Dimensionality Reduction for Registration of High-Dimensional Data Sets

Min Xu, *Member, IEEE*, Hao Chen, *Member, IEEE*, and Pramod K. Varshney, *Fellow, IEEE*

***Abstract*— Registration of two high-dimensional data sets often involves dimensionality reduction to yield a single-band image from each data set followed by pairwise image registration. We develop a new application-specific algorithm for dimensionality reduction of high-dimensional data sets such that the weighted harmonic mean of Cramér-Rao lower bounds for the estimation of the transformation parameters for registration is minimized. The performance of the proposed dimensionality reduction algorithm is evaluated using three remotes sensing data sets. The experimental results using mutual information-based pairwise registration technique demonstrate that our proposed dimensionality reduction algorithm combines the original data sets to obtain the image pair with more texture, resulting in improved image registration.**

***Index Terms*— Dimensionality reduction, Cramer-Rao lower bound, image registration.**

I. Introduction

IMAGE registration is an important step in remote-sensing applications that involve multiple-sensor or multi-date data of a common scene. The primary objective of image registration is to match two or more images that differ in certain aspects, e.g., translation, scaling, and rotation (assuming a rigid-body misalignment), but essentially represent the same scene [1]. A transformation is to be found so that the points and objects in one image can be related to their corresponding points and objects in the other image.

Current registration approaches may broadly be classified as feature-based, Fourier-based, and intensity-based methods. An overview of these image registration approaches can be found in [1] and [3]. Intensity based image registration finds the transformation such that the similarity between the two images is maximized. The similarity measures include correlation, mean square difference and mutual information [4] between the two images. Correlation and mean square difference measures are easy to compute, but are reliable only when the gray-level characteristics of two images are very similar. Different from mean square difference measure that assumes that the two images are the same, and the correlation measure that assumes that the two images have a linear relationship, the mutual information measure has been found to be very robust for registration of multimodal images [5]. Exhaustive search for the optimal solution is quite prohibitive. Thus, instead, some efficient optimization algorithms are usually used. As a result, a multiresolution technique, such as Gaussian pyramid or wavelet transform, is sometimes applied [2].

In remote sensing area, as multiple imaging sensors are increasingly being used to acquire image data that have higher and higher resolution and high dimensionality, automatic and accurate registration of high-dimensional data sets, e.g., multi-spectral and hyper-spectral data, is required which is quite challenging. Multispectral image registration involves the registration of several bands to several bands whereas hyperspectral image registration involves the registration of dozens or hundreds of bands in the two hyperspectral data cubes. It is noted that registration of two high-dimensional data sets in remote sensing applications is different from volume registration in medical image analysis [6]. Volume registration assumes a band-to-band match, i.e., it assumes that the kth band of one data set is matched to the kth band of the other data set while registration in remote sensing does not assume that.

Intensity based image registration has been widely investigated [2] [5] [7] for remote sensing applications. Performing intensity based registration directly on two high-dimensional data sets is extremely computationally expensive. Hoge and Westin [8] have proposed a Fourier transform based algorithm to use a high-order singular value decomposition to decompose the phase correlation between two high-dimensional data sets to find the translational displacements. To extend it to other distortion models such as rigid body transformation, more computational effort is needed. Therefore, it is desired to develop a technique that is able to register two high-dimensional data sets in a computationally efficient manner that yields accurate results. One approach is to perform dimensionality reduction by combining all the bands of each image data set to produce a single-band image, and then perform image registration on the obtained single-band image pair. In this approach, intensity based registration of two high-dimensional data sets includes two steps.

1) Combine the images corresponding to different bands of each data set and obtain a single-band image from each data set.
2) Perform intensity based pairwise image registration for the two resulting single-band images.

Manuscript received June 27, 2012; revised November 5, 2012; accepted February 28, 2013. Date of publication April 25, 2013; date of current version May 24, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Brian D. Rigling.

M. Xu was with the Department of Electrical and Computer Engineering, Syracuse University, Syracuse, NY 13244 USA. She is now with Blue Highway Inc., Syracuse, NY 13244 USA (e-mail: mxu01@syr.edu).

H. Chen is with the Department of Electrical and Computer Engineering, Boise State University, Boise, ID 83725-2075 USA (e-mail: haochen@boisestate.edu).

P. K. Varshney is with the Department of Electrical and Computer Engineering, Syracuse University, Syracuse, NY 13244 USA (e-mail: varshney@syr.edu).

Digital Object Identifier 10.1109/TIP.2013.2253480

In this paper, we focus on the application-specific dimensionality reduction of the data sets prior to pairwise registration to enable registration of high-dimensional data sets. In the literature, very little attention has been paid to explore the dimensionality reduction algorithm in the context of registration of high-dimensional data. Researchers usually consider the dimensionality reduction problem and the pairwise registration problem separately. Existing dimensionality reduction algorithms, such as band selection, averaging, principal component analysis (PCA) [7], [9], maximum-noise-fraction (MNF) transform, also called noise adjusted principal component transform [10], [11], Independent component analysis [12], [13] and wavelet based approaches [14] have been applied to multi-spectral or hyper-spectral data sets to reduce data dimensionality. These dimensionality reduction algorithms are mostly designed for visualization [12], target detection [13] or classification [14], not for the purpose of image registration. For example, the PCA method produces images that contain most information of the data sets, but that do not necessarily lead to higher registration accuracy. The goal of this paper is to design an application-specific dimensionality reduction algorithm, that can achieve more accurate registration performance. In particular, the dimensionality reduction algorithm attempts to minimize the bound on achievable registration performance thereby linking dimensionality reduction with registration performance. This approach is expected to yield improved registration. In Section II, we formulate the problem and introduce the notation. The new algorithm for dimensionality reduction specifically meant for registration is presented in Section III. This algorithm is an enhanced version of the algorithm presented in our preliminary work [15]. Some experimental results are provided in Section IV followed by concluding remarks in Section V.

