

# Endmember Transformation and Replacement in Real Time Hyperspectral Unmixing

Masoud Farzam, Soosan Beheshti

Ryerson University, Department of Electrical Engineering, Toronto, Canada

**Abstract**—For much of the past decade Hyperspectral Imaging (HSI) systems have gained considerable attention among researchers. Recent improvements in Optics have expanded the applications of HSI systems. Real time processing of extensive volumes of Hyperspectral data calls for more efficient and accurate real time algorithms. In current algorithms, speed comes at the expense of accuracy. Nevertheless, our proposed Ultra Fast Transition and Replacement (UFTR) approach shows a substantial improvement to the processing speed while also increasing the accuracy of the present methods. In the UFTR algorithm, Hyperspectral components' signatures, known as Endmembers, are estimated in an iterative approach. In each iteration, a linear transformation of data into the abundance vectors is calculated. This iterative process causes the speed of the algorithm to be extraordinarily fast. To improve the accuracy, a correlation based approach is used to project the estimated Endmembers into the library spectrum. Accurate abundance vectors are calculated using the final transition matrix and the chosen Endmembers from the library. UFTR simulation results show that in low-SNR applications, the accuracy can be improved up to 15% and the speed is 10 to 50 times faster compared to the existing methods for a data cube of 4096 pixel with 224 bands. Furthermore, unlike many existing approaches, UFTR processing time dependency on the noise level is quite low. UFTR is definitely a departure from the trade-off between speed and accuracy and has a great potential for applications in the real time Hyperspectral imaging.

**Keywords**—Hyperspectral imaging, Hyperspectral unmixing, Real time Hyperspectral imaging

## I. INTRODUCTION

Spatial pixel sizes for multispectral and *Hyperspectral* sensors are often large enough that numerous disparate substances can contribute to the spectrum measured from a single pixel. This is known as *mixed pixel* in remote sensing nomenclature. Consequently, many techniques have been exploited to extract constituent materials (known as *Endmember*) in the mixture, from a spectrum, as well as the proportions in which they appear (known as *Abundance*). Multispectral imaging sensors such as Landsat provided the first opportunity to derive multichannel spectral information from large scenes on a pixel-by-pixel basis. Since then, hyperspectral sensors have been developed with hundreds of spectral bands and significantly improved spectral resolution. The predecessors of present day HSI systems were the broad band space-borne multi-spectral imaging systems like MSS and the more recent TM, which provided data in 4 and 7 spectral bands respectively. Present airborne HSI systems, like AVIRIS and HYDICE, operate in the VIS-SWIR spectral region (0.4-2.5  $\mu\text{m}$ ). They acquire data at a higher spectral resolution ( $\sim 10$  nm) and produce data cubes with 224 and 210 bands respectively. Extensive spectral bands of HSI sensors result in large-sized data sets. For

example, an AVIRIS image of size 614 x 512 spatial pixels, occupies about 140 Mbytes in a 2 byte binary storage format. Ultra-spectral sensors with even higher number of channels have placed a correspondingly higher burden on data storage and data manipulation requirements, consequently calling for much faster unmixing algorithms. In this paper, we have introduced a novel method based on a double Least Square Error (LSE) method that uses a transforming matrix for fast calculation of the abundance vectors. Further we have improved the accuracy by means of a library spectrum matching process that recalculates all abundances using the exact data from the library. The concept of Spectral Information Divergence (SID) [2] has been employed to evaluate our method and to compare UFTR with other unmixing approaches.

### A. Linear Mixture Model

In multispectral/hyperspectral imagery, a pixel is generally mixed by a number of materials present in the scene. Two well known models have been proposed in the past [6] to describe such mixing activities. In the macroscopic mixture, a mixed pixel is modeled as a linear combination of materials with relative concentrations. An intimate spectral mixture, is a nonlinear model of mixing materials [6]. Nevertheless, this model can be linearized by a method proposed in [7]. We will consider the linear spectral mixture analysis (LSMA) here due to its effectiveness and simplicity. Observation data on this model for the  $i$ th pixel is formulated as:

$$x_i = A_c S_i + w_i \quad (1)$$

where the elements of  $x_i \in R^l$  are the measured solar radiation at different spectral bands.

$A_c = [a_1 \ a_2 \ \dots \ a_c]$  is an  $\ell \times c$  source matrix (or material signature matrix) with each column  $a_j$  being the spectral signature of endmember  $j$ . The abundance vector  $S_i = [s_{i1}, s_{i2}, \dots, s_{ic}] \in R^c$  consists of the mixing coefficients for the  $i$ th pixel. The last term  $w_i$  takes into account possible errors and sensor noises. We consider a zero-mean Gaussian noise. In order for a LSE-based method to provide accurate and reliable estimates of abundance fractions for material quantification, two constraints must be imposed on the abundance fractions of materials in a pixel: 1) Abundance Sum-to-one Constraint (ASC)  $\sum_{j=1}^c s_{ij} = 1$  and 2) Abundance Nonnegativity Constraint (ANC)  $s_{ij} \geq 0$  for all  $1 < j < c$ . While the ASC is easy to deal with, the ANC is difficult to implement in practice, since the ANC results in a set of inequalities and cannot be solved analytically. In this case, we must rely on numerical methods for optimal solutions. Due

