Discriminant independent component analysis for hyperspectral image classification

Murinto and N R D P Astuti
Departement of Informatics Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
murintokusno@tif.uad.ac.id, rochmahdyah@tif.uad.ac.id

Abstract. A novel approach to discriminative extraction and independence of remote sensing hyperspectral imagery features using Discriminant Independent Component Analysis (DICA) is introduced in this paper. This approach maximizes the variance and negentropy of a given feature. The experimental results show that the proposed method has better classification accuracy when compared to using one of the most commonly used classifiers in the hyperspectral image of SVM. Classification accuracy obtained from DICA dimensionality reduction on Indian Pines dataset, value of average accuracy (AA) is 92.87 %, overall accuracy (OA) is 92.94 %, Kappa index is 0.82, whereas on Washington DC mall dataset, classification result obtained by using DICA dimensionality reduction is average accuracy amounted to 84.89 %, overall accuracy 84.69 % OA, Kappa index is 0.82. From experiment we can see that classification accuracy increases when using DICA approach compared than technique classical SVM.

1. Introduction
Hyperspectral imagery consists of more hundreds bands and have high spatial resolution, range from visible to infrared of areas. Spectral resolution increases with wide spectral range. It will further clarify characterizing the materials. This makes the hyperspectral image very suitable for source separation and classification processes. On the other hand, with very large data dimensions, the computational load will be heavier and can also decrease the result of the classification processes, because of the curse of dimensionality i.e. Huge effect [1]. In many studies have been developed techniques that overcome this problem, but the results still need to be improved again.

The linear transformation of the original data is often used to display data in order to facilitate its computation. Some of linear transformation methods include Principal Component Analysis (PCA) [2], Linear Discriminant Analysis (LDA) [3] and Factor Analysis (FA) [4]. The linear transformation technique that proposed after that is Independent Component Analysis (ICA) in which the components are statistically independent in the displayed data [5, 6]. Although widely used in hyperspectral image processing, the ICA has the disadvantage that independent components extracted on the ICA may be appropriate for representation but may not suitable to classification. Instead, it could use an optimization change on the problem [7]. To solve this problem, one of the solutions offered is to use a single, semi-supervised Discriminant independent component analysis (DICA) linear. In DICA, Fisher linear Discriminant and Negentropy are simultaneously used to process and hidden independent factors are extracted from multivariate data.
Some classifiers such as Support Vector Machine (SVM) [8] and Neural Network (NN) [9] are widely used in hyperspectral image classification, because of their ability to handle high dimensional data problems. However, they have disadvantages that may influence the classification process. The disadvantage of NN techniques is the number of hidden layers that are numerous, as well as the number of iterations that sometimes take too long on the computation process. Although SVM produces relatively good classification accuracy in limited training data sets, SVM also has a major disadvantage of training time, depending on the training sample data. In kernel usage, it is often necessary to find the optimal parameter, usually using a cross-validation approach [10].

In this study we propose a method for Discriminant analysis based on the extraction of independent features obtained by using entropy maximization i.e. Discriminant Independent Component Analysis (DICA) [7, 11]. The value of the fisher criterion and the sum of negentropy from features that independently extracted is maximized simultaneously. The results obtained using the proposed method was then compared with one of the most widely used classifiers in hyperspectral image processing i.e. the support vector machine (SVM).

2. Independent Component Analysis (ICA)

ICA is a linear transformation that minimizes dependencies statistically among the components. It was first introduced in 1994 by Pierre Common and has been used in a variety of applications including Blind Source Separation (BSS) [5]. ICA has been used to reduce high dimensions through the projection of the original hyperspectral image to a lower dimensional space. In this projection, the number of observed signals \( n \) is the original dimension of the data and the number of source \( m \) signals is a low dimensional space corresponding to material present in the hyperspectral image.