II. Problem Description and Notation

Let the two K-band data sets to be registered be $\vec{S}_1$ and $\vec{S}_2$, each dependent on the pixel location (x, y), where, without loss of generality, we have assumed that both high-dimensional data sets are K-band datasets. Let p denote the set of transformation parameters such as translations and rotation. The process of image registration essentially involves the estimation of p. It can be modeled as an optimization problem, where p is the argument of the optimum of some similarity metric S, applied to $\vec{S}_1$ and transformed $\vec{S}_2$. This can be expressed as

$$\hat{p} = arg(opt(S(\vec{S}_1, \vec{S}_2(p)))) \tag{1}$$

In the first step, we need to convert them to grayscale images by linear combination of the bands. This process is described as

$$\begin{aligned} I_1 &= \vec{H}_1^T \vec{S}_1 \\ I_2 &= \vec{H}_2^T \vec{S}_2 \end{aligned} \tag{2}$$

where, I_1 and I_2 are the grayscale images corresponding to the two data sets and $\vec{H}_1$ and $\vec{H}_2$ are the coefficients to be calculated for dimensionality reduction. The objective is to find optimal $\vec{H}_1$ and $\vec{H}_2$ such that I_1 and I_2 can produce the best registration. The problem can then be written as [15]

$$\{\hat{\vec{H}}_1, \hat{\vec{H}}_2\} = \arg\min E\left[p - \hat{p}\left(I_1, I_2\right)\right]^2 \tag{3}$$

The traditional way to register such data sets is either to select a single band from each high-dimensional data set and then register the two selected images, or more generally to apply a linear transform such as principal component analysis (PCA), noise-adjusted principal component (NAPC) transform or maximum noise fraction (MNF) transform to the two high-dimensional data sets, respectively, and then perform registration on the two resulting images which have the largest variance [16].

If $\vec{H}_j$ $(j = 1, 2)$ takes the value $[\underset{1}{0} \cdots \underset{i-1}{0}\ \underset{i}{1}\ \underset{i+1}{0} \cdots \underset{K}{0}]^T$, the ith band is selected. This is the so called band selection algorithm. If $\vec{H}_j$ takes the value $[\frac{1}{K} \ldots \frac{1}{K}]^T$, all the bands are averaged to yield I_j, $j = 1, 2$. This is called the simple averaging algorithm. For the PCA method, the objective is to find an orthonormal transformation vector $\vec{H}_j$ such that the variance of $\vec{S}_j$ is maximized, which can be written as

$$\begin{aligned} \hat{\vec{H}}_{j\_PCA} &= \arg \max_{\vec{H}_j^T H_j = 1} Var\left(\vec{H}_j^T \vec{S}_j\right) \\ &= \arg \max_{\vec{H}_j^T \vec{H}_j = 1} \vec{H}_j^T Var\left(\vec{S}_j\right) \vec{H}_j \end{aligned} \tag{4}$$

The drawback of the PCA method is that it treats the noise level of all the bands the same, and therefore, is sensitive to the scale of the data [10]. For example, if one band is arbitrarily doubled in amplitude, its contribution to the variance of the data will be increased fourfold, and it will, therefore, be found to contribute more to the earlier eigenvalues and eigenvectors [10]. Thus, an improved PCA method, called the NAPC transform or the MNF transform, has been proposed by Lee *et al.* [10] to first weight the bands such that the noise level in each band becomes the same and then a regular PCA is employed. It is equivalent to maximizing the signal to noise ratio, which is written as

$$\hat{\vec{H}}_{j\_MNF} = \arg \max_{\vec{H}_j^T \vec{H}_j = 1} \frac{\vec{H}_j^T Var\left(\vec{S}_j\right) \vec{H}_j}{\vec{H}_j^T \Sigma_N \vec{H}_j} \tag{5}$$

where, Σ_N is the covariance of the noise. The problem with the calculation of the NAPC transformation is that the noise covariance matrix needs to be estimated.

As we can see from (5) when applied directly on the data set, the NAPC does not always produce an image pair that would produce small registration error. In this paper, we want to find an optimal registration-specific dimensionality reduction algorithm such that the resulting image produced from each data set can perform more accurate registration. This approach is developed in the next section.

For notational convenience, we drop the subscript j from $\vec{H}_j$ in the rest of this paper as both datasets are processed in the same manner.

III. Dimensionality Reduction Approach for Registration

As indicated in Eq. (3), our problem is to find an optimal $\vec{H}$ such that a registration-specific estimation-theoretic metric is minimized. However, the actual registration error can not be obtained prior to registration to enable the determination of the optimal $\vec{H}$. Therefore, an alternate means needs to be found. Image registration is essentially an estimation problem. The variance of the estimation error $E\left[p-\hat{p}(I_1,I_2)\right]^2$ is bounded by the performance bound such as the Cramer Rao lower Bound (CRLB) [17], [18] or Ziv Zakai Bound [19]. Such a performance bound for image registration provides us the best achievable performance for any image registration algorithm for the specific image pair in that no image registration algorithm can produce registration error smaller than the bound. These performance bounds, which are independent of the algorithm, can serve as a benchmark for the image dataset and their registrability. Thus, the optimal $\vec{H}$ can be obtained by minimizing the CRLB because the CRLB is relatively easy to compute and can adequately characterize the performance in the asymptotic region. The CRLB has been derived in [17] and [18] for the image registration problem for different misalignment cases. Note that there are different CRLBs for each parameter. We will use the results from [17], [18] here to develop the dimensionality reduction algorithm.