to such mathematical intractability, many LSE-based methods are unconstrained and can produce only suboptimal solutions, e.g., the subspace projection approach [10]. Furthermore, the abundance fractions estimated by the unconstrained LSE do not generally reflect the true and accurate abundance fractions. As a result, they cannot be used for material quantification. Some efforts were made to solve fully constrained linear mixing problems. However, the approaches used to implement these constraints were designed mainly for a small number of material signatures. In [8] a constrained least square solution is obtained by solving an overdetermined system that consisted of  $\ell$  equations with  $c$  unknowns ( $c < \ell$ ), where  $\ell$  is the number of bands and  $c$  is the number of signatures. Since there are no closed-form solutions, one must examine possible solutions in a feasible region bounded by the ASC and ANC. The use of quadratic programming techniques to impose the ASC and ANC was investigated in [9]. Nevertheless, the algorithm used was computationally expensive. Other methods presented also suffered from excessive computational complexity as the number of materials increased. In another note, LSMA-based methods require a priori knowledge of the signatures of materials (endmembers) present in the image scene, which is generally not available. Thus, selection of an appropriate set of material signatures is crucial for successful performance of any LSMA-based method. In the ideal case, these signatures would represent pure spectral signatures of all materials in the image scene. Unfortunately, this is rarely true in practical situations since all the signatures are generally obtained directly from the image scene, in either a supervised or an unsupervised fashion. Retrieving the abundance vectors generally needs the application of the condition of identifiability". This condition basically means that the number of linear independent endmembers cannot exceed the number of linear independent bands. Considering the improvements recently made in optics and imaging systems, the number of bands is always greater than the number of components to be identified, so the condition of identifiability" is mostly met. The UFTR approach uses this fact along with the fact that the sensor measurement  $y_i$  is actually a convex combination of endmember signatures. This fact indicates that in the  $c - 1$  hyperspace, all the mixtures are within a simplex, whose vertices correspond to the endmembers given in the model in 1. This fact makes the UFTR approach valid when using the concept of endmember transformation and invalidates the limitation discussed in [11] about the limited number of bands. To clarify the concept of endmember transformation, Fig. 1 has been provided. We demonstrate a 3D example and will extend the conclusion to multi-dimensional cases. Consider the case that  $x_1$  is a linear combination of three endmembers  $a_1, a_2, a_3$ , such that  $x_1 = \sum_{i=1}^3 a_i s_{1i}$  as shown in Fig. 1. In the absence of noise, vector  $x_1$  will fall on the convex hull formed by the three endmembers. But in reality, the noisy data vector  $x_1$  would fit somewhere in a scattered area (shown as the star marked area in the figure) to simulate the effect of noise. The abundance vector for the data vector  $x_1$  is  $[s_1 \ s_2 \ s_3]^T$ . In this case for

the abundance vector the LSE solution is:

$$\begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = T_3 x_1$$

where:

$$T_3 = (A'_3 A_3)^{-1} A'_3 \quad , \quad A_3 = [a_1 \ a_2 \ a_3] \quad (2)$$

In UFTR, the process of finding a transformation matrix is performed iteratively. The endmember matrix  $A_\kappa$  in  $\kappa$ th iteration is calculated as:

$$A_\kappa = [a_1 \ a_2 \ \dots \ a_\kappa] \quad \kappa < C < \ell \quad (3)$$

Where  $C$  is the number of maximum endmembers available in the spectrum library. In fact,  $A_\kappa$  is a subset of complete library spectrum  $A_C$  in  $\kappa$ th iteration encompassing  $\kappa$  endmembers in the spectrum library. Product of transform matrix and the  $i$ th pixel data  $x_i$  would result to the estimated abundance vector for that pixel in  $\kappa$ th iteration. Transformation matrix always exists regardless of the structure of hyperspectral sensor. Since the transformation matrix is calculated only once for all pixels, the projection process is much faster than the other existing methods. Simulation results has shown that UFTR improves the processing speed up to 50 times faster than some well known unmixing methods. In next section, we will explain how UFTR would be utilized to expedite the process of unmixing.

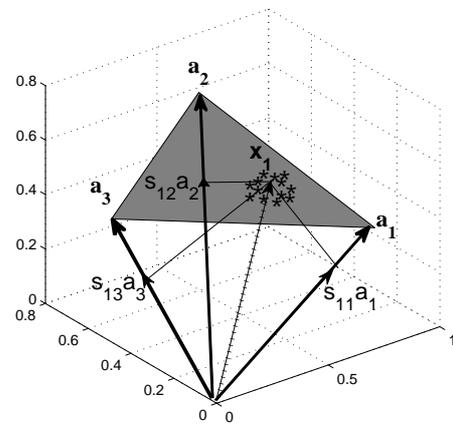


Fig. 1. 3D example of endmember transformation

## II. UFTR UNMIXING ALGORITHM

In this section we describe the UFTR algorithm in four different steps as follows:

### A. Step 1: LSE Based Extraction Method

In order to detect endmembers from the given hyperspectral data cube, we have adopted the LSE-based method in [3] to maximize the fitness between the sensor measurement and the

estimation. The algorithm starts with certain initial endmembers, then finds the most possible abundance distributions of all image pixels based on the constrained least squares. A new endmember is selected as the pixel that generates the largest residual. To suppress the noise effect, it has been proposed to take the average of a set of pixels as the new estimate. This process is repeated until the desired number of endmembers is identified or the predetermined error tolerance is reached. The method starts from initialization of the first two endmembers ( $\kappa = 1, 2$ ). As part of the initialization step, the first endmember is selected as the pixel with the largest magnitude; that is

$$k_1 = \arg \max_i \|x_i\| \quad \hat{a}_1 = x_{k_1} \quad (4)$$

where the operator  $\|\cdot\|$  is the Euclidean norm and  $\hat{a}_i$  is the  $i$ th estimated endmember. The theoretical reason behind choosing this pixel as the first endmember is that no convex combination can yield a vector that is longer than the individual components, thus it must correspond to one of the purest pixels. The second endmember is chosen as the pixel that is the most distinct from  $\hat{a}_1$  based on the Euclidean distance measure:

