Pixel-level multisensor image fusion based on matrix completion and robust principal component analysis

Zhuozheng Wang
J. R. Deller, Jr.
Blair D. Fleet

Pixel-level multisensor image fusion based on matrix completion and robust principal component analysis

Zhuozheng Wang,[a,*] **J. R. Deller Jr.,**[b] **and Blair D. Fleet**[b]
[a]Beijing University of Technology, Department of Electronic and Information Engineering, Beijing 100124, China
[b]Michigan State University, Department of Electrical and Computer Engineering, 428 South Shaw Lane, Room 2120, East Lansing, Michigan 48824-1226, United States

Abstract. Acquired digital images are often corrupted by a lack of camera focus, faulty illumination, or missing data. An algorithm is presented for fusion of multiple corrupted images of a scene using the lifting wavelet transform. The method employs adaptive fusion arithmetic based on matrix completion and self-adaptive regional variance estimation. Characteristics of the wavelet coefficients are used to adaptively select fusion rules. Robust principal component analysis is applied to low-frequency image components, and regional variance estimation is applied to high-frequency components. Experiments reveal that the method is effective for multifocus, visible-light, and infrared image fusion. Compared with traditional algorithms, the new algorithm not only increases the amount of preserved information and clarity but also improves robustness. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JEI.25.1.013007]

1 Introduction

Multisensor image fusion is the process of combining two or more images of a scene to create a single image that is more informative than any of the input images.[1] Image-fusion technology is employed in numerous applications including visual interpretation, image drawing, geographical information gathering, and military target reconnaissance and surveillance. In particular, research into techniques for image fusion by contrast reversal in local image regions has important theoretical and practical significance.[1]

Image-fusion methods are classified as spatial- or transform-domain techniques. Spatial-domain methods are simple, but generally result in images with insufficient detail. Transform-domain strategies based on image-fusion arithmetic and wavelet transformations (WTs) represent the current state of the art. Wavelets can be used to resolve an original image into a series of subimages with different spatial resolutions and frequency-domain characteristics. This representation fully reflects local variations in the original image. In addition, WTs can affect multiresolution analysis,[2,3] perfect refactoring, as well as orthogonal features.[4] Image-fusion arithmetic based on WT coefficients can flexibly resolve multidimensional low-frequency and high-frequency image components. Wavelet transforms can also realize multisensor image fusion using rules that emphasize critical features of the scene.[5,6]

Traditional convolution-based WT methods for multiresolution analysis have been widely applied to image fusion for images with a large number of pixels, but the memory and the computational requirements for these techniques, and their Fourier-domain equivalents, can be substantial.

Attempts to create more efficient algorithms in the transform domain have employed the lifting wavelet transform (LWT).[7–9] Also known as the second-generation WT,[10] the LWT is not dependent upon the Fourier transform. Rather, all operations are carried out in the spatial domain. Image reconstruction is achieved by simply adjusting the calculation and sign orders in the decomposition process,[11] thereby reducing two-dimensional image data computation by half, and the data storage to about 75%.

One important motivation for the use of WTs in image processing is their ability to segregate low-frequency content that is critical for interpretation. Traditional image-fusion methods are based on selecting these significant wavelet decomposition coefficients.[12–14] Even with the effective separation and processing of low-frequency components afforded by WT decomposition, such an approach fails to take into full account the relationships among multiple input images. The result can be adverse fusion effects. Significant information can be lost when local area variance corresponding to pixels across images is small.[8,9]

Other algorithms use principal component analysis (PCA) to estimate the wavelet coefficients. This method works well in low-noise environments, but PCA breaks down when corruption is severe, even if only very few of the observations are affected.[15] For example, consider the two PCA simulation results shown in Fig. 1. Suppose that the light line in Fig. 1(a) represents an object in an image, and that the "✕" markers represent samples of that object that have been corrupted by low-level Gaussian noise. The reconstruction of the object from the samples using the classical PCA approach is shown as a heavy line. The results of a similar experiment are shown in Fig. 1(b) where the PCA reconstruction is seriously in error as the result of a single noise outlier in the sampling process.

*Address all correspondence to: Zhuozheng Wang, E-mail: wangzhuozheng@bjut.edu.cn
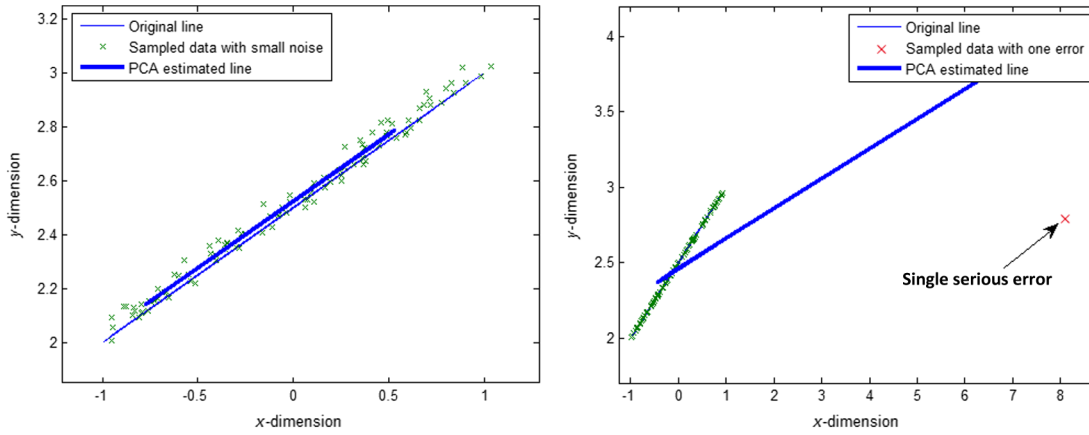
Fig. 1 PCA reconstructions fails when data are corrupted by large errors: (a) samples corrupted by low-level noise and (b) samples include one noise outlier.

To remedy shortcomings in the current methods, this paper presents an improved image-fusion algorithm based on the LWT. For low-frequency image components represented in the LWT decomposition, scale coefficients are determined through matrix completion[16] instead of PCA. For the high-frequency detail and edge information, the LWT coefficients are chosen through self-adaptive regional variance estimation.

