Dependent Component Analysis for Hyperspectral Image Classification

Qian Du [1*], Ivica Kopriva [2]

[1] Department of Electrical and Computer Engineering
Mississippi State University, Mississippi State, MS 39762, USA
[2] Division of Laser and Atomic Research and Development, Rudjer Bošković Institute
Bijenička cesta 54, P.O. Box 180, 10002 Zagreb, Croatia
[*]Corresponding author: du@ece.msstate.edu

ABSTRACT

Independent component analysis (ICA) has been widely used for hyperspectral image classification in an unsupervised fashion. It is assumed that classes are statistically mutual independent. In practice, this assumption may not be true. In this paper, we apply dependent component analysis (DCA) to unsupervised classification, which does not require the class independency. The basic idea of our DCA approaches is to find a transform that can improve the class independency but leave the basis mixing matrix unchanged; thus, an original ICA method can be employed to the transformed data where classes are less statistically dependent. Linear transforms that possess such a required invariance property and generate less dependent sources include: high-pass filtering, innovation, and wavelet transforms. These three transforms correspond to three different DCA algorithms, which will be investigated in this paper. Preliminary results show that the DCA algorithms can slightly improve the classification accuracy.

Keywords: Dependent Component Analysis. Independent Component Analysis. Hyperspectral Imagery. Classification.

1. INTRODUCTION

In many real applications of remote sensing image classification, it may be very difficult or even impossible to get prior information about class signatures. So unsupervised methods have to be applied. The spatial resolution of remote sensing imagery is relatively rough, and the area covered by a single pixel is quite large. In general, there are several different materials resident in this area. So we have to deal with mixed pixels instead of pure pixels as in conventional digital image processing. Linear spectral unmixing analysis is a popularly used approach to handle mixed pixels, which assumes the pixel reflectance is the linear mixture of all the materials resident in the area covered by this pixel. Let L be the number of spectral bands and $\mathbf{r}$ a column pixel vector with dimension L in a hyperspectral image. An element r_i in the $\mathbf{r}$ is the reflectance collected at the i-th band. Let $\mathbf{M}$ denote a matrix containing p independent material spectral signatures (endmembers), i.e., $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_p]$. Let $\boldsymbol{\alpha}$ be the unknown abundance column vector of size $p \times 1$ associated with $\mathbf{M}$, which is unknown and to be estimated. The i-th item α_i in $\boldsymbol{\alpha}$ represents the abundance fraction of $\mathbf{m}_i$ in pixel $\mathbf{r}$. According to the linear mixture model,

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n} \tag{1}$$

where $\mathbf{n}$ is the noise term. When $\mathbf{M}$ is known, the estimation of $\boldsymbol{\alpha}$ can be accomplished by least squares approaches. But when $\mathbf{M}$ is also unknown, i.e., unsupervised analysis, the task is much more challenging since both $\mathbf{M}$ and $\boldsymbol{\alpha}$ need to be estimated. This can be considered as a blind source separation problem.

Independent component analysis (ICA) is a powerful tool for unsupervised classification, which has been successfully applied to blind source separation.[1] Its basic idea is to decompose a set of multivariate signals into a base of statistically independent sources with the minimal loss of information content so as to achieve detection and classification. The standard linear ICA-based data model with additive noise is

$$\mathbf{u} = \mathbf{A}\mathbf{s} + \mathbf{v} \tag{2}$$

where $\mathbf{u}$ is an L dimensional data vector, $\mathbf{A}$ is an unknown mixing matrix, and $\mathbf{s}$ is an unknown source signal vector. Three assumptions are made on the unknown source signals **s**: 1) each source signal is an independent identically distributed (i.i.d.) stationary random process; 2) the source signals are statistically independent at any time; and 3) at most one among the source signals has Gaussian distribution. The mixing matrix **A** although unknown is also assumed to be non-singular. Then the solution to the blind source separation problem is obtained with the scale and permutation indeterminacy, i.e., $\mathbf{Q} = \mathbf{W}\mathbf{A} = \mathbf{P}\Lambda$, where **W** represents the unmixing matrix, **P** is a generalized permutation matrix, and Λ is a diagonal matrix. These requirements ensure the existence and uniqueness of the solution to the blind source separation problem (except for ordering, sign, and scaling). Comparing to many conventional techniques, which use up to the second order statistics only, ICA exploits high order statistics that makes it a more powerful method in extracting irregular features in the data.