$$k_2 = \arg \max_i \|\hat{a}_1 - x_i\| \quad \hat{a}_2 = x_{k_2} \quad (5)$$

Further, we calculate the first estimated endmember matrix and abundance vector using the LSE solution:

$$\hat{A}_2 = [\hat{a}_1 \ \hat{a}_2] \quad (6)$$

$$T_2 = (\hat{A}'_2 \hat{A}_2)^{-1} \hat{A}'_2 \quad (7)$$

$$S_{i,2} = T_2 x_i \quad (8)$$

Given the two initial endmembers, the UFTR algorithm is applied to all pixels to find the abundances related to each pixel using a double-LSE based method. In each iteration, the estimated endmember matrix  $A_\kappa$  is calculated by augmenting the previous endmember matrix and the new estimated endmember:

$$\hat{A}_\kappa = [\hat{A}_{\kappa-1} \ \hat{a}_\kappa] \quad (9)$$

In the  $\kappa$ th iteration, the new endmember  $\hat{a}_\kappa$  is selected from the pixels yielding the largest residual (First LSE solution). Maximum residual based endmember extraction is implemented as:

$$i^* = \arg \max_{i \in 1..N} \|x_i - \hat{A}_{\kappa-1} S_{i,\kappa-1}\| \quad (10)$$

$$\hat{a}_\kappa = x_{i^*} \quad (11)$$

### B. Step 2: Second LSE ( $\kappa > 2$ )

In previous step we calculated  $A_\kappa$ . The goal is to find the related abundance vectors using the transformation matrix. In the second step, the transformation matrix will be extracted from the second LSE solution:

$$S_{i,\kappa} = \arg \min_{S_{i,k}} \|x_i - A_\kappa S_{i,k}\|^2 \quad (12)$$

where:

$$S_{i,\kappa} = T_\kappa x_i \quad (13)$$

$$T_\kappa = (\hat{A}'_\kappa \hat{A}_\kappa)^{-1} \hat{A}'_\kappa \quad (14)$$

Calculation of the endmembers and abundance vectors will be continued iteratively until all  $c$  endmembers are extracted from the hyperspectral data cube.

### C. Step 3: Suppressing the Noise Effect

Since the averaging scheme that is normally adopted tends to smooth out the spectral details when the noise level is low, the UFTR approach first estimates the noise level of the given data. If the signal-to-noise ratio (SNR) is higher than a predetermined threshold, only the pixel with the largest LSE is used as the new endmember. Otherwise, the method takes the average of a set of pixels with the largest LSEs as the new endmember. In this way, the noise effect is successfully suppressed when the SNR is low and the smoothness effect is avoided when the SNR is high. The estimation of SNR is given by:

$$SNR = 10 \log \frac{P_D}{P_X - P_D} \quad (15)$$

where  $P_X = E[X^T X]$ , and  $P_D = E[X^T U_c U_c^T X]$ .  $U_c$  is  $l \times c$  matrix formed by the leftmost singular vectors of singular value decomposition (SVD) of the given data matrix.

### D. Step 4: Library Spectral Matching

In LSE-Based endmember Extraction method we calculate the estimation of endmembers  $\hat{A}_\kappa$  in each iteration which is not necessarily a subset of the real spectrum library,  $A_C$ . However, it is easy to find the exact endmember match in the library as long as the reference library set is available (which is mostly the case). The result shows that the error performance will be improved at least by 5% to 10% compared to the previous cases when replacement from library was not considered. This process has been shown to be much more effective for a high noise level of  $SNR < 15$ . The process of finding the exact matched endmembers is performed by comparing the identified endmembers with the spectral library data, based on their spectral correlations. Correlation between two random variables  $V$  and  $W$  is defined as:

$$\rho_{V,W} = \frac{E(V - \mu_V)(W - \mu_W)}{\sigma_V \sigma_W} \quad (16)$$

Where  $\sigma_V, \sigma_W$  denote the standard deviation and  $\mu_V, \mu_W$  denote expected values of  $V, W$ . Using the estimation of the spectral correlation in [14], the library spectrum with the highest correlation would be selected as the improved version of the estimated endmember:

$$\rho_{a_i, \hat{a}_j} = \frac{(\hat{a}_j - \hat{\beta})^T (a_i - \beta)}{\|\hat{a}_j - \hat{\beta}\| \|a_i - \beta\|} \quad i = 1..C, \quad j = 1..c \quad (17)$$

$$improved(\hat{a}_j) = \arg \max_{a_i} \rho_{a_i, \hat{a}_j} \quad (18)$$

where  $\hat{a}_j, a_i$  correspond to the estimated spectrum and the library spectrum of each endmember. endmember vectors  $a_i$

are selected from the spectrum library  $A_C$ . The mean value of all estimated spectra and all library spectra, denoted by  $\hat{\beta} = \frac{1}{c} \sum_{i=1}^c \hat{a}_i$  and  $\beta = \frac{1}{c} \sum_{i=1}^c a_i$ , are subtracted from the original spectra to give more accurate correlation coefficients.

### III. ESTIMATING COMPONENT DIMENSIONALITY ( $c$ )

In previous sections we assumed that the number of endmembers is known a-priori. In fact, Intrinsic Dimensionality (ID) refers to the minimum number of parameters required to account for the observed properties of the data. Component dimensionality is defined by the number of components in a sample data vector. The true dimensionality of hyperspectral data is difficult to determine in practice, since its ID cannot be simply determined by the component dimensionality [5]. In general, the ID is expected to be much smaller than the component dimensionality. This is specially valid for hyperspectral imagery. A Neyman-Pearson detection theory-based eigen-threshold method, referred to as the HFC method [13], was previously developed to determine the number of endmembers in hyperspectral data. The method first calculates the sample correlation matrix  $R_{\ell \times \ell}$  and sample covariance matrix  $K_{\ell \times \ell}$  and then finds the difference between their corresponding eigenvalues.