2 Matrix Completion and Robust Principal Component Analysis

2.1 Overview

The matrix completion problem has been the subject of intense research in recent years. Candés et al.[17] verify that the ℓ_0-norm optimization problem is equal to ℓ_1-norm optimization under a restricted isometry property. Candés and Recht[16] demonstrate exact matrix completion using convex optimization. The "nuclear norm" of the matrix $\mathbf{X} \in \mathbb{R}^{N \times N}$,

$$\|\mathbf{X}\|_* = \sum_k \sigma_k(\mathbf{X}), \tag{1}$$

in which $\sigma_k(\mathbf{X})$ denotes the k'th largest singular value, can be used to approximate the matrix rank, $\rho(\mathbf{X})$. The method yields a convex minimization problem for which there are numerous efficient solutions. Candés and Recht[16] prove that if the number, S, of sampled entries obeys

$$S \geq CN^{1.2}\rho(\mathbf{X}) \log N \tag{2}$$

for some positive constant C, then $N \times N$ matrix $\mathbf{X}$ can be perfectly recovered with probability ≈ 1, by solving a simple convex optimization problem.

Lin and Ma[15] report a fast, scalable algorithm for solving the robust PCA (RPCA) problem. The method is based on recovering a low-rank matrix with an unknown fraction of corrupted entries. The mathematical model for estimating the low-dimensional subspace is to find a low-rank matrix. The algorithm proceeds as follows: given a matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ with $\rho(\mathbf{A}) \ll \min(M, N)$, the rank is the target dimension of the subspace. The observation matrix $\mathbf{D}$ is modeled as

$$\mathbf{D} = P_\Omega(\mathbf{A}) + \mathbf{E}, \tag{3}$$

in which $P_\Omega(\cdot)$ is a subsampling projection operator and $\mathbf{E}$ represents a matrix of unmodeled perturbations that is assumed sparse relative to $\mathbf{A}$.

2.2 Matrix Completion

The objective of matrix completion is to recover in the low-dimensional subspace the truly low-rank matrix $\mathbf{A}$ from $\mathbf{D}$, under the working assumption that $\mathbf{E}$ is zero. That is, we seek

$$\mathbf{A} = \underset{\mathbf{A}' \in \mathbb{R}^{N \times M}}{\operatorname{argmin}} \|\mathbf{A}'\|_*, \quad \text{subject to } P_\Omega(\mathbf{A}) = \mathbf{D}. \tag{4}$$

It has been shown that the solution to this convex relaxation represents an exact recovery of the matrix $\mathbf{A}$ under quite general conditions.[16] Further, the recovery is robust to noise with small magnitude bounds; that is, when the elements of $\mathbf{E}$ are small and bounded. For example, if $\mathbf{E}$ is a white noise matrix with standard deviation σ, and Frobenius norm $\|\mathbf{E}\|_\mathcal{F} < \epsilon$, then the recovered $\mathbf{D}$ will be in a small neighborhood of $\mathbf{A}$ with high probability if $\epsilon^2 \leq (M + \sqrt{8M})\sigma^2$.[18]

2.3 Robust Principal Component Analysis

Conventional PCA is often used to estimate a low-dimensional subspace via the following constrained optimization problem: In the observation model Eq. (5), minimize the difference in the matrices $\mathbf{A}$ and $\mathbf{D}$ by solving

$$\min_{\mathbf{A}, \mathbf{E}} \|\mathbf{E}\|_\mathcal{F}, \quad \text{subject to } \rho(\mathbf{A}) \leq r, \quad \mathbf{D} = \mathbf{A} + \mathbf{E}, \tag{5}$$

where $r \ll \min\{M, N\}$ is the target dimension of the subspace, and the use of the Frobenius norm represents an assumption that the matrix elements are corrupted by additive i.i.d. Gaussian noise. PCA works well in practice as long as the magnitude of noise is small. To use PCA, the singular value decomposition (SVD) of $\mathbf{D}$ is used to project the columns of $\mathbf{D}$ onto the subspace spanned by the r principal left singular vectors of $\mathbf{D}$.

RPCA employs an identity operator $P_\Omega(\cdot)$ and sparse matrix $\mathbf{E}$ which differ from those in the matrix completion and PCA approach. Wright et al.[19] and Candés et al.[20] have shown that, for a sufficiently sparse error matrix, a low-rank

matrix $\mathbf{A}$ can be recovered exactly from the observation matrix $\mathbf{D}$ by solving the following convex optimization problem:

$$\mathbf{A} = \underset{\mathbf{A}'}{\arg\min}\{\|\mathbf{A}'\|_* + \lambda\|\mathbf{E}\|_1\}, \quad \text{subject to } \mathbf{D} = \mathbf{A} + \mathbf{E}, \tag{6}$$

where λ is a positive weighting parameter. RPCA has been used for background modeling, removing shadows from face images, alignment of the human face, and video denoising.[21,22]

In the present paper, RPCA is coupled with the "inexact augmented Lagrange multiplier" (IALM)[15] method to determine the low-frequency LWT coefficients for fusion of corrupted images. The IALM method is described in Sec. 3.2 after introducing the general procedure.

3 Frequency-Domain Fusion Rules

3.1 Overview

By adopting separate fusion strategies for high- and low-frequency components, the WT can differentially preserve the critical features that accompany these separate bands. The procedure that exploits this property is shown in Fig. 2. The source images are converted to frequency-domain coefficients by the LWT. Frequency-band-dependent fusion rules are applied to the low- and high-frequency components of each image. The inverse lifting wavelet transform (ILWT) is used to reconstruct the fused image.

3.2 Low-Frequency Fusion Based on Inexact Augmented Lagrange Multiplier

Weighted average coefficients are often employed to fuse low-frequency wavelet coefficients. This method is effective when the coefficients of the fused images are similar. However, when contrast reversal occurs in local regions of an image, this procedure results in a loss of image detail in the fused image due to reduced contrast. Further, erroneous or missing regions of corrupted images strongly affect PCA results. These inadequacies of the weighted average method and PCA provide the motivation for using RPCA to determine the weighting of low-frequency coefficients.

There is ordinarily little difference in the low-frequency coefficient values extracted by the LWT from different images of the same scene. RPCA coefficients are used to represent low-frequency content in an attempt to preserve fidelity and coherency between the subbands. Algorithms

Fig. 2 Image fusion processing based on wavelet transform.

have been developed in this research to solve the RPCA problem that is the basis for the recovery of the low-rank matrix $\mathbf{A}$ and the estimation of the sparse matrix $\mathbf{E}$ from the observation matrix $\mathbf{D}$. We employ the IALM method to compute the low-frequency subband coefficients. The method is sketched as follows.