Several researchers have explored the ICA to remote sensing image classification.[2-5] In general, when we intend to apply the ICA approach to classify optical multispectral/hyperspectral images, the linear mixture model in Eq. (1) needs to be reinterpreted so as to fit in the ICA model in Eq. (2). Specifically, the pixel vector **r** is denoted as **u** in Eq. (2), the noise term **n** is denoted as **v** in Eq. (2), the endmember matrix **M** in Eq. (1) corresponds to the unknown mixing matrix **A** in Eq. (2), and abundance fractions in α in Eq. (1) correspond to source signals in **s** in Eq. (2). Moreover, the abundance fractions are considered as unknown random quantities specified by random signal sources in ICA model (2) rather than unknown deterministic quantities as assumed in the linear mixture mode (1). With these interpretations and abovementioned assumptions, we will use the model (2) to replace model (1) thereafter. The advantages offered by using model (2) in remote sensing image classification are: 1) no prior knowledge of the endmembers in the mixing process is required; 2) the spectral variability of the endmembers can be accommodated by the unknown mixing matrix **A** since the source signals are considered as scalar and random quantities; and 3) higher order statistics can be exploited for better feature extraction and pattern classification.

However, it has been argued that the assumption of source independence may not be true in many situations. For instance, endmembers may be correlated spatially and spectrally. Thus, in this paper, we will propose the use of dependent component analysis (DCA) for image restoration, which does not require that the sources are independent. The experimental result demonstrates that DCA outperforms ICA under this circumstance.

2. Derivation of DCA algorithms

Very few papers in the literature discuss the problem of DCA.[6] Here we adopt some previous studies conducted in [7][8]. The basic idea is to find a transform T that can improve the statistical independence between the sources but leave the basis matrix unchanged, i.e.,

$$T(\mathbf{G}) = T(\mathbf{A}\mathbf{S}) \cong \mathbf{A}T(\mathbf{S}). \tag{3}$$

Because the sources in this new representation space will be less statistically dependent, any standard ICA algorithm, such as FastICA, derived for the original blind source separation problem can be used to learn the basis matrix **A**. Once the basis matrix **A** is estimated, the sources **S** can recovered by applying the inverse of **A** on the multi-frame image **G** in (2).

Linear transforms that possess such a required invariance property and generate less dependent sources in the new representation space include: 1) high-pass filtering, 2) innovation, and 3) wavelet transforms. These three transforms correspond to three different DCA algorithms, which are described as follows.

High-pass filtering (HP)

A high-pass filter, such as the Butterworth high-pass filter, is applied to preprocess the observed signals **G** and then apply a standard ICA algorithm, such as JADE, on the filtered data in order to learn the mixing matrix **A**. This is motivated by the fact that high-pass filtered signals are usually more independent than original signals that include low frequency components. Meanwhile, this approach is computationally very efficient, making it attractive for DCA

problems with statistically dependent sources. In this case the transform T in Eq. (3) is the high-pass filtering operator that can be seen as a special case of the filter bank approach.[9-10]

Innovation (IN)