Let  $\gamma_1 > \gamma_2 > \dots > \gamma_\ell$  and  $\lambda_1 > \lambda_2 > \dots > \lambda_\ell$  be two sets of eigenvalues generated by  $R_{\ell \times \ell}$  and  $K_{\ell \times \ell}$ , called correlation eigenvalues and covariance eigenvalues matrices, respectively. By assuming that signal sources are non-random unknown positive constants and noise is white with zero mean, we can expect that the eigenvalues in the  $l$ th spectral channel can be related by:

$$\gamma_\ell > \lambda_\ell > \sigma_{nl}^2 \quad \text{for } l = 1, \dots, c \quad (19)$$

$$\gamma_\ell \approx \lambda_\ell \approx \sigma_{nl}^2 \quad \text{for } l = c + 1, \dots, \ell \quad (20)$$

where  $\sigma_{nl}^2$  is the noise variance in the  $l$ th spectral channel.

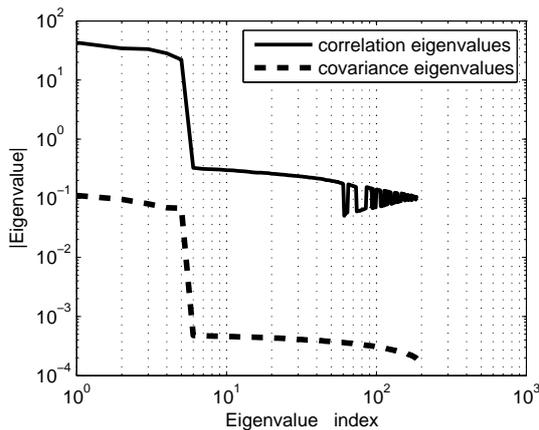


Fig. 2. The relation between Eigenvalues of correlation and covariance matrix for SNR=10 and #EM=6

Now consider the matrix  $X = [x_1 \ x_2 \ \dots \ x_N]$  as our observation data matrix. We can estimate the sample covariance matrix and sample correlation matrix for the column-vector valued random variables  $X$  to determine the dimensionality  $c$ .

Harsanyi [13] formulated the problem of determination of  $c$  as a binary hypothesis problem as follows:

$$H_0 : z_l = \gamma_l - \lambda_l \approx 0 \quad (21)$$

$$H_1 : z_l = \gamma_l - \lambda_l > 0, \quad \text{for } l = 1, 2, \dots, \ell \quad (22)$$

where the null hypothesis  $H_0$  and the alternative hypothesis  $H_1$  represent the case that the correlation eigenvalue is equal to its corresponding covariance eigenvalue and the case that the correlation eigenvalue is greater than its corresponding covariance eigenvalue, respectively. In other words, when  $H_1$  is true implies that there is an endmember contributing to the correlation eigenvalue in addition to noise, since the noise energy represented by the eigenvalue of  $R_{\ell \times \ell}$  in that particular component is same as the one represented by the eigenvalue  $R_{\ell \times \ell}$  of its corresponding component. Fig. 2 shows this fact in an example where the hyperspectral data has been created synthetically using six endmembers considering a Gaussian noise with  $SNR = 10$ . Eigenvalues have been sorted in a descending order and the horizontal axis represents the index of eigenvalues in that order. It can be seen that the sample Correlation eigenvalues drastically drop when the index becomes equal to the number of constituent endmembers. This is also true for the sample covariance matrix. This behavior originates from the fact that eigenvalues of the sample correlation matrix are an indication of the contribution of endmember components and noise. In this paper, we use this concept to calculate the number of endmembers available in the scene.

### IV. UFTR ALGORITHM

We can summarize the steps taken by UFTR to implement a fast unmixing approach in a pseudo-code as explained in Algorithm 1.

---

#### Algorithm 1: UFTR Algorithm

---

**Data:** Mixture data  $X \in \mathbb{R}^{\ell \times N}$  ( $N$  is the number of pixels,  $\ell$  is the number of spectral bands).

**Result:** The estimated source matrix  $\hat{A}_c \in \mathbb{R}^{\ell \times c}$  and the estimated abundance matrix  $S \in \mathbb{R}^{c \times N}$  where  $c$  is the number of endmembers

//Initialization

$\kappa = 2$ ;  $A_\kappa \leftarrow [a_1 \ a_2]$ , where  $a_1$  and  $a_2$  are chosen based on Eq. 4 and 5

//Main loop

while  $\kappa \leq c$  do

Set  $S_\kappa$  as a  $\kappa \times N$  zero matrix

Calculate Transformation matrix  $T_\kappa$  from 14

//Calculation of residuals

**for** every pixel  $x_i$  ( $i=1:N$ ) in the image

Calculate  $S_i = T_\kappa x_i$

Calculate the residual of pixel  $x_i$  as  $\|x_i - A_\kappa S_{i,\kappa}\|^2$

Save  $S_i$  as the  $i$ th column of matrix  $S_\kappa$

**end**

//LSE-based endmember selection

Identify a new endmember  $a_{\kappa+1}$  using  
the LSE-based method  
Find the exact match of  $a_{\kappa+1}$  in the library using the  
spectral correlation method and include it into  $A_{\kappa+1}$   
Increase  $\kappa$  by 1

**end**

$$\hat{A}_c \leftarrow A_\kappa, S \leftarrow S_\kappa$$

## V. SID EVALUATION CRITERION

The SID criterion for two pixels  $x_1 = [x_{11} \ x_{12} \dots \ x_{1\ell}]^T$  and  $x_2 = [x_{21} \ x_{22} \dots \ x_{2\ell}]^T$  is defined as:

$$SID(x_1, x_2) = D(x_1 \parallel x_2) + D(x_2 \parallel x_1) \quad (23)$$

Where:

$$D(x_1 \parallel x_2) = \sum_{i=1}^{\ell} p_i \log(p_i / q_i) \quad (24)$$

$$p_i = x_{1i} / \sum_{i=1}^{\ell} x_{1i}, \quad q_i = x_{2i} / \sum_{i=1}^{\ell} x_{2i} \quad (25)$$

It can be seen from the definition of SID that this criterion is symmetric. Lower SID values indicate lower divergence or in other words higher correlation between estimated and true measurements. Therefore, we use this value as one of the key evaluation criteria in our simulation results to compare UFTR with other methods.