Usually there are multiple transformation parameters to be estimated. It is not feasible to obtain a single solution that simultaneously minimizes each CRLB in general. The most intuitive approach is to combine all of the parameters' CRLBs into a single functional form. The well-known combinations are arithmetic mean, geometric mean and harmonic mean. The harmonic mean of a set of numbers tends strongly toward the least elements of the list, it tends (compared to the arithmetic mean) to mitigate the impact of large outliers and increase the impact of small ones [20]. Thus, in the case of multiple transformation parameters, we choose to find an optimal $\vec{H}$ such that the weighted harmonic mean of the CRLBs of all the parameters is minimized. In practice, we can actually obtain a single solution when we use the weighted harmonic mean [20] of the CRLBs of all the parameters as the objective function. The weights are chosen such that the CRLBs of all the parameters have the same order of magnitude. Thus, assuming n transformation parameters to be estimated and the weight is w_i, the general dimensionality reduction algorithm can be written as

$$\hat{\vec{H}}=\arg\min\frac{w_1+\cdots+w_n}{\frac{w_1}{CRLB(p_1)}+\frac{w_2}{CRLB(p_2)}+\cdots+\frac{w_n}{CRLB(p_n)}} \tag{6}$$

We will assume an additive noise structure

$$\vec{S}=\vec{F}+\vec{N} \tag{7}$$

We also assume that $\vec{F}$ and $\vec{N}$ are uncorrelated signal and noise components. Thus

$$Var\vec{S}=\Sigma_S=\Sigma_F+\Sigma_N \tag{8}$$

where Σ_F and Σ_N are the covariance matrices for $\vec{F}$ and $\vec{N}$ respectively. Dimensionality reduction produces the single image as $I=\vec{H}^T\vec{S}$. The covariance of the resulting image I can be obtained by $\vec{H}^T\Sigma_S\vec{H}=\vec{H}^T\Sigma_F\vec{H}+\vec{H}^T\Sigma_N\vec{H}$, where $\vec{H}^T\Sigma_N\vec{H}$ is the covariance of the noise component.

More specifically, in the simplest case, when there is only one parameter to be estimated, the dimensionality reduction algorithm in Eq. (6) is reduced to

$$\hat{\vec{H}}=\arg\min CRLB\left(p_1\right) \tag{9}$$

The CRLB of estimate $\hat{p}_1$ is given by [17] as

$$\begin{aligned}CRLB\left(p_1\right)&=\frac{\vec{H}^T\Sigma_N\vec{H}}{\frac{\partial\left\{\vec{H}^T\vec{S}\right\}}{\partial p_1}\frac{\partial\left\{\vec{H}^T\vec{S}\right\}}{\partial p_1}^T}\\&=\frac{\vec{H}^T\Sigma_N\vec{H}}{\vec{H}^T\frac{\partial\vec{S}}{\partial p_1}\frac{\partial\vec{S}^T}{\partial p_1}\vec{H}}\end{aligned} \tag{10}$$

Substituting Eq. (10) into Eq. (9), we have

$$\begin{aligned}\hat{\vec{H}}&=\arg\min_{\vec{H}^T\vec{H}=1}\frac{\vec{H}^T\Sigma_N\vec{H}}{\vec{H}^T\frac{\partial\vec{S}}{\partial p_1}\frac{\partial\vec{S}^T}{\partial p_1}\vec{H}}\\&=\arg\max_{\vec{H}^T\vec{H}=1}\frac{\vec{H}^T\left(\frac{1}{M}\frac{\partial\vec{S}}{\partial p_1}\frac{\partial\vec{S}^T}{\partial p_1}\right)\vec{H}}{\vec{H}^T\Sigma_N\vec{H}}\end{aligned} \tag{11}$$

where, M is the total number of pixels in one band. Let R denote the correlation matrix, which is defined as $R=\frac{1}{M}\frac{\partial\vec{S}}{\partial p_1}\frac{\partial\vec{S}^T}{\partial p_1}$. It yields

$$\hat{\vec{H}}=\arg\max_{\vec{H}^T\vec{H}=1}\frac{\vec{H}^TR\vec{H}}{\vec{H}^T\Sigma_N\vec{H}}. \tag{12}$$

To solve Eq. (12), we first want to find the linear transformation $\vec{D}$ such that

$$\vec{D}^T\Sigma_N\vec{D}=I \tag{13}$$

where, I is the identity matrix. Let $\vec{E}=\vec{D}^{-1}\vec{H}$, Eq. (12) is equivalent to maximization of

$$\frac{\vec{E}^T\vec{D}^TR\vec{D}\vec{E}}{\vec{E}^T\vec{E}} \tag{14}$$

It is well known [10] that the principal components of $\vec{D}^TR\vec{D}$ maximize (14). Thus the solution of (12) can be obtained by

$$\hat{\vec{H}}=\vec{D}\vec{E} \tag{15}$$

Further, we observe that the solution of (12) is essentially performing a noise-adjusted principal component (NAPC) transform on the first derivative of images $\partial\vec{S}/\partial p_1$. That is equivalent to a two-stage transformation in which the data $\partial\vec{S}/\partial p_1$ are first transformed so that the noise covariance matrix is transformed to an identity matrix and the second stage is the conventional PCA. Then, the multiplication of the principal eigenvector that corresponds to the maximal eigenvalue of the noise adjusted correlation matrix and the transformation that normalizes the noise covariance matrix is our estimate of the gain vector. For different misalignment cases, the optimal dimensionality reduction algorithm is equivalent to performing PCA on the noise-adjusted correlation matrices corresponding to misalignment models. These different form of correlation matrices for different misalignment cases will be presented below.

When the data sets contain only translation distortion along the x axis x_0, the misalignment error in this case has the mapping function

$$\begin{aligned}u\left(x,y\right)&=x+x_0\\v\left(x,y\right)&=y\end{aligned} \tag{16}$$

where, u and v are the transformed coordinates and x and y are the original coordinates. The correlation matrix, in this case, is given by

$$R = \frac{1}{M}\frac{\partial \vec{S}}{\partial x}\frac{\partial \vec{S}^T}{\partial x}. \tag{17}$$

When the data sets only have a constant translation along the y axis y_0, the mapping function in this case is

$$\begin{aligned} u(x, y) &= x \\ v(x, y) &= y + y_0 \end{aligned} \tag{18}$$

And the correlation matrix is given by

$$R = \frac{1}{M}\frac{\partial \vec{S}}{\partial y}\frac{\partial \vec{S}^T}{\partial y}. \tag{19}$$

When the data sets contain misalignment of translations along both axes, the mapping function is

$$\begin{aligned} u(x, y) &= x + x_0 \\ v(x, y) &= y + y_0 \end{aligned} \tag{20}$$