Another computationally very efficient approach to solve the blind separation of statistically dependent sources is based on the use of innovation. The arguments for using innovation are that they are usually more independent from each other and more non-Gaussian than original processes.[11] The innovation process is referred to as prediction error [12], which is defined as:

$$e_m(r) = s_m(r) - \sum_{i=1}^{l} b_i s_m(r-i), \quad m = 1, \cdots, M \tag{4}$$

where $s_m(r-i)$ is the i-th sample of a source process $s_m(r)$ at location $(r-i)$ and b's are prediction coefficients. $e_m(r)$ represents the new information that $s_m(r)$ has but is not contained in the past l samples. It is proved in [11] that if **G** and **S** follow the linear mixture model (2), their innovation processes $\mathbf{E_G}$ and $\mathbf{E_S}$ (in matrix form) follow the same model as well, i.e.,

$$\mathbf{E_G} = \mathbf{A}\mathbf{E_S}. \tag{5}$$

In this case the transform T in Eq. (3) is the linear prediction operator.

Subband decomposition independent component analysis (SDICA)

The SDICA approach assumes that wideband source signals can be dependent but some of their narrowband sub-components are less dependent.[9-10] Thus, SDICA extends applicability of standard ICA through the relaxation of the independence assumption. In this case the transform T in Eq. (3) is any kind of the filter-bank-like transform used to implement the sub-band decomposition scheme.

A wavelet transform-based approach to SDICA was developed in [8][13] to obtain adaptive subband decomposition of the wideband signals through the computationally efficient implementation in a form of iterative filter bank. Computationally efficient small cumulant-based approximation of mutual information is used for automated selection of the subband with the least dependent components, to which an ICA algorithm is applied. The potential disadvantage of this approach is high computational complexity if 2D wavelet transform is used for image decomposition. Hence, a reformulation can be accomplished based on dual tree complex wavelets.[14] Dual tree complex wavelets are approximately computationally efficient as decimated wavelet packets but as accurate as the shift-invariant wavelet packet approach.[15-16]

3. EXPERIMENT

The HYDICE Forest subimage scene of size 128×64 shown in Figure 1(a) was collected in Maryland in 1995 from a flight altitude of 10,000 ft with about 1.5 m spatial resolution. The spectral coverage is 0.4-2.5 μm. The water absorption bands and low SNR bands were removed reducing the data dimensionality from 210 to 169. This scene includes 30 panels arranged in a 10×3 matrix. The three panels in the same row p_{ia}, p_{ib}, p_{ic} were made from the same material of sizes $3\text{m} \times 3\text{m}$, $2\text{m} \times 2\text{m}$, and $1\text{m} \times 1\text{m}$, respectively, which can be considered as one class, P_i. The ground truth map in Figure 1(b) shows the precise locations of the panel center pixels. These panel classes have very close spectral signatures, and it is difficult to discriminate them from each other. The total number of pure panel pixels is 38.

FastICA method was chosen for ICA, and it was also used in DCA. Figures 2-5 show the classification result using FastICA, DCA (HP), DCA (IN), and DCA (SDICA), respectively. From visual inspection, all the results were comparable; there were some misclassifications between P2 and P4, P3 and P7 and P8, P4 and P6, P9 and P10. To quantify the performance, each classification map was normalized to [0 1] and converted to a binary map by changing a threshold η from 0.1 to 0.9. The accurately detected panel pixels were counted as N_D and false alarm pixels as N_F.

Table 1 lists N_D and N_F when $\eta = 0.5$. We can see that the values of N_D were similar, but NFs from DCA approaches were smaller.

Since N_D is a non-decreasing function of N_F and vice versa, we also counted the largest N_D when $N_F = 0$ and the smallest N_F when $N_D = 38$. For better performance, the former should be larger and the latter should be smaller. Table 2 lists the largest N_D when $N_F = 0$. ICA, DCA (HP), DCA (IN), and DCA (SDICA) detected 16, 20, 23 and 18 out of 38 panel pixels. By slightly decreasing the threshold η, all the 38 pixels could be detected but the false alarm was not zero any more. Table 3 lists the smallest number of N_F when all the 38 panel pixels were detected. The values of N_F from the three DCA approaches were significantly lower than that from ICA.

The three DCA approaches performed similarly. Overall, DCA (HP) provided slightly better result than the other two in this experiment.