## VI. SIMULATION RESULTS

In this section, we provide two different rank estimation examples. In the first example, we consider the synthetic data generated using an available spectrum library as the test endmembers and a set of random numbers as the abundance vectors. In the second example, we use the real data collected by AVIRIS [15].

### A. Experiment 1: Experimental Results Using Synthetic Data

The spectral reflectances used in the subsequent experiments are selected from the USGS digital spectral library which contains 224 spectral bands covering wavelengths ranging from 0.38  $\mu$ m to 2.5  $\mu$ m. A set of four spectral profiles is selected as the endmembers to create the mixture as shown in Fig. 3 (indicated as true EM). To create linear mixtures, we randomly selected positive abundance vectors then multiplied them to spectral endmembers and added a Gaussian noise. The resulting image was then degraded by a spatial  $\eta \times \eta$  average filter to produce mixed pixels ( $\eta$  controls the degree of mixing). With a small  $\eta$ , only the pixels close to the block boundary are mixed, so the mixture data are very likely to contain pure pixels. Using Equ. 19 and 20, we compared the eigenvalues of sample covariance matrix and correlation matrices to achieve the dimensionality of the hyperspectral data. The comparison revealed that there are four distinct components in the mixed data which match the correct value. The result of unmixing is shown in Fig. 3 for SNR=5. As we can see all four spectral endmembers have been successfully

extracted from the image. To show the abundance values we selected a gray scale mapping so the brightness of the pixels is proportional to the abundances values. We calculated the MSE error, for all  $N$  pixels, between the noiseless data and the estimated data. The MSE error was equal to 0.012. Furthermore, the results from improving the error performance showed an MSE error equal to 0.01 which indicates 15% improvement in total error. Fig. 4(a) shows a comparison of computational speed (based on Matlab code) between UFTR and three other methods. It can be seen that GDME and FCLS are 10 to 20 times slower than the UFTR approach. The VCA approach used in [4] offers close competition to UFTR. The SID evaluation criterion explained in Section V was used for this comparison. Fig. 4(b) shows how SID is changed with different SNR values for the UFTR and VCA approaches. It can be seen that the SID variation for UFTR is almost three times less than that for VCA. This indicates that estimations in UFTR are closer and correlated to the real spectrums. Fig. 4(c) shows the line of true and estimated abundances for all four endmembers in 32 consecutive pixels. It can clearly be seen that estimated abundances more closely follow the true abundances compared to results for VCA. Also the divergence in amplitude is much more for VCA. At some points it is even higher, up to 50%. This implies that UFTR is an improvement not only in speed but also in accuracy.

### B. Experiment 2: Experimental Results Using a Real Hyperspectral Scene

The real hyperspectral data used in this paper were collected by the AVIRIS sensor over Cuprite, NV. Cuprite is a relatively undisturbed acid-sulfate hydrothermal system in volcanic rocks with well exposed alteration mineralogy, located approximately 200 km northwest of Las Vegas along U.S. Highway 95. The relevant minerals consist of kaolinite, alunite, chalcedony, muscovite and other materials listed in [16].

This site has been extensively used for remote sensing experiments since the 1980s, and much research has been published [4]. The image shown in Fig. 5 is a subimage at the eastern center ( $250 \times 190$  pixels and 188 bands) of a data set acquired on the AVIRIS flight June 19, 1997, in which the noisy bands as well as the water vapor absorption bands (including bands 1, 2, 104-113, 148-167, 221-224) were removed from the original 224-band data cube to improve the detection performance.

In order to find the number of endmembers in Cuprite data, the HFC method has been used. A noise whitening process was used as a preprocessing step to remove the second-order statistical correlation. Sample correlation matrix and sample covariance matrices were calculated to perform the eigenvalue comparison. The result of comparison for the cuprite image showed an estimate of endmember quantity equal to 13 with a false-alarm rate of 0.001. Based on the ground truth provided in [16], we know that there are more than 20 materials present in the real hyperspectral scene. Analyzing the signal energy showed that the first eight eigenvalues cover 99.72% of the total signal energy. This means that most of the endmembers occur only in a small set of mixed pixels leading to