ICA is based on the problem BSS recovering unknown source signal by simplifying the assumption of \( n \) independent signal that can be write as \( s(t) = s_1(t), ..., s_n(t) \) and observed mixture of the signals are denoted as \( x(t) = x_1(t), ..., x_n(t) \). Suppose mixture of signal as linear mixing equation can be given as equation (1).

\[
x(t) = As(t)
\]

Where \( t \) is time, \( A \) is unknown mixing matrix. \( S(t) \) is a source signal for transformation into a mixed signal. If mixing matrix, \( A \) is invertible there is no source \( M \) less than or equal to number of sensors \( P \), then source can be split directly and output from \( y(t) \) can be write as equation (2).

\[
y(t) = Bx(t) = Bas(t)
\]

Where \( B \) is mixing matrix obtained from matrix \( A \).

ICA is a linear approximation that maximizes a Gaussian measure or minimizes objective function. In general, the objective function used in ICA algorithm is mutual information of \( y(t) \), which is written in equation (3).

\[
I(y_t, B) = \sum_i H(y_{t_i}) - H(y_t)
\]

3. Discriminant Independent Component Analysis

Multivariate data with lower dimensions and independent features in Discriminant Independent Component Analysis (DICA) were obtained by using entropy maximization [7]. Here the Fisher criterion and marginal summation of negentropy extracted independent features are maximized simultaneously. Optimization function that maximize the incorporation of negentropy and the function of classification performance measurement at constrains normalize vector \( v_i \) given through equations (4).

\[
\hat{L}(W) = \hat{L}(W) + k\phi(W, Z, C)
\]
Where k is constant and $\phi(W, Z, C)$ is measurement function classification performance from features Y given C, while $L(W)$ is a constraint of Lagrangian formula as given in [7].

Adaptation of $w_t$ can be improved by using gradient ascent algorithm along the function in equation (4), then followed by symmetric orthonormalization of W. Learning rule as given in equation (5).

$$\Delta w_t = \eta \left( y_i (E (Z g^2 (w_t^T Z)) + y_i k E (Z g^2 (W^T Z)) \frac{\partial \phi(w_t, Z, C)}{\partial w_i} + 2 \beta_i w_i \right)$$

$$W \leftarrow (WW^T)^{-1/2} W$$

(5)

Value of $\beta_i$ in equation (5) can be obtained from completion of stationary point of equation (4) and is given as:

$$\beta_i = -\frac{1}{2} \left( (y_i (k_1 E (y_i g^1 (y_i)) + y_i k_2 E (y_i g^2 (y_i)) + k w_i \frac{\partial \phi(w_t, Z, C)}{\partial w_i}) \right)$$

(7)

Function measuring classification perform $\phi(W, Z, C)$, Which is corresponds to demixing vector W given by the equation (8).

$$\phi(W, Z, C) = \sum_{c=1}^C \log \sum_{i=1}^R N_c (\mu_{ic} - \mu_i)^2$$

(8)

Where $\mu_{ic} = E (z_i | c)$ is feature mean i class c, and $N_c$ is the number of samples corresponding to class c, while $\mu_i = E (z_i)$ is feature mean i-th class. While optimal linear transformation can be write as equation (9).

$$\bar{W} = arg \max_W \phi(W, Z, C) = arg \max_W \sum_{c=1}^C \log (\sum_{i=1}^R N_c (\mu_{ic} - \mu_i)^2$$

(9)

Partial derivatives from $\phi(W, Z, C)$ Which corresponds to the vector basis $w_t$ can be written as an equation (10).

$$\frac{\partial \phi(W, Z, C)}{\partial w_i} = 2 \sum_{c=1}^C \sum_{n \epsilon class c} \frac{(\mu_{ic} - \mu_i)}{\sum_{c=1}^C N_c (\mu_{ic} - \mu_i)^2} z_{in}$$

(10)