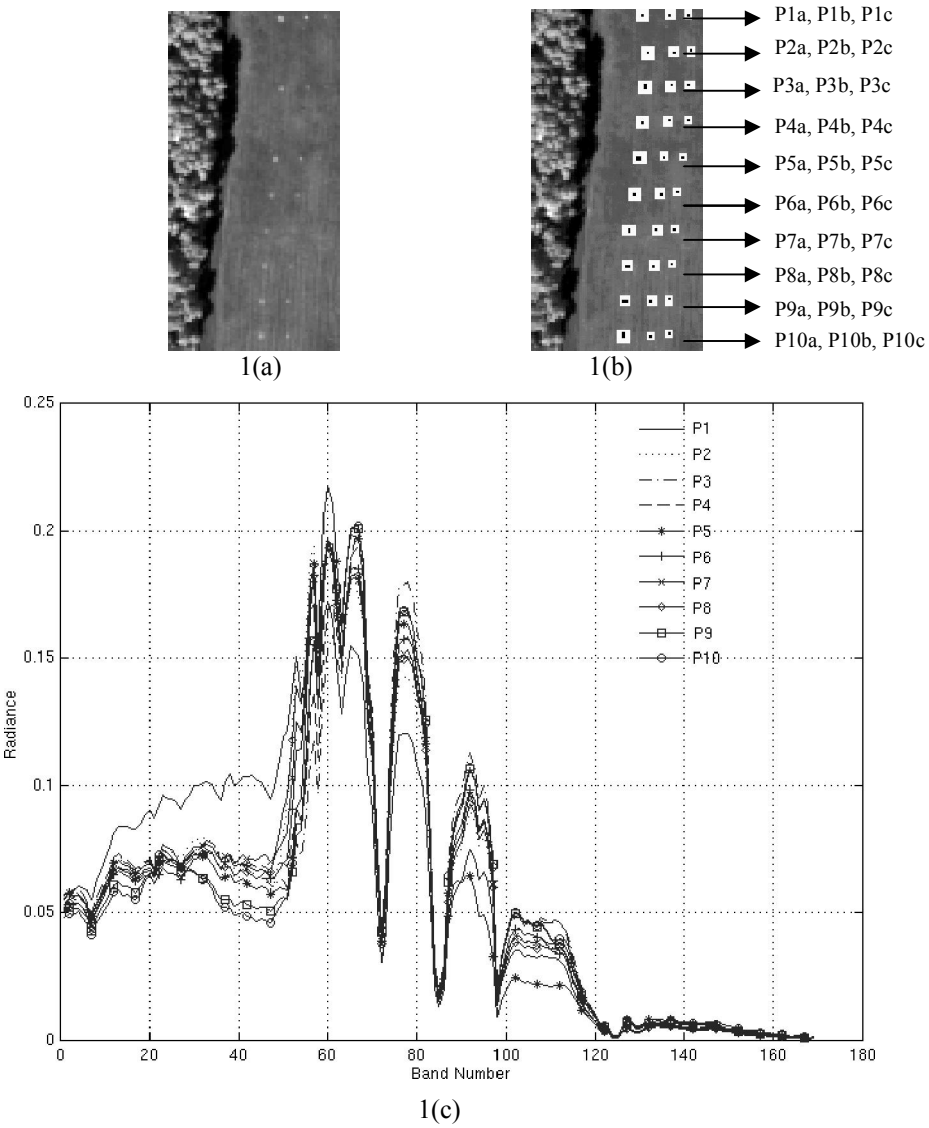

Figure 1: (a) A HYDICE image scene with 30 panel; (b) Spatial locations of 30 panels provided by ground truth; (c) spectral signatures of the 10 panel classes.