There are two parameters to be estimated, x_0 and y_0. The CRLBs for the estimation of x_0 and y_0 usually have the same order of magnitude, and therefore, they are equally weighted. By minimizing the harmonic mean of the CRLB for the estimation of the two parameters, the estimates of $\vec{H}$ can be obtained by Eq. (12) with the correlation matrix given by

$$R = \frac{1}{M}\frac{\partial \vec{S}}{\partial x}\frac{\partial \vec{S}^T}{\partial x} + \frac{1}{M}\frac{\partial \vec{S}}{\partial y}\frac{\partial \vec{S}^T}{\partial y}. \tag{21}$$

In the case of rotation only distortion case, the transformation parameter is θ_0, and the mapping function is

$$\begin{aligned} u(x, y) &= \cos\theta_0 x + \sin\theta_0 y \\ v(x, y) &= -\sin\theta_0 x + \cos\theta_0 y \end{aligned} \tag{22}$$

Therefore, the correlation matrix is given by

$$\begin{aligned} R &= \frac{1}{M\Delta_\theta}\int_0^{\Delta_\theta} A d\theta_0 \\ A &= \left(\frac{\partial \vec{S}}{\partial x} \circ \frac{dU}{d\theta_0} + \frac{\partial \vec{S}}{\partial y} \circ \frac{dV}{d\theta_0}\right)\left(\frac{\partial \vec{S}}{\partial x} \circ \frac{dU}{d\theta_0} + \frac{\partial \vec{S}}{\partial y} \circ \frac{dV}{d\theta_0}\right)^T \end{aligned} \tag{23}$$

where $\circ$ denotes the element-by-element multiplication matrix operation, $dU/d\theta_0(i, (x, y)) = -\sin\theta_0 x + \cos\theta_0 y$ and $dV/d\theta_0(i, (x, y)) = -\cos\theta_0 x - \sin\theta_0 y$.

In the case of general rigid body transformation distortion, the mapping function is

$$\begin{aligned} u(x, y) &= \cos\theta_0 x + \sin\theta_0 y + x_0 \\ v(x, y) &= -\sin\theta_0 x + \cos\theta_0 y + y_0 \end{aligned} \tag{24}$$

There are three transformation parameters x_0, y_0 and θ_0. The correlation matrix is given by

$$R = w\frac{1}{M}\frac{\partial \vec{S}}{\partial x}\frac{\partial \vec{S}^T}{\partial x} + w\frac{1}{M}\frac{\partial \vec{S}}{\partial y}\frac{\partial \vec{S}^T}{\partial y} + (1-2w)\frac{1}{M\Delta_\theta}\int_0^{\Delta_\theta} A d\theta_0 \tag{25}$$

where A is given in Eq. (23); the weight w is chosen such that the CRLB for the estimation of translation misalignment has the same order of magnitude as that of rotation misalignment.

Skew distortion is a non-rigid transformation with the mapping function

$$\begin{aligned} u(x, y) &= x + c_x y \\ v(x, y) &= c_y x + y \end{aligned} \tag{26}$$

The transformation parameters are c_x and c_y. The correlation matrix is given by

$$\begin{aligned} R = &\frac{w_1}{M}\left[\frac{\partial \vec{S}}{\partial x} \circ U_y\right]\left[\frac{\partial \vec{S}}{\partial x} \circ U_y\right]^T \\ &+ \frac{w_2}{M}\left[\frac{\partial \vec{S}}{\partial x} \circ U_x\right]\left[\frac{\partial \vec{S}}{\partial x} \circ U_x\right]^T \end{aligned} \tag{27}$$

where, $U_x(i, (x, y)) = x$ and $U_y(i, (x, y)) = y$.

Affine transformation distortions are the most general form of geometric deformation mapping parallel lines to parallel lines [18]

$$\begin{aligned} u(x, y) &= c_1 x + c_2 y + c_3 \\ v(x, y) &= c_4 x + c_5 y + c_6 \end{aligned} \tag{28}$$

The transformation parameters are $c_1, \ldots c_6$. In this case, the correlation matrix is obtained by

$$\begin{aligned} R = &\frac{w_1}{M}\left[\frac{\partial \vec{S}}{\partial x} \circ U_x\right]\left[\frac{\partial \vec{S}}{\partial x} \circ U_x\right]^T \\ &+ \frac{w_2}{M}\left[\frac{\partial \vec{S}}{\partial x} \circ U_y\right]\left[\frac{\partial \vec{S}}{\partial x} \circ U_y\right]^T \\ &+ \frac{w_3}{M}\left[\frac{\partial \vec{S}}{\partial y} \circ U_x\right]\left[\frac{\partial \vec{S}}{\partial y} \circ U_x\right]^T \\ &+ \frac{w_4}{M}\left[\frac{\partial \vec{S}}{\partial y} \circ U_y\right]\left[\frac{\partial \vec{S}}{\partial y} \circ U_y\right]^T \\ &+ \frac{w_5}{M}\frac{\partial \vec{S}}{\partial x}\frac{\partial \vec{S}^T}{\partial x} + \frac{w_6}{M}\frac{\partial \vec{S}}{\partial y}\frac{\partial \vec{S}^T}{\partial y} \end{aligned} \tag{29}$$

From the above, we observe that the correlation matrices for different misalignment cases are essentially the correlation matrices of linear combination of gradients of data sets. Performing PCA on the noise adjusted correlation matrices of gradients of the data set is to find the linear transformation for keeping the subspace that has the largest correlation of gradients of the data set. Images with small gradients, i.e., a smooth texture, such as those containing lakes and sky, are difficult to register resulting in a lower registration accuracy and registration of textured images with many edges will achieve higher registration accuracy. Thus, the result produced by our dimensionality reduction algorithm will keep the most gradient information of the original data set and, therefore, is expected to yield higher registration accuracy. The central problem in this algorithm is the estimation of the covariance matrix of the noise Σ_N. We used the approach mentioned in [11] to estimate the noise as the residual in a simultaneous autoregression (SAR) model involving the W and N neighbors of the current pixel. The complete procedure in the general case is described as follows:

Step 1: For each data set, compute the correlation matrix R based on the distortion model considered.

Step 2: Estimate the noise as the residual in a SAR model that describes a pixel in terms of its N, W neighbors, and then calculate the covariance matrix of the noise Σ_N.

