolation*. Nearest-neighbor interpolation is simple and easily computed but produces images with visible blockiness and low fidelity. Nearest-neighbor interpolation reconstructs but does not restore. Figs. 9–11 are the portions of tiles with centers at scan-angles 0.0826, 0.5170, and 0.9428 radians, respectively, resampled by nearest neighbor interpolation. The artifacts of nearest-neighbor interpolation are clearest in Fig. 11. Table IV presents the actual RMS errors for the 4096 × 4096 nearest-neighbor reconstructed tiles.

More sophisticated reconstruction techniques, such as linear interpolation or cubic convolution, yield better results than nearest-neighbor interpolation, and techniques that restore as well as reconstruct can yield much better fidelity [1], [9], [10]. Our technique described in this paper not only reconstructs but also restores from overlapping images.

For each tile, the optimal restoration and reconstruction kernel determined by (26) weights the 4 × 4 nearest pixels. The weights of the optimal kernel depend on the spatial locations in the support set. The spatial locations in the support set for different output pixels may be different because input pixels in the overlapping image may not be uniformly spaced. The 4 × 4 nearest pixels may come from one swath (if there is no overlap in the area) or come from two swaths (if there is overlap). Fig. 4 shows some input pixels from two overlapping swaths of the tile at 0.5170 radians. Fig. 12 illustrates three kernels for three different output locations in Fig. 4. They have the same across-track locations. In Fig. 12(a), the output pixel is between Swath 1 Row 2 and Swath 0 Row 37. Its value is determined by the nearest pixels from two overlapping swaths. In Fig. 12(b), the output pixel is between Swath 1 Row 5 and Swath 1 Row 6. Its value is determined by the pixels from two partially overlapping swaths. In Fig. 12(c), the output pixel is between Swath 1 Row 7 and Swath 1 Row 8. Its value is determined by the pixels from one swath. In these figures, the coordinates are relative to the output pixel location and normalized to the GSI.

The kernel weights are influenced both by the distance to the pixel and by the presence of other pixels. For example, in Fig. 12(a), the pixel at $(-0.3086, -0.3171)$ is the nearest of the 16 weighted pixels and so has the largest positive weight 0.7250 in the kernel. In Fig. 12(b), the pixel at $(-0.3086, 0.3171)$ has the same distance, but another pixel at $(-0.3086, 0.0732)$ is much closer in the same general direction. So in the example of Fig. 12(b), the closer pixel has the largest positive weight 1.7120, and the pixel at $(-0.3086, 0.3171)$ has a negative

TABLE IV
ACTUAL RMSE AND FIDELITY FOR NEAREST-NEIGHBOR RECONSTRUCTION AND FOR RESTORATION AND RECONSTRUCTION

| View angle (Radians)	RMS Error (Nearest-Neighbor)	Fidelity (Nearest-Neighbor)	RMS Error (Restoration & Reconstruction)	Fidelity (Restoration & Reconstruction)
0.0826	24.9006	0.8357	19.9008	0.8950
0.2426	26.7903	0.8098	20.3627	0.8901
0.3890	27.5015	0.7995	21.4217	0.8784
0.5170	28.4765	0.7851	22.5139	0.8656
0.6254	29.5014	0.7693	23.9047	0.8485
0.7156	31.3717	0.7391	25.9771	0.8211
0.7900	31.9900	0.7287	27.2762	0.8028
0.8511	33.4045	0.7042	29.0693	0.7760
0.9014	34.6324	0.6821	30.8708	0.7474
0.9428	36.1878	0.6529	32.6182	0.7180
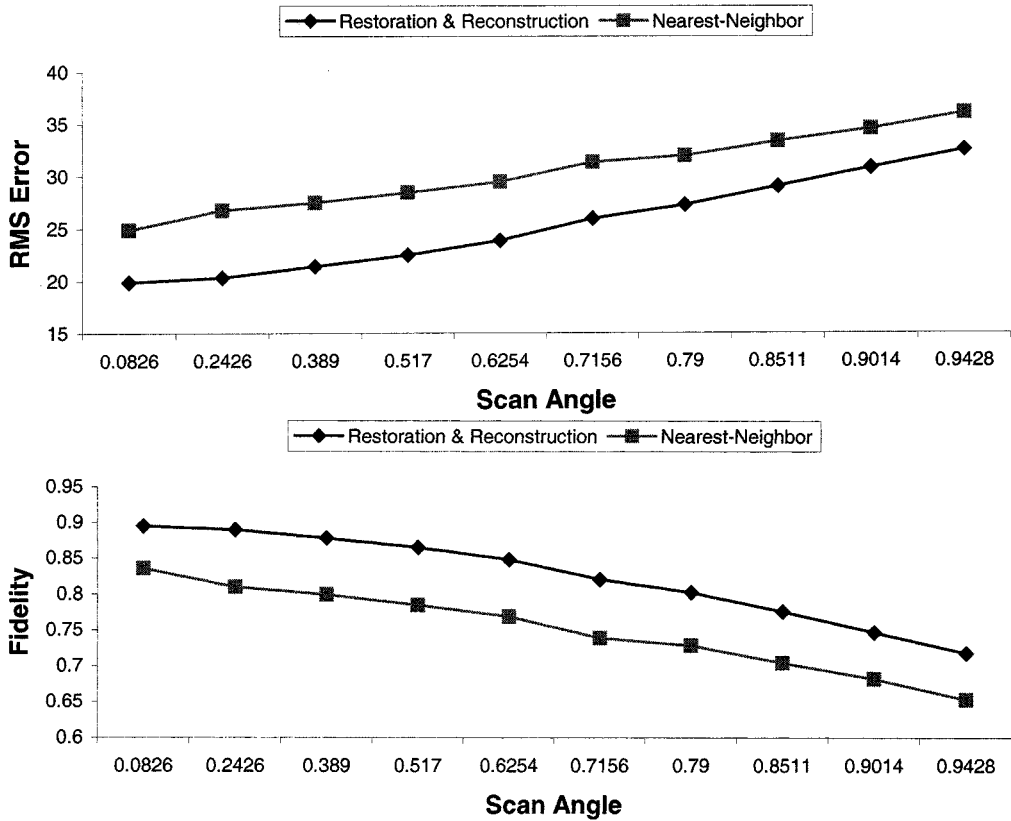
Fig. 17.   RMSE and fidelity for nearest-neighbor reconstruction and for restoration and reconstruction.