Let $\Gamma = \{\mathbf{I}_k \in \mathbb{R}^{N_1 \times N_2}\}_{k=1}^K$ denote a set of corrupted images from K sensors, and let $\tilde{\Gamma} = \{\tilde{\mathbf{I}}_k \in \mathbb{R}^{(N_1 \times N_2)/4^L}\}_{k=1}^K$ be the corresponding set of low-frequency subimages computed using the LWT. L is the number of LWT layers. For simplicity, we assume square images so that $N_1/4^L = N_2/4^L \stackrel{\text{def}}{=} N$. Stack all N columns of each $\tilde{\mathbf{I}}_k$ into a single vector of dimension N^2, then use these vectors as K columns of a matrix $\tilde{\mathbf{I}}_D$. After normalizing the data, we denote by $i_{\ell k}$ the (ℓ, k) element of $\tilde{\mathbf{I}}_D$,

$$\tilde{\mathbf{I}}_D = \begin{pmatrix} i_{11} & i_{12}\ldots & i_{1K} \\ i_{21} & i_{22} & \cdots & i_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ i_{N^21} & i_{N^22} & \cdots & i_{N^2K} \end{pmatrix}. \tag{7}$$

The cumulative low-frequency subimage matrix is modeled similarly to Eq. (3),

$$\tilde{\mathbf{I}}_D = \tilde{\mathbf{I}}_A + \tilde{\mathbf{I}}_E, \tag{8}$$

in which $\tilde{\mathbf{I}}_A \in \mathbb{R}^{N^2 \times K}$ denotes the noise-free and integrated low-frequency subimage sequence matrix, and $\tilde{\mathbf{I}}_E \in \mathbb{R}^{N^2 \times K}$ denotes the sparse error matrix from which high-frequency content has been attenuated by the selection of LWT coefficients. The low-frequency LWT coefficients are similar across multiple subimages of the same scene. According to the model, $\tilde{\mathbf{I}}_A$ is noise-free and will ideally, therefore, consist of K identical columns. Accordingly, $\tilde{\mathbf{I}}_A$ will be of low rank as required by the matrix completion procedure. Thus, $\tilde{\mathbf{I}}_A$ can be estimated via matrix completion and RPCA by solving

$$\min_{\tilde{\mathbf{I}}_A, \tilde{\mathbf{I}}_E} \|\tilde{\mathbf{I}}_A\|_* + \lambda\|P_\Omega(\tilde{\mathbf{I}}_E)\|_1 \quad \text{subject to } \tilde{\mathbf{I}}_A + \tilde{\mathbf{I}}_E = \tilde{\mathbf{I}}_D, \tag{9}$$

where the augmented Lagrange multiplier is

$$L(\tilde{\mathbf{I}}_A, \tilde{\mathbf{I}}_E, \mathbf{Y}, \mu) = \|\tilde{\mathbf{I}}_A\|_* + \lambda\|P_\Omega(\tilde{\mathbf{I}}_E)\|_1 \\ + \text{Tr}\{\mathbf{Y}, \tilde{\mathbf{I}}_D - \tilde{\mathbf{I}}_A - \tilde{\mathbf{I}}_E\} + \frac{\mu}{2}\|\tilde{\mathbf{I}}_D - \tilde{\mathbf{I}}_A - \tilde{\mathbf{I}}_E\|_{\mathcal{F}}^2. \tag{10}$$

In this equation, λ is an estimated positive weighting parameter representing the proportion of the sparse matrix $\tilde{\mathbf{I}}_E$ in the low-rank matrix $\tilde{\mathbf{I}}_A$. The default value for this fraction is $1/N$. μ is a positive tuning parameter balancing accuracy and computational effort. $\text{Tr}\{\mathbf{A}, \mathbf{B}\}$ is the trace of the product $\mathbf{A}^T\mathbf{B}$ and $\mathbf{Y}$ is the iterated Lagrange multiplier.

A flowchart of the IALM algorithm is shown in Fig. 3. Definitions of the notation used in the flowchart appear in Table 1. The algorithm is recursive with superscript j indicating the iteration number. The quantity $\tilde{\mathbf{I}}_A^{(j')} \in \mathbb{R}^{N^2 \times K}$ is the recovered low-rank matrix for some sufficiently large j, say j'. A reasonable strategy for transforming the resulting $\tilde{\mathbf{I}}_A^{(j')}$ to the final low-frequency subimage is to unwrap its first
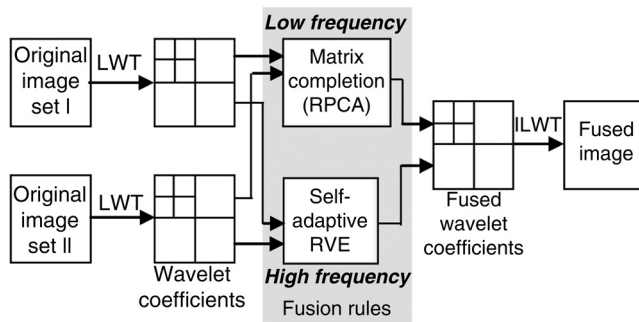
Fig. 3 Flowchart of operations in the IALM algorithm.

Table 1 Notation used in the IALM$_\partial$ algorithm.

Notation	Definition
$\tilde{\mathbf{I}}_D$	Low-frequency subimage observation matrix
$\tilde{\mathbf{I}}_E^{(j)}$	Error (sparse) matrix, iteration j
$\tilde{\mathbf{I}}_A^{(j)}$	Recovered low-rank subimage matrix, iteration j
$\mathbf{Y}^{(j)}$	Lagrange multiplier matrix, iteration j
τ	Mean-squared-error tolerance bound
$\nabla(\mathbf{X})$	Singular value decomposition (SVD) of general matrix $\mathbf{X}$
$\mathbf{U}$ and $\mathbf{V}$	Customary notation for orthogonal matrices of SVD
$\mathbf{S}$	Customary notation for diagonal matrix of singular values
$\mathbf{S}_\varepsilon[x]$	Soft-shrinkage operator applied to scalar x[15] $$\mathbf{S}_\varepsilon[x] \overset{\text{def}}{=} \begin{cases} x - \varepsilon, & x > \varepsilon \\ x + \varepsilon, & x < -\varepsilon \\ 0, & \text{otherwise} \end{cases}, x \in \mathbb{R}, \varepsilon > 0$$

regions across subimages vary widely, a high-frequency fusion rule for selecting the source image of greatest variance has been shown to be effective at preserving image features.[8,9] However, if the local variances of two source images are similar, this method can result in the loss of information by discarding subtle variations among different subimages. An empirical procedure has been developed in which a thresholding procedure is used to segregate local areas that have sufficiently large variance. This allows the entire set to be represented by the single maximum-variance set member. The selection of this difference threshold, ξ, is discussed below.