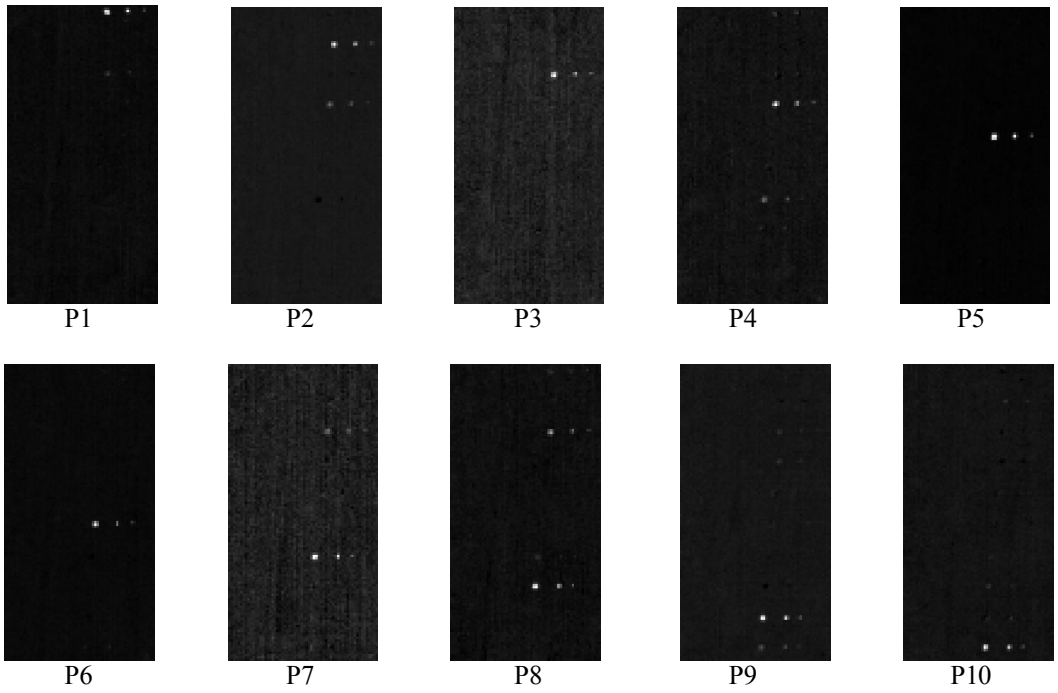

Figure 2: Classification result using ICA

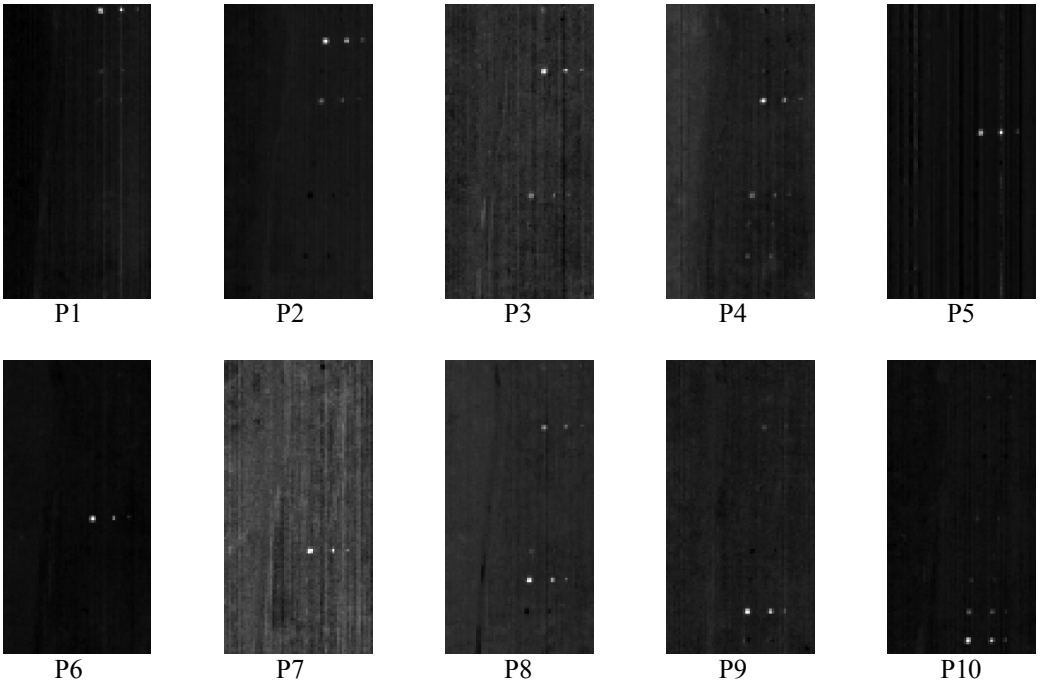