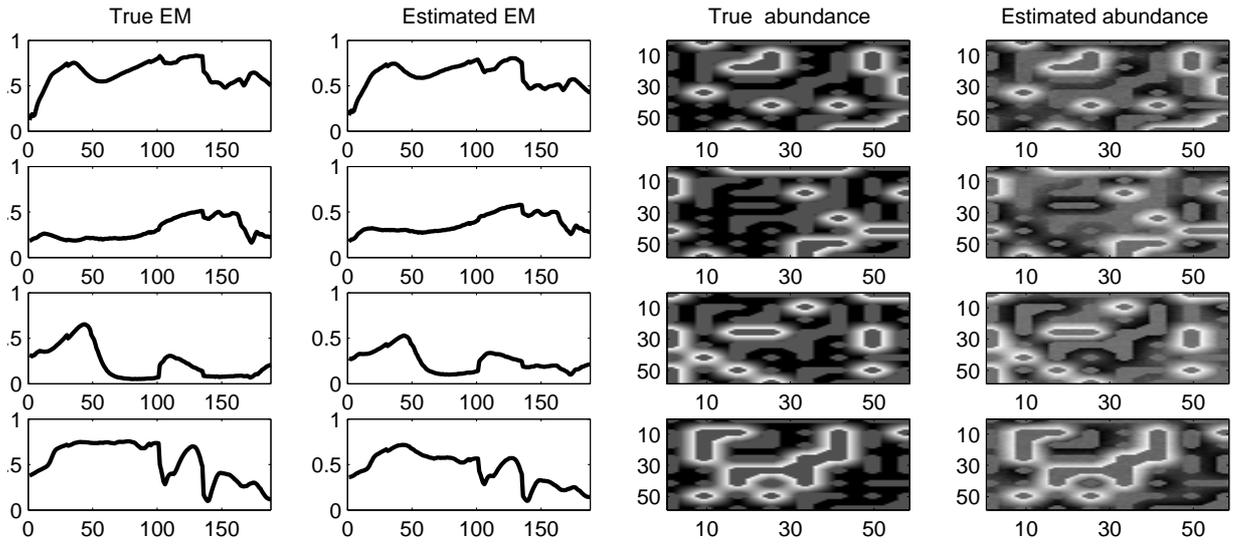


Fig. 3. UFTR estimation of endmembers and abundances

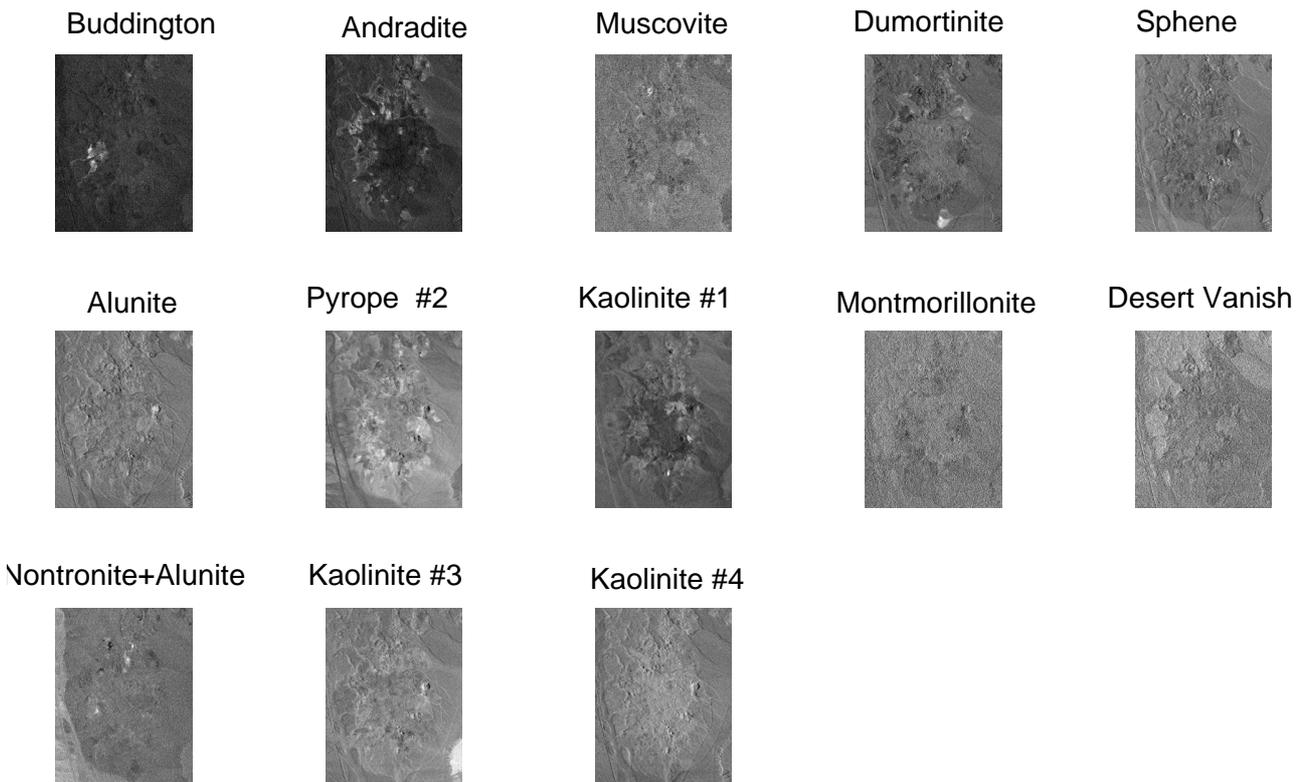


Fig. 6. Abundance fractions estimated with the UFTR algorithm

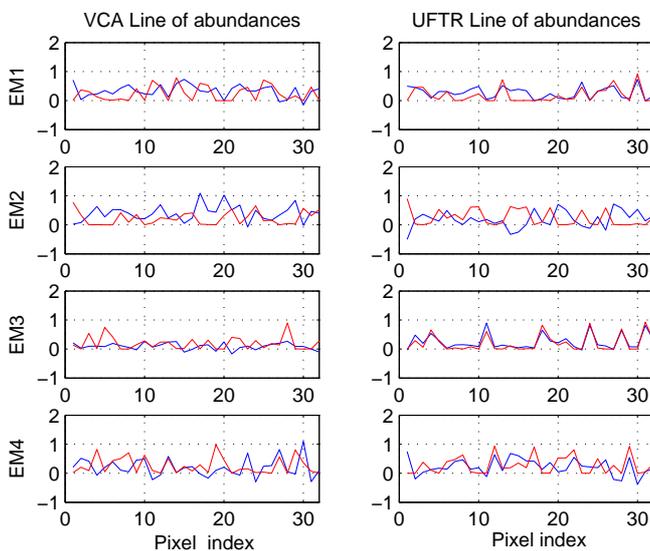
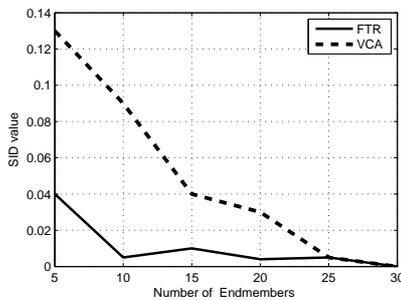
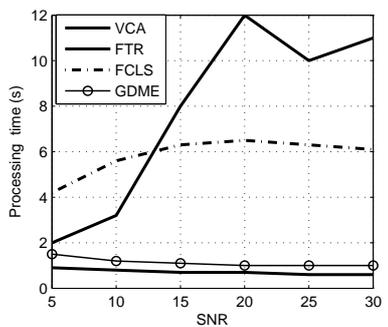


Fig. 4. a) Processing time (based on Matlab) b) SID graph c) Line of true and estimated abundances

underestimating the number of endmembers in cuprite data. Accurate estimation of the number of endmembers is very important in mixed-pixel decomposition. Underestimating the number of endmembers would lead to overestimation of the abundance of certain endmembers. On the other hand, if this value is more than the actual number of endmembers, the abundance estimate tends to spread out among extra spurious endmembers resulting in underestimation of abundances. The HFC method was able to estimate the number of endmembers in cuprite data equal to 13. Accordingly, UFTR has identified 13 signatures as shown in Figures 6 and 7. It can be seen from the figures that the estimated endmembers match the laboratory data. A visual comparison between the UFTR results on the Cuprite data set and the ground truth presented

TABLE I  
'SPECTRAL ANGLE DISTANCE BETWEEN EXTRACTED ENDMEMBERS AND LABORATORY REFLECTANCE FOR UFTR, N-FINDR, AND PPI ALGORITHMS'

	UFTR	N-Finder	PPI
Pyrope #1	3.9	4.0	4.0
Andradite	3.9	4.4	3.7