Step 3: Find the transformation $\vec{D}$ such that $\vec{D}^T \Sigma_N \vec{D} = I$.

Step 4: Perform PCA on $\vec{D}^T R \vec{D}$, and obtain the eigenvector $\vec{E}$ that corresponds to the maximum eigenvalue of the data matrix.

Step 5: Let $\vec{H} = \vec{D}\vec{E}$.

Step 6: Obtain the two images by Eq. (3).

Step 7: Perform pairwise image registration.

Our proposed registration approach for the high-dimensional data has the following advantages. First, it is fast since intensity based image registration algorithm is performed only on a single image pair. Second, it is accurate because the image pair we obtained by our dimensionality reduction algorithm is application-specific that is designed to yield good registration performance. Third, it can be applied to the data sets with different modalities, and therefore, it is quite robust.

IV. Experimental Results

We provide three examples using remote sensing datasets to illustrate and evaluate the registration performance of our proposed algorithm using two hyperspectral data sets and one multispectral data set. For comparison purposes, we employ five other dimensionality reduction algorithms to obtain two single-band images prior to pairwise registration.

1) "Averaging": Take the average of all the bands of the data set.
2) "Selection": Manually select the band with the highest contrast from each data set.
3) "Entropy": Select the band with the largest entropy from each data set.
4) "FA": Perform factor analysis algorithm to reduce the dimension to 1 for each data set.
5) "PCA_orig": Perform PCA directly on the noise adjusted intensities of the data set (NAPC algorithm).
6) "PCA_tran": Our proposed dimensionality reduction algorithm.

After two images are obtained from the two data sets, respectively, we apply the mutual information (MI) based image registration algorithm [5] to register the two images. This pairwise MI based image registration algorithm attempts to find the transformation parameters which maximize the mutual information between the two images. The mutual information between the two images I_1 and I_2 is defined as

$$MI\,(I_1, I_2) = \sum f\,(I_1, I_2) \log \left(\frac{f\,(I_1, I_2)}{f\,(I_1)\, f\,(I_2)} \right) \tag{30}$$

where $f\,(I_1, I_2)$ is the joint probability density function of the images I_1 and I_2, and $f\,(I_1)$ and $f\,(I_2)$ are the marginal probability density functions of the images I_1 and I_2, respectively.

TABLE I
RATIO OF CRLB WITH ORIGINAL DATA SET AND THAT WITH DIFFERENT ALGORITHMS

Averaging	Selection	Entropy
0.00675	0.0031	0.0172
FA	PCA_orig	PCA_tran
0.0165	0.1611	0.8205

TABLE II
REGISTRATION ERROR IN TRANSLATION ALONG X-AXIS-ONLY CASE FOR EXPERIMENT 1 (SAME MODALITY)

Averaging	Selection	Entropy
6.4e–5	2.5e–5	2.0e–5
FA	PCA_orig	PCA_tran
2.2e–5	1.3e–4	1.0e–5

More details of the MI based image registration algorithm can be found in [5].

A. Experiment 1

For experimental verification, we use a hyperspectral data set available in the HYDICE dataset. We first evaluate the registration performance for the data sets with the same modality. We add two realizations of Gaussian noise to the first 90 bands to emulate two data sets with the same modality at SNR level of 10 dB. We then implement the six dimensionality reduction algorithms and use the MI based registration approach [5] to estimate the transformation parameters for different misalignment cases.

As mentioned in an earlier section, CRLB is known to be a good measure of registrability of the data. We apply the result of CRLB for pairwise image registration [17], [18] to the case of registration of high-dimensional data sets in this section. We calculate the ratio of harmonic mean of CRLBs of the original high-dimensional data set and that of the image with the reduced dimension. This ratio can be used as a measure of registrability remaining in the data after dimensionality reduction. The closer the ratio is to 1, the less loss of registrability dimensionality reduction has on the data. Table I shows the values of this measure after using the six dimensionality reduction algorithms for the case of translation estimation along x axis. It is observed that the use of our procedure yields the largest value of the ratio. This implies that our dimensionality reduction algorithm results in the closest CRLB compared with the CRLB corresponding to the original data set and, therefore, our proposed algorithm preserves the registrability of the original data set the most. The whole process of dimensionality reduction and pairwise registration is repeated 100 times. Then we compute the mean square error (MSE) of the translation and rotation estimates. Tables II–IV show the MSE results for different misalignment cases. PCA_tran algorithm is observed to produce the smallest registration error among the six dimensionality reduction algorithms in all the misalignment cases, which demonstrates the superior performance of our proposed algorithm.

TABLE III
REGISTRATION ERROR IN TRANSLATION ALONG BOTH AXES CASE FOR EXPERIMENT 1 (SAME MODALITY)

Averaging		Selection	
x_0	y_0	x_0	y_0
6.8e–5	1.8e–6	1.9e–5	5.4e–3
Entropy		FA	
x_0	y_0	x_0	y_0
2.0e–5	1.4e–6	2.3e–5	1.2e–5
PCA_orig		PCA_tran	
x_0	y_0	x_0	y_0
1.3e–4	3.6e–6	1.2e–5	8.6e–7

TABLE IV
REGISTRATION ERROR IN RIGID BODY TRANSFORMATION CASE FOR EXPERIMENT 1 (SAME MODALITY)

Averaging			Selection		
x_0	y_0	θ_0	x_0	y_0	θ_0
6.7e–5	3.8e–5	8.3e–6	2.4e–5	5.8e–3	1.6e–5
Entropy			FA		
x_0	y_0	θ_0	x_0	y_0	θ_0
2.0e–5	2.1e–5	5.4e–6	1.5e–4	4.0e–4	9.9e–6
PCA_orig			PCA_tran		
x_0	y_0	θ_0	x_0	y_0	θ_0
1.7e–4	8.4e–5	1.8e–6	1.1e–5	8.9e–7	9.2e–7