Let us return to the original set of images $\Gamma = \{\mathbf{I}_k \in \mathbb{R}^{N_1 \times N_2}\}_{k=1}^K$. Denote by $\mathbf{I}_k(x, y)$ the gray-scale value at pixel (x, y) in the k'th image. Also let $\mathbf{V}_k \in \mathbb{R}^{N_1 \times N_2}$ denote a matrix associated with image $\mathbf{I}_k$ in which matrix element $\mathbf{V}_k(x, y)$ contains the normalized sample variance of the 3×3 window of pixels centered on pixel (x, y). The normalized sample variance means that all variance values are in the interval $[0,1]$. Without loss of generality, we select images $\mathbf{I}_1$ and $\mathbf{I}_2$ with which to describe the steps of the high-frequency fusion algorithm:

1. Compute normalized sample variance matrices $\mathbf{V}_1$ and $\mathbf{V}_2$. Then $\mathbf{V}_k(x, y)$ denotes the normalized variance value of pixel (x, y) in image $\mathbf{I}_k$ for $k = 1, 2$.
2. Implement the LWT over $L = 2$ layers against $\mathbf{I}_1$, $\mathbf{I}_2$, $\mathbf{V}_1$, and $\mathbf{V}_2$. Multiresolution structures for each matrix are obtained: $\mathbf{I}_1^\theta$, $\mathbf{I}_2^\theta$, $\mathbf{V}_1^\theta$, and $\mathbf{V}_2^\theta$, in which the superscript θ takes one of three designators of direction—horizontal (h), vertical (v) or diagonal (d)—associated with structure matrix

$$\Delta_V^\theta(x, y) = \mathbf{V}_1^\theta(x, y) - \mathbf{V}_2^\theta(x, y). \tag{11}$$

Let $\Delta_V(x, y)$ denote the sum of the differences in the horizontal, vertical, and diagonal directions

column to form the original $N \times N$ image structure. The final low-frequency subimage is denoted $\tilde{\mathbf{I}}_\partial$.

In this process, $\mathbf{Y}^{(0)}$ is initialized to $\tilde{\mathbf{I}}_D / \max(\|\tilde{\mathbf{I}}_D\|_2, \|\tilde{\mathbf{I}}_D\|_\infty)$; and $\tilde{\mathbf{I}}_E^{(0)}$ is initialized to zero matrix as the same size of $\tilde{\mathbf{I}}_D$; λ is initialized to $1/\sqrt{m}$ where m is the column size of $\tilde{\mathbf{I}}_D$; tolerance for stopping criterion τ is initialized to 1×10^{-7}; and j is set to zero for loop computation.

3.3 High-Frequency Fusion Based on Self-Adapting Regional Variance Estimation

Processing of high-frequency wavelet coefficients has a direct effect on salient details which affect the overall clarity of the image. As the variance of a subimage characterizes the degree of gray level change in a corresponding image region, the variance is a key indicator in processing of high-frequency components. In addition, there is generally a strong correlation among adjacent pixels in a local area, so that there is significant amount of shared information among neighboring pixels. When variances in corresponding local

$$\Delta_V(x, y) = [\mathbf{V}_1^h(x, y) - \mathbf{V}_2^h(x, y)] + [\mathbf{V}_1^v(x, y) \\ - \mathbf{V}_2^v(x, y)] + [\mathbf{V}_1^d(x, y) - \mathbf{V}_2^d(x, y)], \qquad (12)$$

in which $\mathbf{V}_k^\theta(x, y)$ indicates the normalized variance of the k'th image in direction θ.

3. Compare the threshold value and $|\Delta_V(x, y)|$. If $|\Delta_V(x, y)| \geq \xi$ take the pixel value with bigger variance as the wavelet coefficient after fusion; otherwise use a weighted sum to compute the wavelet coefficient, D_F^θ is the multiresolution structure after fusion, namely

$$D_F^\theta(x, y) = \begin{cases} \mathbf{I}_1^\theta(x, y), & \text{when } \Delta_V(x, y) > 0 \text{ and } |\Delta_V(x, y)| \geq \xi, \\ \mathbf{I}_2^\theta(x, y). & \text{when } \Delta_V(x, y) < 0 \text{ and } |\Delta_V(x, y)| \geq \xi, \\ \mathbf{V}_1^\theta(x, y)\mathbf{I}_1^\theta(x, y) + \mathbf{V}_2^\theta(x, y)\mathbf{I}_2^\theta(x, y), & \text{when } |\Delta_V(x, y)| < \xi. \end{cases} \qquad (13)$$

In this study, the value of ξ is set to 0.8. This means that when the normalized variance of the pixel (x, y) in one image is much greater than another, the source image of greater variance is selected. Otherwise, the coefficient is obtained by averaging as in Eq. (13). This fusion rule for high-frequency subimages not only results in the retention of details, but it also prevents the loss of image information caused by redundant data. It ensures the consistency of the fused image.

In summary, IALM is used to determine the low-frequency component to be fused, and self-adapting regional variance is employed to estimate the high-frequency contribution. The fused wavelet coefficients are combined by ILWT to create the final result.

4 Experimental Results and Analysis

4.1 Comparison of Robust Principal Component Analysis Algorithms

To validate the new procedure, four groups of experiments results are reported. The objective of the first is to compare the performance of RPCA algorithms with that of IALM. The results are shown in Table 2. Two mainstream algorithms are compared—singular value thresholding (SVT), accelerated proximal gradient with IALM.