Figure 3: Classification result from DCA (HP)

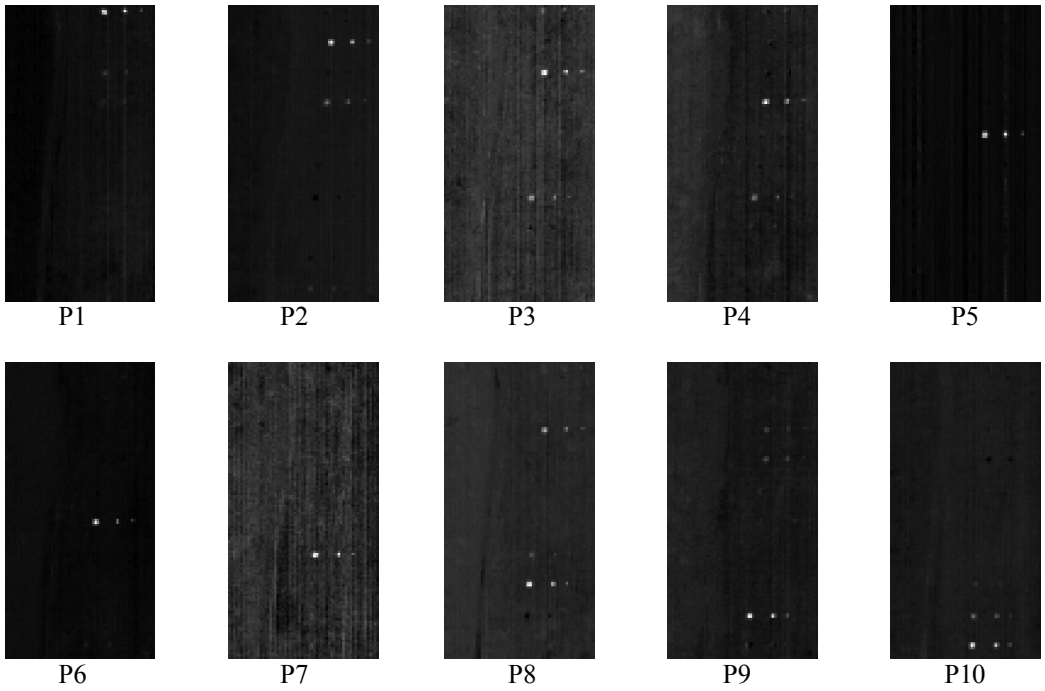

Figure 4: Classification result from DCA (IN)