Pyrope #2	3.3	3.4	3.2
Dumortinite	4.1	4.1	5.6
Kaolinite #1	4.5	4.8	4.8
Alunite	4.0	4.0	4.2
Kaolinite #2	5.0	5.2	5.2
Buddington	2.8	-	-
Sphene	2.7	2.9	3.1
Natronite	3.9	4.0	3.8
Moscovite	5.3	5.2	6.1
Kaolinite #3	2.4	2.6	2.6

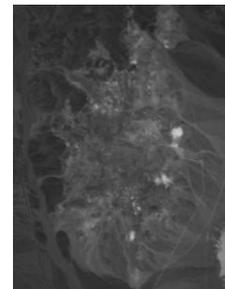


Fig. 5. Band 100 in Cuprite Site data

in [16] can identify the exact type of endmember materials matched to the library. Minerals shown in the figures match the library available for cuprite listed in [16]. In some cases the comparison of the abundance maps will be more accurate due to the close correlation of some spectral signatures. Table I compares the spectral angles between extracted endmembers and laboratory reflectance for UFTR, N-Finder [18], and pixel purity index (PPI) [17] algorithms. It can be seen from the table that with the exception of Moscovite, UFTR achieves lower spectral angles indicating a better estimation. For the case of Buddington there was no data available for the N-finder and PPI to be compared with UFTR. The displayed results follow the pattern of behavior shown in the simulations, where UFTR outperformed other methods.

## VII. CONCLUSION

A novel fast and accurate unmixing approach, UFTR, was introduced in this paper. UFTR is based on the concept of endmember transformation and spectral matching. The first concept substantially improves the processing speed while the second greatly improves the error performance. It was