In this table, the input dataset named observation matrix $\mathbf{D}$ of Eq. (6) is of dimension $N \times N$. It has some random missing or broken pixels. For fair comparison, we set r, the rank of $\mathbf{A}$, to 0.05 N, and define the normalized mean squared error (NMSE) as

$$\text{NMSE} = \frac{\|\mathbf{D} - \mathbf{A} - \mathbf{E}\|_\mathcal{F}}{\|\mathbf{D}\|_\mathcal{F}}. \qquad (14)$$

In Table 2, the column labeled #SVD indicates the number of iterations. The "times" column displays the number of seconds to run the algorithm. The oversampling rate (p/d_r) is six, implying $\sim$60% downsampling of the data appearing in the observation matrix, in which, d_r indicates the number of degrees of freedom in the rank r matrices: $d_r = r(2N - r)$. p elements from $\mathbf{A}$ are then sampled uniformly to form the known samples in $\mathbf{D}$.[16]

Among the three algorithms, IALM exhibits superiority performance in all three measures. The results indicate that time increases proportionately with N^2. Note, however, that #SVD is not dependent upon N.

4.2 Fusion of Clean Images

For convenience, we will refer to the new algorithm as IALM$_\partial$. The next two groups of experiments involve processing of left-focus–right-focus images and visible-light–infrared-light images, comparing different image-fusion algorithms with IALM$_\partial$. The source images are not corrupted by noise or errors. The spline 5/3 wavelet basis[23] was

Table 2 Comparison of RPCA algorithms.

| N	Algorithm	r	NMSE	#SVD	Time (s)
500	SVT	25	1.35×10^{-4}	78	13.72
500	APG	25	2.33×10^{-5}	56	10.34
500	IALM	25	4.73×10^{-7}	34	3.32
600	SVT	30	1.27×10^{-4}	77	19.02
600	APG	30	2.11×10^{-5}	58	16.92
600	IALM	30	4.61×10^{-7}	34	5.64
700	SVT	35	1.36×10^{-4}	74	24.77
700	APG	35	2.25×10^{-5}	58	26.25
700	IALM	35	4.62×10^{-7}	34	8.41
800	SVT	40	1.26×10^{-4}	75	33.95
800	APG	40	2.14×10^{-5}	59	42.14
800	IALM	40	4.30×10^{-7}	34	12.09
900	SVT	45	1.27×10^{-4}	75	42.52
900	APG	45	2.03×10^{-5}	60	59.24
900	IALM	45	4.45×10^{-7}	34	16.78
1000	SVT	50	1.25×10^{-4}	73	52.65
1000	APG	50	2.16×10^{-5}	60	72.26
1000	IALM	50	4.45×10^{-7}	34	22.54
2000	SVT	100	1.30×10^{-4}	71	257.17
2000	APG	100	2.05×10^{-4}	64	387.42
2000	IALM	100	4.39×10^{-7}	34	154.43
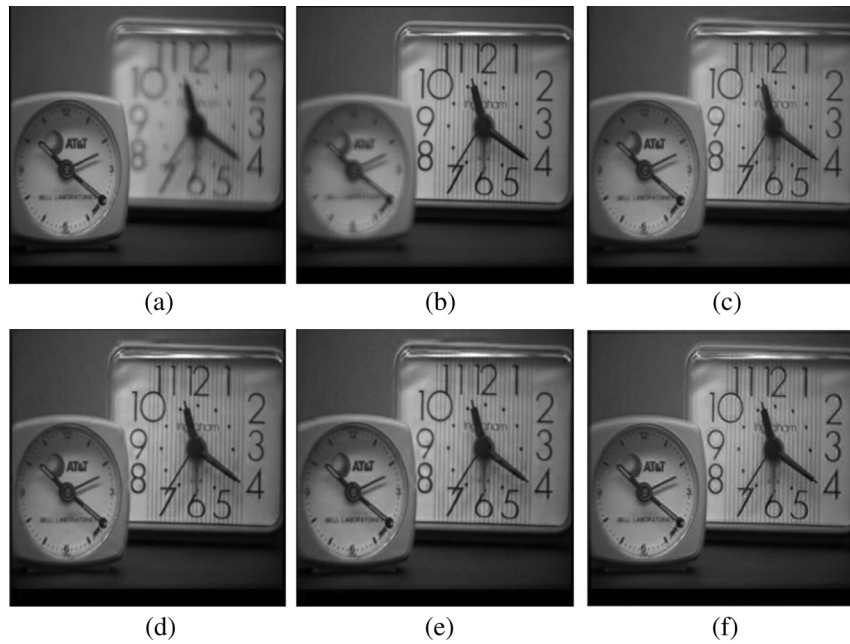
Fig. 4 Multifocus image-fusion experiment: (a) left-focus image, (b) right-focus image, (c) WA_LM, (d) PCNN, (e) PCA$_\partial$, and (f) IALM$_\partial$.

selected for the LWT process. Through factorization, the equivalent lifting wavelet was obtained. The experimental results are shown in Figs. 4 and 5.

The first group of source images involves those with eccentric focus, the second contains images of visible contrasting and infrared light. Fig. 4(a) shows a left-focused source image, whereas Fig. 4(b) is right-focused; Fig. 5(a) is a visible-light source image, while Fig. 5(b) uses an infrared source; in Figs. 4(c)–4(f) and 5(c)–5(f) are, respectively, the fusion results by the weighted average over low frequencies and the absolute value maximum method over high frequencies (WA_AM), weighted average over low frequencies and the local area maximum method over high frequencies (WA_AM), improved pulse-coupled neural networks (PCNN) method,[24,25] and PCA-weighted over low frequencies, the self-adaptive regional variance

estimation method over high frequencies (PCA$_\partial$), and the algorithm developed in this paper (IALM$_\partial$).