To simulate the data sets with different modalities, we divide the whole data set into two parts to form two data sets, where the first set includes 90 bands and the second set includes 46 bands, to emulate data coming from two different modalities. Since the two data sets we consider are from the same data set, they do not have geometrical distortion between each other, which means that the true rotation and true translations along both axes are all zero, that is $\theta_0 = 0$, $x_0 = 0$, $y_0 = 0$. We again implement the six dimensionality reduction algorithms for the two data sets and then use the MI based registration approach [5] to estimate the transformation parameters for different misalignment cases. Figs. 1 and 2 show the two images obtained from the two data sets using the six dimensionality reduction algorithms. The results shown in Figs. 1(f) and 2(f) are produced by PCA_tran algorithm assuming a rigid body transformation misalignment in the two data sets. It is observed that the image obtained by PCA_tran algorithm from the first data set preserves more texture than the other algorithms. In the second data set, the resulting images obtained by PCA based algorithms look very similar. Tables V–VII give the registration errors using the six dimensionality reduction algorithms. We observe that registration using our proposed dimensionality reduction algorithm achieved the best performance, i.e., the smallest registration error. In particular, when there is only a single translation parameter to be estimated, such as the translation along x axis only distortion case, PCA_tran algorithm produces much smaller registration error than the other five dimensionality reduction algorithms. In the case of translation errors along both axes, PCA_tran algorithm still produces smaller translation error than the other five algorithms. In the case of rigid body transformation distortion case also, where there is a combination of translations and rotation distortions, PCA_tran algorithm obtains the smallest translation errors and rotation error.

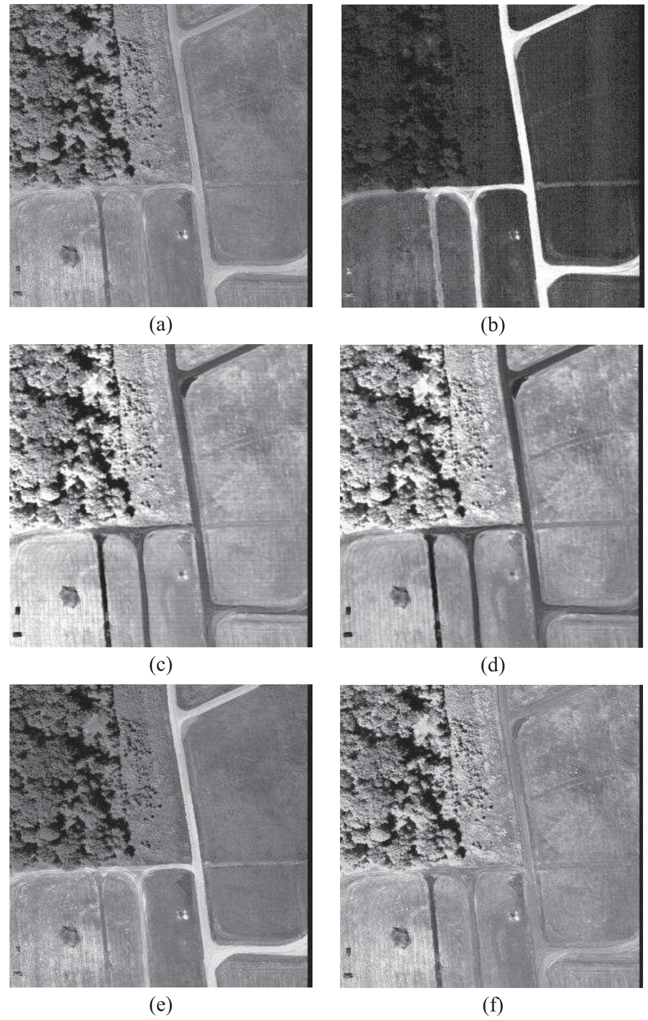

Fig. 1. Fusion results produced by the six algorithms for Data Set 1 in Experiment 1. (f) Fusion result using "PCA_tran" algorithm assuming a rigid body transformation misalignment in the data set.

B. Experiment 2

This example investigates the performance of the proposed registration algorithm for a multispectral image registration application. This set of digital images was collected by the sensors Landsat ETM+ in 1999. It is a multi-spectral image with spatial resolution 30 m. A portion consisting of (221*327) pixels covering a part of Syracuse area is selected. Six bands of data are used in our experiment. This data set contains six bands and encompasses a diverse environment, including lakes, creeks, a well-developed downtown area, residential areas, parks, golf courses, urban, forest etc.

We consider the first three bands and the other three bands as two separate data sets, and shift the second data sets along both axes by 2 pixels. We then implement the six dimensionality reduction algorithms and use the MI based registration approach [5] to estimate the transformation parameters in

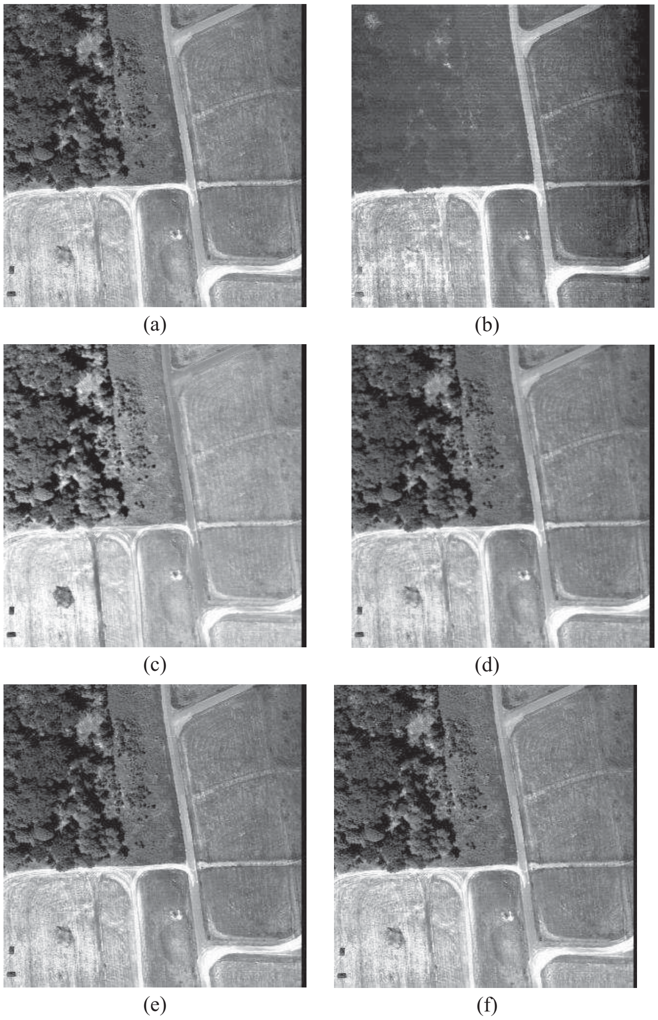

Fig. 2. Fusion results produced by the six algorithms for Data Set 2 in Experiment 1. (f) Fusion result using "PCA_tran" algorithm assuming a rigid body transformation misalignment in the data set.