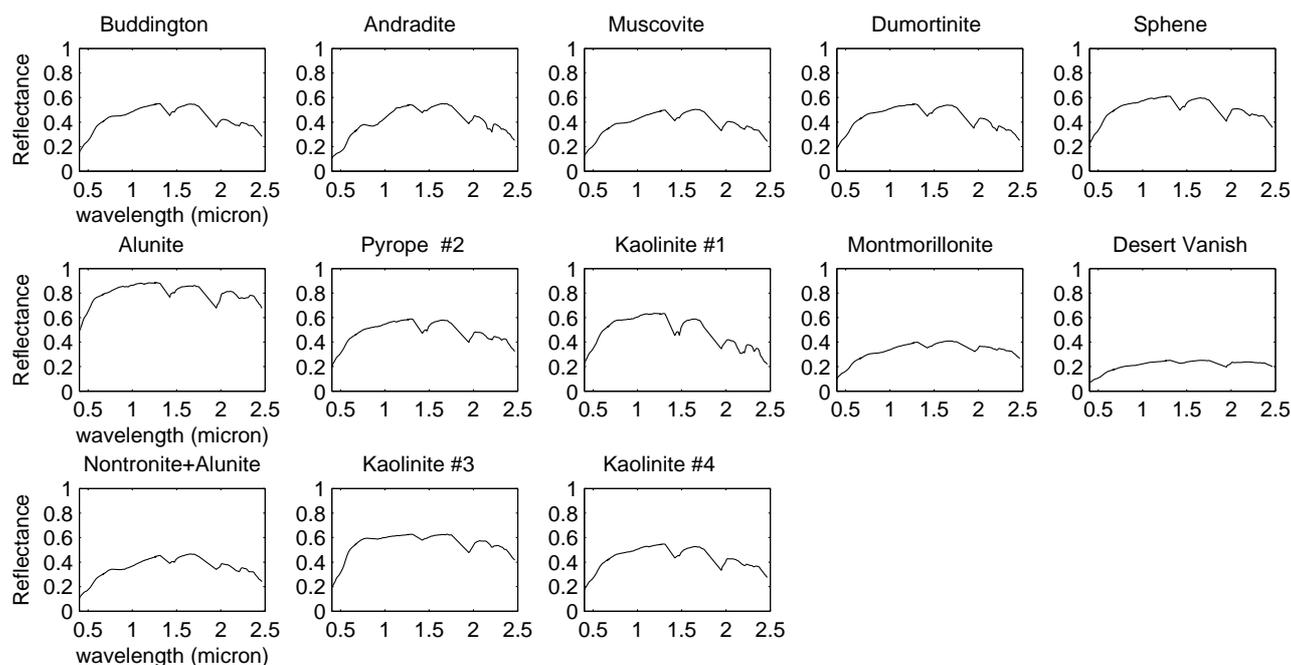


Fig. 7. Extracted signatures using UFTR algorithm

shown that existing fast approaches such as VCA estimate the pixels information with much higher spectral information divergence (SID) compared to UFTR. Speed improvement in UFTR enables this method to be used in real time hyperspectral applications. In other applications where a high level of accuracy is a requirement (e.g., medical applications), UFTR could be an alternative solution. As a future work, application of this new algorithm in non-linear mixing models could be explored and some other accurate dimension estimation methods could be used.

## REFERENCES

- [1] M. Farzam and S. Beheshti, "A fast transition and replacement (FTR) algorithm for real time hyperspectral imaging applications," *Int. Remote Sensing Conference (REMOT'E'08), Venice, Italy, Nov. 21-23, 2008*.
- [2] J. E. Ball, T. West, S. Prasad, and L. M. Bruce, "Level set hyperspectral image segmentation using spectral information divergence-based best band selection," *Geoscience and Remote Sensing Symposium, IGARSS 2007*.
- [3] L. Miao, H. Qi and H. Szu, "A maximum entropy approach to unsupervised mixed-pixel decomposition," *IEEE Trans. on Image Processing, vol. 16, No. 4, Apr. 2007*.
- [4] J. M. P. Nascimento and J. M. B. Dias, "Vertex Component Analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. on Geoscience and Remote Sensing, vol. 43, No. 4, Apr. 2005*.
- [5] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing, vol. 42, No. 3, Mar. 2004*.
- [6] B. Hapke, "Bidirectional reflectance spectroscopy," *Journal of Geophysical Res., vol. 86, pp. 3039-3054, 1981*.
- [7] P. Johnson, M. Smith, S. Taylor-George and J. Adams, "A semi-empirical method for analysis of the reflectance spectra of binary mineral mixtures," *Journal Geophys. Res., vol. 88, pp. 3557-3561, 1983*.
- [8] Y. E. Shimabukuro and J. A. Smith, "The least-squares mixing models to generate fraction images derived from remote sensing multispectral data," *IEEE Trans. Geoscience Remote Sensing, vol. 29, pp. 16-20, Jan. 1991*.
- [9] C.-I. Chang, T. L. E. Sun and M. L. G. Althouse, "An unsupervised interference rejection approach to target detection and classification for hyperspectral imagery," *Opt. Eng., vol. 37, pp. 735-743, Mar. 1998*.
- [10] C.-I. Chang, X. Zhao, M. L. G. Althouse and J.-J. Pan, "Least squares subspace projection approach to mixed pixel classification in hyperspectral images," *IEEE Trans. Geoscience Remote Sensing, vol. 36, pp. 898-912, May 1998*.
- [11] J. Luo, R. L. King and N. Younan "An unmixing algorithm based on vicinal information," *Geoscience and Remote Sensing Symposium, IGARSS 2002*.
- [12] C.-I. Chang, "Spectral information divergence for hyperspectral image analysis," *Geoscience and Remote Sensing Symposium, IGARSS 1999*.
- [13] J. Harsanyi, W. Farrand and C.-I. Chang, "Determining the number and identity of spectral endmembers," *Proc. 9th Thematic Conf. Geologic Remote Sensing, Feb. 1993*.
- [14] D. Landgrebe, "Multispectral data analysis: A signal theory perspective," *Purdue Univ., West Lafayette, Tech. Rep., 1998*.
- [15] G. Vane, R. Green, T. Chrien, H. Enmark, E. Hansen and W. Porter, "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ., vol. 44, No. 2/3, pp. 127-143, May/June. 1993*.
- [16] Swayze, "The hydrothermal and structural history of the cuprite mining district, southwestern Nevada: An integrated geological and geophysical approach," *Ph.D. dissertation, Univ. Colorado, Boulder, 1997*.
- [17] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," *Summaries 4th Annu. JPL Airborne Geoscience Workshop, vol. 1, JPL Pub. 93-26, pp. 11-14, 1993*.
- [18] M. E. Winter, "N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE Conf. Imaging Spectrometry V, pp. 266-275, 1999*.
- [19] E. Choi, S. Lee and C. Lee, "Effects of compression on the classification of hyperspectral images," *12th WSEAS International Conference on Systems, Heraklion, Greece, Jul. 22-24, 2008*.
- [20] K. Jusoff, "Advanced processing of UPM-APSBs AISA airborne hyperspectral images for individual timber species identification and

mapping;" *International Journal of Systems Applications, Engineering and Development vol 2, Issue 1, 2007.*

- [21] P. Toivanen, J. Mielikainen , "Lossless compression of 3D hyperspectral sounding data via statistical image characteristics ," *WSEAS Transactions on Systems, 3 (6), 2368-2373, 2004.*