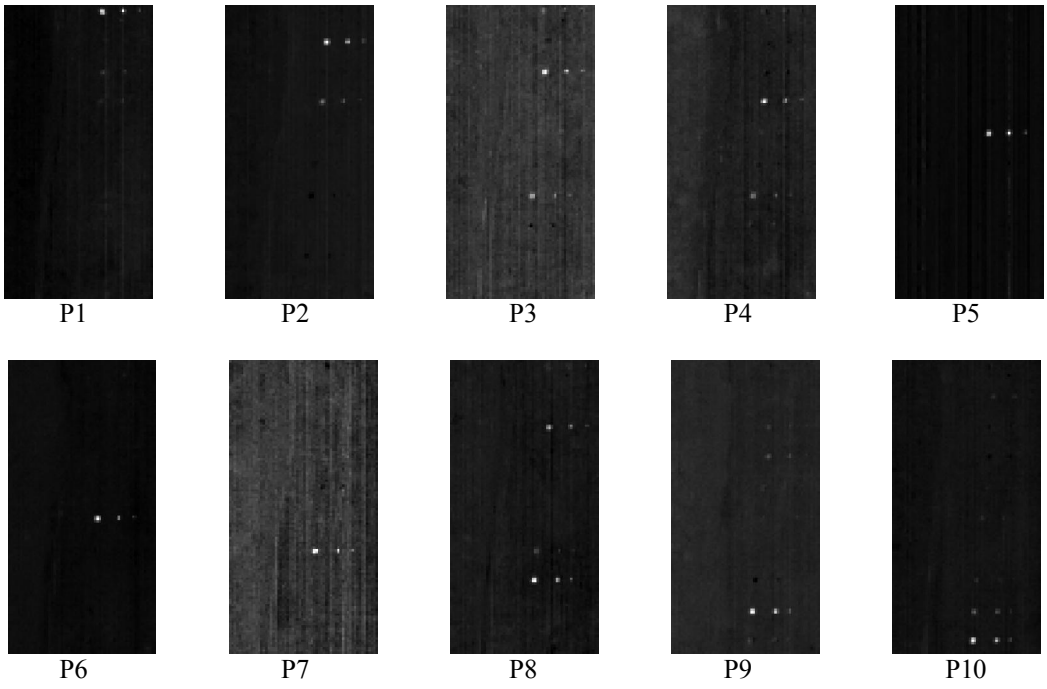

Figure 5: Classification result from DCA (SDICA)

Table 1: Number of false alarm pixels (N_F) and detected pixels (ND) when $\eta = 0.5$.

# of Pure Pixels		ICA		DCA (HP)		DCA (IN)		DCA (SDICA)	
		N_D	N_F	N_D	N_F	N_D	N_F	N_D	N_F
P1	3	2	0	2	0	2	0	2	0
P2	3	2	0	2	0	2	0	2	0
P3	4	3	0	3	0	3	1	4	5
P4	3	2	0	2	0	2	0	2	0
P5	6	5	0	4	0	5	0	5	0
P6	3	0	25	1	0	1	0	1	0
P7	4	3	0	3	0	3	0	3	0
P8	4	3	1	3	1	3	1	3	2
P9	4	3	0	3	0	3	0	3	0
P10	4	3	0	3	0	3	0	3	0
Total	38	26	26	26	1	27	2	28	7

Table 2: (Largest) Number of detected pixels (N_D) when no false alarm exists ($N_F = 0$).

# of Pure Pixels		ICA		DCA (HP)		DCA (IN)		DCA (SDICA)	
		N_D	N_F	N_D	N_F	N_D	N_F	N_D	N_F
P1	3	2	0	2	0	2	0	2	0
P2	3	1	0	1	0	1	0	1	0
P3	4	1	0	3	0	3	0	3	0
P4	3	1	0	1	0	1	0	1	0
P5	6	3	0	1	0	4	0	1	0
P6	3	0	0	1	0	1	0	1	0
P7	4	3	0	3	0	3	0	3	0
P8	4	2	0	2	0	2	0	2	0
P9	4	2	0	3	0	3	0	3	0
P10	4	1	0	3	0	3	0	1	0
Total	38	16	0	20	0	23	0	18	0

Table 3: (Smallest) Number of false alarm pixels (N_F) when all target pixels are detected ($N_D = 38$).