TABLE V

REGISTRATION ERROR IN TRANSLATION ALONG X-AXIS-ONLY CASE FOR EXPERIMENT 1 (DIFFERENT MODALITIES)

Averaging	Selection	Entropy
6.4e–3	6.4e–3	6.4e–3
FA	PCA_orig	PCA_tran
1e–3	8.4e–4	1.7e–4

the translation distortion only case and rigid body transformation distortion case. The registration errors of the six algorithms are shown in Tables VIII and IX. We observe that registration using our proposed dimensionality reduction algorithm achieved the best performance. In particular, the translation error along y axis of PCA_tran algorithm is much smaller than that of the other five algorithms in the translation distortion only case as shown in Table VIII. Selection algorithm (manual selection) in this experiment selected the image pair with the smallest rotation error and largest translation errors. The rotation error in rigid body transformation distortion case of the other five algorithms are the same. Note that

TABLE VI

REGISTRATION ERROR IN TRANSLATION ALONG BOTH AXES CASE FOR EXPERIMENT 1 (DIFFERENT MODALITIES)

Averaging		Selection	
x_0	y_0	x_0	y_0
6.4e–3	6.4e–3	6.4e–3	6.4e–3
Entropy		FA	
x_0	y_0	x_0	y_0
6.4e–3	6.4e–3	1.0e–3	6.4e–3
PCA_orig		PCA_tran	
x_0	y_0	x_0	y_0
8.4e–4	3.6e–3	1.7e–4	2.3e–3

TABLE VII

REGISTRATION ERROR IN RIGID BODY TRANSFORMATION CASE FOR EXPERIMENT 1 (DIFFERENT MODALITIES)

Averaging			Selection		
x_0	y_0	θ_0	x_0	y_0	θ_0
6.4e–3	6.4e–3	2.0e–5	6.4e–3	6.4e–3	1.6e–4
Entropy			FA		
x_0	y_0	θ_0	x_0	y_0	θ_0
6.4e–3	6.4e–3	1.2e–5	1.0e–3	6.4e–3	4.1e–5
PCA_orig			PCA_tran		
x_0	y_0	θ_0	x_0	y_0	θ_0
8.4e–4	3.6e–3	4.1e–7	1.7e–4	2.3e–3	4.1e–7

TABLE VIII

REGISTRATION ERROR IN TRANSLATION ALONG BOTH AXES CASE FOR EXPERIMENT 2

Averaging		Selection	
x_0	y_0	x_0	y_0
2.6e–1	2.3e–3	6.4e–1	6.4e–3
Entropy		FA	
x_0	y_0	x_0	y_0
2.3e–3	2.3e–3	2.3e–3	2.3e–3
PCA_orig		PCA_tran	
x_0	y_0	x_0	y_0
2.6e–2	2.3e–3	4.1e–3	1.6e–6

in rigid body transformation distortion case, there are three transformation parameters to estimate. Our algorithm attempts to minimize the harmonic mean of the registration errors of the three parameters, which cannot guarantee a solution that minimizes the estimation error of all the three parameters. Considering the registration errors of all the three parameters, PCA_tran algorithm achieved the best overall performance.

C. Experiment 3

This experiment uses a hyperspectral image data set which is publicly available at https://engineering.purdue.edu/~biehl/MultiSpec/index.html. This data set contains 220 bands and is from the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) built by JPL and flown by NASA/Ames on June 12, 1992. The scene is over an area 6 miles west of West Lafayette, Indiana. We consider the

TABLE IX
REGISTRATION ERROR IN RIGID BODY TRANSFORMATION CASE FOR EXPERIMENT 2

Averaging			Selection		
x_0	y_0	θ_0	x_0	y_0	θ_0
2.5e–1	2.3e-3	6.4e–3	6.4–1	6.4e–3	4.1e–5
Entropy			FA		
x_0	y_0	θ_0	x_0	y_0	θ_0
2.3e–3	2.3e–3	6.4e–3	2.3–3	1.0e–3	6.4e–3
PCA_orig			PCA_tran		
x_0	y_0	θ_0	x_0	y_0	θ_0
2.6e–2	1.0e–3	6.4e–3	2.3e–3	1.0e–3	6.4e–3

TABLE X
REGISTRATION ERROR IN TRANSLATION ALONG BOTH AXES CASE FOR EXPERIMENT 3

Averaging		Selection	
x_0	y_0	x_0	y_0
2.6e–1	6.4e–3	6.4e–1	6.4e–3
Entropy		FA	
x_0	y_0	x_0	y_0
6.4e–3	6.4e–3	5.0e–4	6.4e–3
PCA_orig		PCA_tran	
x_0	y_0	x_0	y_0
6.4e–3	6.4e–3	1.6e–6	6.4e–3

TABLE XI
REGISTRATION ERROR IN RIGID BODY TRANSFORMATION CASE FOR EXPERIMENT 3

Averaging			Selection		
x_0	y_0	θ_0	x_0	y_0	θ_0
5.0e–4	1.6e–4	6.4e–3	1.0–3	6.4e–3	6.4e–3
Entropy			FA		
x_0	y_0	θ_0	x_0	y_0	θ_0
1.0e–3	6.4e–3	2.3e–3	5.0e–4	1.6e–4	6.4e–3
PCA_orig			PCA_tran		
x_0	y_0	θ_0	x_0	y_0	θ_0
6.4e–3	6.4e–3	6.4e–3	3.7e–4	1.6e–4	3.2e–3

first 110 bands and the other 110 bands as two separate data sets. We use the same procedure as that in the previous two experiments. We implement the six dimensionality reduction algorithms and perform the MI based registration approach to estimate the transformation parameters for the translation distortion only case and rigid body transformation distortion case. The registration errors of the six algorithms are shown in Tables X and XI. According to the results, our proposed algorithm achieved the smallest translation errors in both distortion cases and the second smallest rotation error in the rigid body transformation distortion case. The entropy based algorithm has the smaller rotation error while much larger translation errors compared with PCA_tran algorithm. Thus, PCA_tran algorithm achieved the best overall registration performance.