The processed images empirically suggest that a clearer fused image is obtained through (IALM$_\partial$). More detailed information is evident, e.g., in Figs. 4(e) and 4(f) in which the image information on the left edge of the large alarm clock is apparently richer than the same feature in the other three fused images. This also means that algorithm IALM$_\partial$ is equally effective to algorithm PCA$_\partial$, even though the algorithm IALM$_\partial$ has more detailed information (Table 2). Furthermore, the new algorithm achieves a fusion result with finer detail. For example, the barbed wire in Fig. 5(d) is more clearly visible than the same feature in (c). In Fig. 5, the person in 5(c) is better defined than in 5(d), while in 5(e) and 5(f), both the barbed wire and the person, and even the smoke in the upper-right corner of the image, are easier
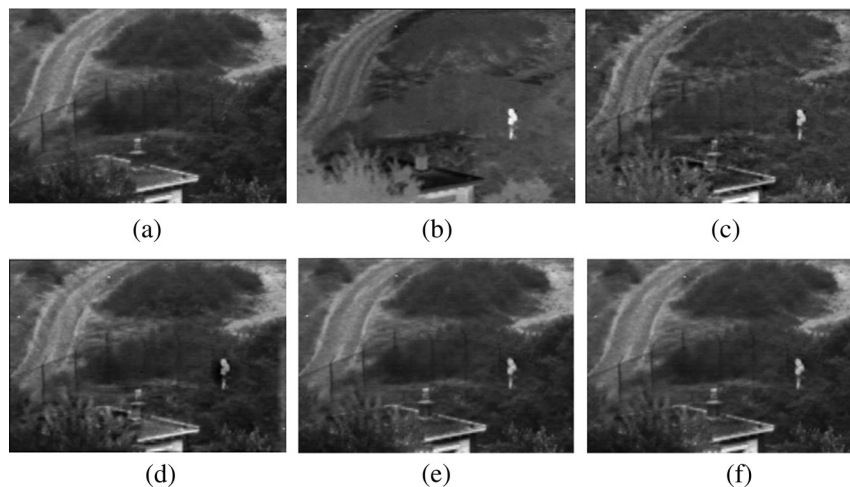
Fig. 5 Visible light and infrared image-fusion experiment: (a) visible-light image, (b) infrared image, (c) WA_LM, (d) PCNN, (e) PCA$_\partial$, and (f) IALM$_\partial$.

to identify than in the others. This enhanced clarity admits more effective subsequent processing.

The following objective criteria were evaluated:

1. The "mutual information" (MI) is a measure of statistical dependence that can be interpreted as the amount of information transmitted from the source images to the fused image.[26] To assess the MI between source image I_1 and the fused image, say I_F, we use the estimator

$$M_{1,F} = \sum_{l_1, l_F} h_{1,F}(l_1, l_F) \log \frac{h_{1,F}(l_1, l_F)}{h_1(l_1) h_F(l_F)}, \quad (15)$$

where $h_1(l_1)$ and $h_F(l_F)$ represent the normalized histogram of source image I_1 and fused image I_F, respectively. l_1 and l_F each take integers indicating one of 2^8 gray levels $\{0, 1, \ldots, 255\}$. $h_{1,F}(l_1, l_F)$ denote the jointly normalized histogram of I_1 and image I_F. Similarly, $M_{2,F}$ denotes the mutual information between image I_2 and the fused image I_F. The MI between the source images I_1 and I_2 and the fused image I_F is

$$M_{1,2,F} = M_{1,F} + M_{2,F}. \quad (16)$$

A larger MI value indicates that the fused image includes more information from the original images.

2. The "average gradient" (AG), or "clarity," reflects the preservation of gray level changes in the image. With dimensions $N_1 = N_2 = N$, larger values of AG imply greater clarity and edge preservation. Gray-level differentials are important, e.g., in texture rendering. The AG is defined as

$$\nabla \bar{g} = \frac{1}{N^2} \sum_{x=1}^{N-1} \sum_{y=1}^{N-1} \sqrt{[\Delta I_x^2(x,y) + \Delta I_y^2(x,y)]/2}, \quad (17)$$

where $\Delta I_x(x,y) = I(x+1,y) - I(x,y)$ and $\Delta I_y(x,y) = I(x,y+1) - I(x,y)$ are the gray value differentials in the coordinate x and y directions, respectively.

3. The "correlation coefficient" (CC) is used to compare two images of the same object (or scene). CC, which measures the correlation (degree of linear coherence) between the original and the fused images, is defined as

$$C_{F,1} = \frac{\sum_{x,y}[(I_F(x,y) - \bar{I}_F)][(I_1(x,y) - \bar{I}_1)]}{\sqrt{\sum_{x,y}[I_F(x,y) - \bar{I}_F]^2 \sum_{x,y}[I_1(x,y) - \bar{I}_1]^2}}, \quad (18)$$

where $I_F(x,y)$ and $I_1(x,y)$ are the gray levels at pixel (x,y) in the fused and original images, and $\bar{I}_F$ and $\bar{I}_1$ denote the average gray levels in the two images.

4. The "degree of distortion" (DD), a direct indicator of image fidelity, is defined as

$$D_{F,1} = \frac{1}{N_1 \times N_2} \sum_{x=1}^{N_1} \sum_{y=1}^{N_2} |I_F(x,y) - I_1(x,y)|, \quad (19)$$

in which $I_F(x,y)$ and $I_1(x,y)$ are as defined above.

5. The $Q^{AB/F}$ metric quantifies the amount of edge information transferred from two source images I_A and I_B to a fused image I_F.[26] It is calculated as

$$Q^{AB/F}$$
$$= \frac{\sum_{x=1}^{N_1} \sum_{y=1}^{N_2} [Q^{AF}(x,y) w^A(x,y) + Q^{BF}(x,y) w^B(x,y)]}{\sum_{x=1}^{N_1} \sum_{y=1}^{N_2} [w^A(x,y) + w^B(x,y)]}, \quad (20)$$

where each image is of size $N_1 \times N_2$. α and β represent, respectively, the edge strength and orientation. $Q^{AF}(x,y)$ is the product of $Q_\alpha^{AF}(x,y)$ and $Q_\beta^{AF}(x,y)$ which represent, respectively, how well the edge strength and orientation values of a pixel are represented in the fused image I_F. Similarly, $Q^{BF}(x,y)$ is computed as the product of $Q_\alpha^{BF}(x,y)$ and $Q_\beta^{BF}(x,y)$ which represent, respectively, how well the edge strength and orientation values of a pixel (x,y) in I_2 are represented in the fused image I_F. $w^A(x,y)$ and $w^B(x,y)$, respectively, denote the proportion of $Q^{AF}(x,y)$ and $Q^{BF}(x,y)$, which reflect the importance of $Q^{AF}(x,y)$ and $Q^{BF}(x,y)$. The dynamic range of $Q^{AB/F}$ is between [0 1], and it should be as close to 1 as possible.