# of Pure Pixels		ICA		DCA (HP)		DCA (IN)		DCA (SDICA)	
		N_D	N_F	N_D	N_F	N_D	N_F	N_D	N_F
P1	3	3	5	3	0	3	1	3	1
P2	3	3	2670	3	1	3	11	3	7
P3	4	4	7175	4	17	4	2074	4	4721
P4	3	3	419	3	5	3	17	3	205
P5	6	6	0	6	0	6	0	6	0
P6	3	3	8086	3	0	3	0	3	0
P7	4	4	6445	4	1261	4	2053	4	5223
P8	4	4	61	4	6	4	82	4	15
P9	4	4	812	4	0	4	4	4	13
P10	4	4	119	4	3	4	5	4	5
Total	38	38	25792	38	1293	38	4247	38	10190

4. CONCLUSION

In this paper, we proposed DCA for unsupervised hyperspectral image classification. It finds a transform that can improve the statistical independence between the sources but leave the basis matrix unchanged, where an original ICA method can be applied. The preliminary result demonstrates that the three DCA approaches investigated can provide more accurately classification result than the original ICA.

REFERENCES

[1] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[2] C.-I Chang, S.-S Chiang, J. A. Smith, and I. W. Ginsberg, "Linear spectral random mixture analysis for hyperspectral imagery," *IEEE Trans. Geoscience Remote Sensing*, vol. 40, no. 2, pp. 375-392, 2002.

[3] X. Zhang and C. H. Chen, "New independent component analysis method using high order statistics with application to remote sensing images," *Optical Engineering*, vol. 41, no. 7, pp. 1717-1728, 2002.

[4] M. S. Naceur, M. A. Loghmari, and M. R. Boussema, "The contribution of the sources separation method in the decomposition of mixed pixels," *IEEE Trans. Geoscience Remote Sensing*, vol. 42, no. 11, pp. 2642-2653, 2004.

[5] Q. Du, I. Kopriva, and H. Szu, "Independent component analysis for hyperspectral remote sensing imagery classification," *Optical Engineering*, vol. 45, no.1, pp. 017008-1:13, 2006.

[6] A. K. Barros, *The independence assumption: dependent component analysis*, in: Advances in Independent Component Analysis, Springer, London, 2000.

[7] I. Kopriva. *Blind Signal Deconvolution as an Instantaneous Blind Separation of Statistically Dependent Sources*, Lecture Notes in Computer Science 4666, pp. 504-511, Springer-Verlag, Proceedings of the Seventh International Conference on Independent Component Analysis and Blind Source Separation, eds. M.E. Davies et al., London, UK, September 9-12, 2007.

[8] I. Kopriva and D. Seršić, "Wavelet packets approach to blind separation of statistically dependent sources," *Neurocomputing*, vol. 71, pp. 1642-1655, 2008.

[9] A. Cichocki and P. Georgiev, "Blind source separation algorithms with matrix constraints," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E86-A, pp. 522-531, 2003.

[10] T. Tanaka and A. Cichocki, "Subband decomposition independent component analysis and new performance criteria," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (ICASSP 2004), Montreal, Canada, pp. 541-544, 2004.

[11] A. Hyvärinen, "Independent component analysis for time-dependent stochastic processes," *Proceedings of the International Conference on Artificial Neural Networks* (ICANN'98), pp. 541-546, Skovde, Sweden, 1998.

[12] S. J. Orfanidis, *Optimum Signal Processing – An Introduction*, 2nd ed. MacMillan Publishing Comp., New York, 1988.

[13] I. Kopriva, "Approach to blind image deconvolution by multiscale subband decomposition and independent component analysis," *Journal of Optical Society of America A*, vol. 24, pp. 973-983, 2007.

[14] I. Kopriva and D. Seršić, "Robust Blind Separation of Statistically Dependent Sources Using Dual Tree Wavelets," *Proceedings of IEEE International Image Processing Conference*, vol. I, pp. 433-436, San Antonio, TX, USA, September 16-19, 2007.

[15] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, pp. 234-253, 2001.

[16] W. Selesnick, R. G. Baraniuk, and N.G. Kingsbury, "The dual-tree complex wavelet transform," *Signal Processing Magazine*, vol. 22, pp. 123-151, 2005.