V. CONCLUSION

For remote sensing image registration with high-dimensional data sets, directly processing two data sets for intensity based registration is computationally expensive. Thus, it is desired to select two single-band images from the two data sets to perform pairwise registration. In this paper, we proposed a PCA based dimensionality reduction algorithm to obtain two images such that a registration-specific metric is minimized. Different from the traditional NAPC based dimensionality reduction algorithm that applies PCA directly on the noise adjusted intensities of data set, minimization of the CRLB resulted in PCA being applied on the noise adjusted gradients of the data set. As a result, the obtained image pair contains more texture information from the data sets and is expected to yield smaller registration error, since an image pair with smaller gradients, i.e., smooth texture, tends to produce larger registration error. Three examples using remote sensing data have demonstrated more accurate registration performance of our proposed algorithm. Our approach should be applied to multi-temporal and multi-modality data sets to further evaluate its performance. In our proposed dimensionality reduction algorithm for registration of remote sensing data, the algorithm is derived directly from the CRLB of the transformation parameter estimates, which implies that the algorithm is applicable only when the CRLB of transformation parameter estimates can be obtained. Unfortunately, it is difficult to derive the CRLB of transformation parameter estimates in some more complicated misalignment cases such as the non-rigid body transformation misalignment case. Thus, in such cases, new dimensionality reduction algorithms need to be developed in the future.

ACKNOWLEDGMENT

The authors would like to thank J. Walton of the USDA Forest Service Northeastern Research Center SUNY-ESF for sharing the remote sensing data used in Example 2 of this paper.

REFERENCES

[1] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, 1992.

[2] B. Zitov'a and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.

[3] J. L. Moigne, N. S. Netanyahu, and R. D. Eastman, *Image Registration for Remote Sensing*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[4] S. Suri and P. Reinartz, "Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 939–949, Feb. 2010.

[5] H. M. Chen, P. K. Varshney, and M. K. Arora, "Performance of mutual information similarity measure for registration of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2445–2454, Nov. 2003.

[6] W. M. Wells, III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–51, 1996.

[7] C. E. Mandujano and S. Mitra, "Cross-power spectrum phase for automated registration of multi/hyperspectral data-cubes for efficient information retrieval," in *Proc. IEEE Southwest Symp. Image Anal. Interpretat.*, Apr. 2002, pp. 111–115.

[8] W. Hoge and C. Westin, "Identification of translational displacements between n-dimensional data sets using the high-order SVD and phase correlation," *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 884–889, Jul. 2005.

[9] P. K. Varshney and M. K. Arora, *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*. New York, NY, USA: Springer-Verlag, 2004.
[10] J. Lee, A. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 3, pp. 295–304, May 1990.
[11] A. A. Nielsen and R. Larsen, "Restoration of geris data using the maximum noise fractions transform," in *Proc. 1st Int. Airborne Remote Sens. Conf. Exhibit.*, 1994, pp. 557–568.
[12] Y. Zhu, P. Varshney, and H. Chen, "Dimensionality reduction of hyperspectral images for color display using segmented independent component analysis," in *Proc. IEEE Int. Conf. Image Process.*, Sep.–Oct. 2007, pp. IV-97–IV-100.
[13] J. Wang and C. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.
[14] L. Bruce, C. Koger, and J. Li, "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2331–2338, Oct. 2002.
[15] M. Xu, H. Chen, and P. K. Varshney, "Registration of high-dimensional remote sensing data based on a new dimensionality reduction rule," in *Proc. 16th IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 581–584.
[16] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2127–2137, Sep. 2005.
[17] D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1185–1199, Sep. 2004.
[18] I. S. Yetik and A. Nehorai, "Performance bounds on image registration," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1737–1749, May 2006.
[19] M. Xu, H. Chen, and P. K. Varshney, "Ziv–Zakai bounds on image registration," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1745–1755, May 2009.
[20] I. M. Gelfand and A. Shen, *Algebra*. Cambridge, MA, USA: Birkhäuser, 1993.

Min Xu (S'07–M'10) received the B.S. degree from the University of Science and Technology of China (USTC), Hefei, China, in 2002, and the M.S. and Ph.D. degrees from Syracuse University, Syracuse, NY, USA, in 2005 and 2009, respectively, both in electrical engineering.

She has been a Researcher with Blue Highway Inc., Syracuse, since 2009. Her current research interests include statistical signal processing and image processing.

Hao Chen (S'06–M'08) received the Ph.D. degree in electrical engineering from Syracuse University, Syracuse, NY, USA, in 2007.

He was a Post-Doctoral Research Associate from 2007 to 2010 and then a Research Assistant Professor with Syracuse University. Since August 2010, he has been an Assistant Professor with the Department of Electrical and Computer Engineering and Computer Science, Boise State University, Boise, ID, USA. His current research interests include statistical signal and image processing, and communications.

Pramod K. Varshney (S'72–M'77–SM'82–F'97) received the B.S. degree in electrical engineering and computer science and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 1972, 1974, and 1976, respectively.

He has been with Syracuse University since 1976, where he is currently a Distinguished Professor of electrical engineering and computer science and the Director of CASE: Center for Advanced Systems and Engineering. His current research interests include distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing and remote sensing.

Dr. Varshney was a recipient of numerous awards including the IEEE Judith A. Resnik Award in 2012. He was the 2001 president of International Society of Information Fusion.