6. The "peak signal-to-noise ratio" (PSNR) is an expression for the ratio between the maximum possible power of a signal and the power of distorting noise that affects the quality of its representation. This objective metric is used to compare the effectiveness of algorithms by measuring the proximity of the fused image and the original image. The PSNR is computed as

$$\text{PSNR} = 10 \lg \frac{(L-1)^2}{\text{RMSE}^2}, \quad (21)$$

where RMSE denotes the root mean square error between the reference and fused images. $L = 256$ is the number of gray levels used in representing an image. A larger PSNR value indicates a better fusion result.

Tables 3 and 4 report the objective performance evaluation measures for the four fusion algorithms.

Table 3 Experimental objective evaluation measures of Fig. 4.

Evaluation indicator	WA_LM	PCNN	PCA$_\partial$	IALM$_\partial$
MI	6.1604	7.0814	7.2788	7.5191
AG	4.2067	6.8089	6.8096	6.8089
CC	0.9768	0.9749	0.9836	0.9927
DD	3.9762	3.9089	3.6406	3.5743
$Q^{AB/F}$	0.6133	0.6897	0.6987	0.6929
PSNR	22.6195	28.1095	28.1270	31.3846

Table 4 Evaluation comparison of Fig. 5.

Evaluation indicator	WA_LM	PCNN	PCA_∂	$IALM_\partial$
MI	2.7595	3.7565	3.8953	3.8938
AG	6.8556	7.8666	7.9206	8.1982
CC	0.7873	0.8729	0.8808	0.8976
DD	17.1008	11.0016	10.9259	10.2100
$Q^{AB/F}$	0.5988	0.6978	0.6798	0.7548
PSNR	20.4271	24.3396	25.1234	25.3540

Relative to the other algorithms, $IALM_\partial$ obtains the largest MI and AG for the fused images, suggesting that this algorithm can provide fused images with higher information content and better clarity. The objective indicators of fidelity to the source image also favor the IALM and self-adaptive regional variance estimation algorithm performance.

4.3 Fusion of Corrupted Images

To assess whether $IALM_\partial$ is robust to missing data and image corruption, we continue to use clean, multifocus clock images for processing. At a 0.15 error rate, 15% of the pixels of the original image are corrupted, and an additional 15% are missing (gray-level values set to zero). This implies an effective data corruption rate or 30%. The results of the test of the four algorithms are shown in Fig. 6. Figures 6(a) and 6(b) show, respectively, Fig. 4(a) with errors and Fig. 4(b) with errors. Figure 6(c) shows the result of using PCA_∂ without a denoising filter, while Fig. 6(d), labeled $PCA_{\partial,\mathcal{F}}$, shows the result of using PCA_∂ with an adaptive median filter. The result of using PCNN with an adaptive median filter is labeled $PCNN_\mathcal{F}$ and appears in Fig. 6(e). To achieve this outcome, we use the adaptive median filtering strategy proposed by Chen and Wu[27] to identify pixels corrupted by impulsive noise and replace each damaged pixel by the median of its neighborhood. The adaptive median filter can employ varying window sizes to accommodate different noise conditions and to reduce distortions like excessive thinning or thickening of object boundaries. Figure 6(f) shows results using $IALM_\partial$ without denoising. The clarity of result 6(f) relative to those in 6(c), 6(d), and 6(e) is quite apparent. The empirical image quality tracks the improvement in PSNR as reported in the captions. Figures 6(g) and 6(h) show 400% blow ups of portions of 6(e) and 6(f).

These results demonstrate the ability of $IALM_\partial$ to recover the missing or erroneous data, while preserving image detail in both corrupted and clean images.
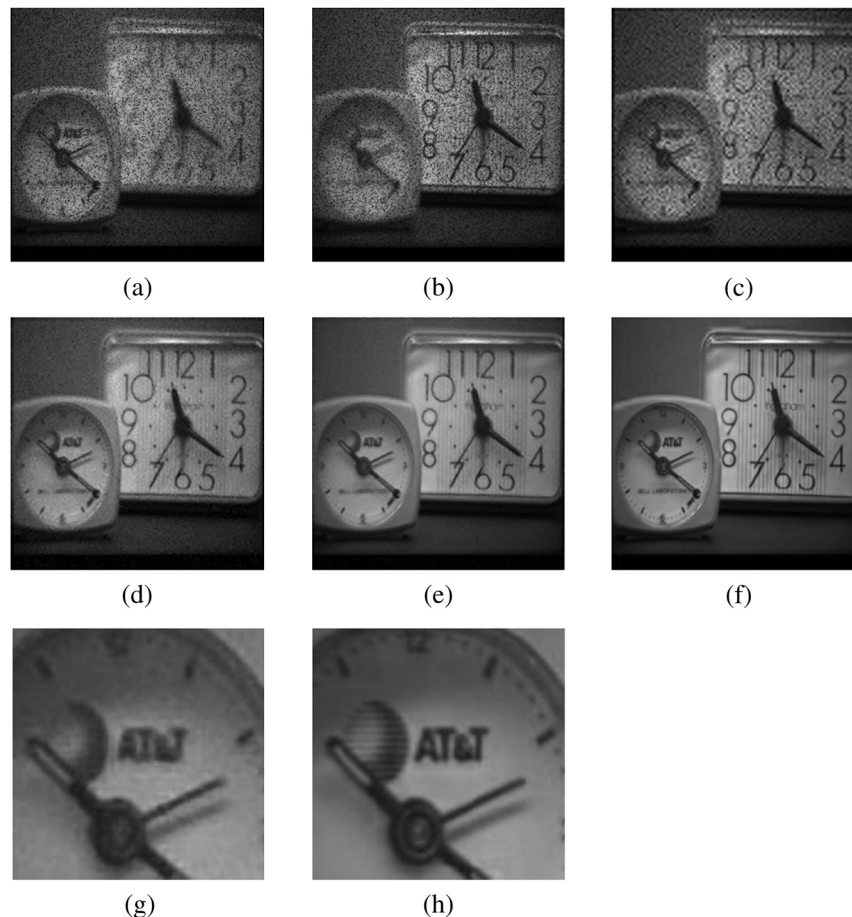
Fig. 6 Multifocus corrupted image-fusion experiment: (a) Fig. 4(a) with errors; (b) Fig. 4(b) with errors; (c) PCA_∂ (PSNR = 17.82); (d) $PCA_{\partial,\mathcal{F}}$ (PSNR = 19.37) (e) $PCNN_\mathcal{F}$ (PSNR = 20.76); (f) $IALM_\partial$ (PSNR = 30.33); (g) zoom out 400% of (e); and (h) zoom out 400% of (f).