weight −0.4846, effectively extending the difference between the pixels' values to sharpen the restored image.

Fig. 13 illustrates the result of restoring and reconstructing the simulated image in Fig. 5. Each tile of the image is resampled to the size of original scene with pixel values to be determined at 4096 × 4096 locations. Figs. 14–16 are the portions of restored and reconstructed tiles with centers at scan-angles 0.0826, 0.5170, and 0.9428 radians, respectively. The visual artifacts related to the bow-tie effect in Figs. 6–8 are removed and the visual quality is better than in Figs. 9–11. Table IV compares the results of our restoration and reconstruction technique to the results of nearest-neighbor interpolation. Fig. 17 compares the RMSE and the fidelity. Our method for restoration and reconstruction achieves much better fidelity than nearest-neighbor interpolation.

V. CONCLUSION

Based on a comprehensive end-to-end system model that accounts for overlapping images, the one-pass restoration and reconstruction technique developed in this paper yields high fidelity and mean-square optimal resampling. The restoration and reconstruction filter is derived subject to user-defined and data availability constraints on spatial support for spatial domain processing.

The restoration and reconstruction technique in this paper is described in terms of its applicability to overlapping swaths in images from a wide angle scanning imager such as MODIS. The approach has broader applicability to multi-image fusion [11], the more general problem of combining images from multiple data sources to form a single image. Other multi-image

fusion problems may be complicated by sources with different imaging geometries, spatial resolution, spectral response, time of acquisition, or other variables. With respect to these issues, the problem of resampling MODIS images on a uniform grid is a fairly straightforward problem but provides a useful arena for developing, demonstrating, and evaluating various approaches.

REFERENCES

[1] J. F. Moreno and J. Melia, "An optimum interpolation method applied to the resampling of NOAA AVHRR data," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 131–151, Jan. 1994.

[2] MODIS Science Data Support Team, "MODIS Level 1A Earth location: Algorithm theoretical basis document Version 3.0," http://modarch.gsfc.nasa.gov/MODIS/ATBD/atbd_mod28_v3.pdf.

[3] J. L. Barker *et al.*, "MODIS Level 1 geolocation, characterization and calibration algorithm theoretical basis document," Tech. Rep. NASA TM 104594 MODIS, vol. 2, May 1994.

[4] R. E. Wolfe, D. P. Roy, and E. Vermote, "MODIS land data storage, gridding and compositing methodology: Level 2 grid," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 1324–1338, July 1998.

[5] E. H. Linfoot, "Transmission factors and optical design," *J. Opt. Soc. Amer.*, vol. 46, no. 9, pp. 740–752, 1956.

[6] F. O. Huck, C. L. Fales, N. Halyo, R. W. Samms, and K. Stacy, "Image gathering and processing: Information and fidelity," *J. Opt. Soc. Amer. A*, vol. 2, no. 10, pp. 1644–1666, 1985.

[7] S. E. Reichenbach and S. K. Park, "Small convolution kernels for high-fidelity image restoration," *IEEE Trans. Signal Processing*, vol. 39, pp. 2263–2274, Oct. 1991.

[8] S. E. Reichenbach, S. K. Park, R. Alter-Gartenberg, and Z. Rahman, "Artificial scenes and simulated imaging," *Proc. SPIE*, vol. 1569, pp. 422–433, 1991.

[9] S. E. Reichenbach, D. E. Koehler, and D. W. Strelow, "Restoration and reconstruction of AVHRR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 997–1007, July 1995.

[10] S. E. Reichenbach and K. Haake, "Cubic convolution for one-pass restoration and resampling," in *Int. Symp. Geoscience and Remote Sensing*. Piscataway, NJ: IEEE, 1996, pp. 1597–1599.

[11] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*. San Diego, CA: Academic, 1997.

Stephen E. Reichenbach received the B.S. degree from the University of Nebraska, Lincoln (UNL), the M.Sc. degree in computer science from Washington University, St. Louis, MO, and the Ph.D. degree in computer science from the College of William and Mary, Williamsburg, VA.

He is an Associate Professor, Computer Science and Engineering Department (CSE), UNL. From 1996 to 2000, he served as CSE Department Chair, UNL. His research interests include digital image processing and image information systems.

Dr. Reichenbach held a National Research Council Research Associateship in the Visual Information Processing Laboratory, NASA Langley Research Center, Langley, VA, and an ASEE Research Fellowship in the Landsat 7 Project Science Office.

Jing Li received the B.A. degree in computer science from the Chang Chun Institute of Optics and Fine Mechanics, China, in 1994, and the M.S. degree in computer science from the University of Nebraska, Lincoln, in 1998.