5 Conclusions

Traditional convolution-based wavelet transform processing for image fusion has shortcomings including large memory requirements and high computational complexity. The approach to fusion taken in this research uses different fusion rules for low-frequency and high-frequency decomposition components represented on a lifting wavelet basis set. Low-frequency components are characterized by the matrix completion and RPCA methods: IALM, whereas the high-frequency components critical for image details are represented by taking into account the variance differences among proximal neighborhoods. Furthermore, strong correlation between pixels in a local area is captured by a self-adaptive regional variance assessment.

Experimental results show that the new algorithm not only improves the amount of information and the correlation between the fused and source images, but also reduces the level of distortion. Significant clarity improvement relative to state-of-the-art methods is also demonstrated for corrupted images.

References

1. B. Khaleghi et al., "Multisensor data fusion: a review of the state-of-the-art," *Inf. Fusion* **14**, 28–44 (2013).
2. G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Inf. Fusion* **4**(4), 259–280 (2003).
3. Y. Chai, H. Li, and Z. Li, "Multifocus image fusion scheme using focused region detection and multiresolution," *Opt. Commun.* **284**(19), 4376–4389 (2011).
4. R. K. Sharma and M. Pavel, *Probabilistic Model-Based Multisensor Image Fusion*, pp. 1–35, Oregon Graduate Institute of Science and Technology (1999).
5. Y. Zheng, "An orientation-based fusion algorithm for multisensor image fusion," *Proc. SPIE* **7710**, 77100K (2010).
6. R. Nava, B. Escalante-Ramírez, and G. Cristóbal, "A novel multi-focus image fusion algorithm based on feature extraction and wavelets," *Proc. SPIE* **7000**, 700028 (2008).
7. C. Ramesh and T. Ranjith, "Fusion performance measures and a lifting wavelet transform based algorithm for image fusion," *Inf. Fusion* **1**, 317–320 (2002).
8. G. Liu and C. Liu, "A novel algorithm for image fusion based on wavelet multi-resolution decomposition," *J. Optoelectron.* **15**, 334–347 (2004).
9. Z. Qiang and J. Peng, "Remote sensing image fusion based on small wavelet transform's local variance," *J. Huazhong Univ. Sci. Technol.* **6**, 89–91 (2003).
10. W. Sweldens, "The lifting scheme: a construction of second generation wavelets," *SIAM J. Math. Anal.* **29**(2), 511–546 (1998).
11. M. Chen and H. Di, "Study on optimal wavelet decomposition level for multi-focus image fusion," *Opto-Electron. Eng.* **31**, 64–67 (2004).
12. Q. Lin and F. Gui, "A novel image fusion algorithm based on wavelet transforms," *Proc. SPIE* **7001**, 70010M (2008).
13. L. S. Arivazhagan, L. Ganesan, and T. Kumar, "A modified statistical approach for image fusion using wavelet transform," *Signal Image Video Process.* **3**(2), 137–144 (2009).
14. S. El-Khamy et al., "Regularized super-resolution reconstruction of images using wavelet fusion," *Opt. Eng.* **44**(9), 097001 (2005).
15. Z. Lin and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Urbana-Champaign Technical Report 2, University of Illinois (2011).
16. E. J. Candés and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.* **9**, 717–772 (2009).
17. E. J. Candés, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.* **59**, 1207–1223 (2006).
18. E. J. Candés and Y. Plan, "Matrix completion with noise," *Proc. IEEE* **98**, 925–936 (2010).
19. J. Wright et al., "Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization," *Proc. Neural Inf. Process. Syst.* **3**, 1–9 (2009).
20. E. J. Candés et al., "Robust principal component analysis?" *J. ACM* **58**, 11 (2011).
21. W. Tan, G. Cheung, and Y. Ma, "Face recovery in conference video streaming using robust principal component analysis," in *Proc. IEEE Int. Conf. on Image Processing*, pp. 3225–3228 (2011).
22. H. Ji et al., "Robust video denoising using low rank matrix completion," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1791–1798 (2010).
23. A. Z. Averbuch and V. A. Zheludev, "Image compression using spline based wavelet transforms," *Wavelets Signal Image Anal.* **19**, 341–376 (2001).
24. X. Qu et al., "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain," *Acta Autom. Sin.* **34**, 1508–1514 (2008).
25. Y. Chai, H. F. Li, and M. Y. Guo, "Multifocus image fusion scheme based on features of multiscale products and PCNN in lifting stationary wavelet domain," *Opt. Commun.* **284**, 1146–1158 (2011).
26. C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.* **36**, 308–309 (2000).
27. T. Chen and H. Wu, "Adaptive impulse detection using center-weighted median filters," *IEEE Signal Process. Lett.* **8**, 1–3 (2001).

Zhuozheng Wang is an associate professor at Beijing University of Technology and a visiting scholar at Michigan State University sponsored by the China Scholarship Council. He received his MS and PhD degrees in electronic engineering from Beijing University of Technology in 2005 and 2013. He is the first author of more than 10 academic papers and has written one book chapter. His current research interests include image processing, electroencephalography, and virtual reality technology. He has been a reviewer and is a member of SPIE.

J. R. Deller Jr. is an IEEE fellow and professor of electrical and computer engineering at Michigan State University, where he received the distinguished faculty award in 2004. He received a PhD in biomedical engineering in 1979, an MS degree in electrical and computer engineering in 1976, and an MS degree in biomedical engineering in 1975 from the University of Michigan, and his BS degree in electrical engineering (summa cum laude) in 1974 from Ohio State University. His research interests include statistical signal processing with applications to speech and hearing, genomics, and other aspects of biomedicine.

Blair D. Fleet received her BS degree (summa cum laude) from Morgan State University, Baltimore, MD, in 2010, and her MS degree from Michigan State University in 2012, both in electrical engineering. She is a National Science Foundation graduate research fellowship award recipient, as well as a GEM (the National Consortium for graduate degrees for Minorities in Engineering and Science, Inc.) fellow. She is currently pursuing her PhD in electrical engineering at Michigan State University. Her research interests include merging signal/image processing with the evolutionary computation fields to solve challenging engineering processing problems, especially in the biomedical domain.