Where $z_{in}$ is sample n$^a$ from whitened observation. By substituting the equation (10) into equation (5), it produces a discriminative feature extraction algorithm that maximizes sum of marginal negentropy among classes plotted on each feature. The algorithm for extraction of DICA features on asymmetric distribution is given as follows [7, 12]:

1. Center the data, $X : X \leftarrow X - E(X)$.
2. Whitening the data center X to get orthonormal features, Z using equation (4).
3. Initialize $W = W_0 : \|W_0\| = I_R$.
4. Obtain DICA features, $Y = W^T Z, Y \in R^R$.
5. Update W using equations (10) and (5).
6. Symmetric orthogonalization $W' : W \leftarrow (WW^T)^{-1/2} W$.
7. If the sum negentropy y and discriminant function $\phi$ is convergent then iteration stopped. If other, back to step 4.

4. Experiment results

This experiment uses two hyper spectral image datasets obtained from Aviris and Hydice sensors. Data from Aviris sensors used in this experiment is Indian pines hyperspectral image, while Hydice sensor is hyperspectral image of Washington DC Mall. First image is Indian Pines hyperspectral image obtained from Aviris sensor that flew over the north western Indiana area in 1992. This image has a size of its spectral resolution of 145 x 145 pixels. Where in this image consists of 220 spectral bands (Noise already omitted) [12]. Ground truth Indian pines hyper spectral image consists of 16 classes. Second image is hyper spectral image obtained from the Hydice sensor. This sensor gets the
image of a flight over the Washington DC Mall. The hyperspectral image of Washington DC Mall consists of 210 spectral bands. Band noise is eliminated and only 191 spectral bands are used. This hyperspectral image was taken in August 1995, in which each band consists of 1280 lines, with each band consisting of 307 pixels. Ground truth image Washington DC Mall consists of 6 classes. In the experiment, 35 pixels were taken randomly from each class and used as training samples [8]. The test data in this experiment are each taken from each class on the image data of Indian pines and Washington DC Mall. Complete data of training and testing are shown in table 1. In this experiment training sample was taken 20 pixels randomly from each class.

**Table 1.** Samples of train-test for two hyper spectral image datasets.

| No. | Class         | Indian Pines | Washington DC Mall |
|-----|---------------|--------------|--------------------|
|     |               | Class | Train | Test | Class | Train | Test |
| #1  | Alfalfa       | 20    | 54    | 35   | 3794  |       |      |
| #2  | Corn no-till  | 20    | 1434  | 35   | 376   |       |      |
| #3  | Corn-min till | 20    | 834   | 35   | 135   |       |      |
| #4  | Corn          | 20    | 234   | 35   | 1888  |       |      |
| #5  | Grass-pasture | 20    | 497   | 35   | 365   |       |      |
| #6  | Grass-Trees   | 20    | 747   | 35   | 1184  |       |      |
| #7  | Grass-Mowed   | 20    | 26    | 35   | 57    |       |      |
| #8  | Hay-windowed  | 20    | 489   |      |       |       |      |
| #9  | Oats          | 20    | 20    |      |       |       |      |
| #10 | Soybean-no till | 20 | 968   |      |       |       |      |
| #11 | Soybean-min till | 20 | 2468  |      |       |       |      |
| #12 | Soybean-clean | 20    | 614   |      |       |       |      |
| #13 | Wheat         | 20    | 212   |      |       |       |      |
| #14 | Woods         | 20    | 1294  |      |       |       |      |
| #15 | Bldg-Trees-Drive | 20 | 380   |      |       |       |      |
| #16 | Stone-Steel-Tower | 20 | 95    |      |       |       |      |

Results of experiments conducted in this study using DICA then compared with results obtained using SVM, with polynomial kernels. Compared performance is in terms of Overall Accuracy (OA), Kappa index (K) and Average Accuracy (AA) of both techniques used. OA represents percentage of sample correctly classified. AA represents the average of classification accuracy for each class in the data set, while Kappa index (K) is a correct estimate of the percentage estimation parameters without the amount due chance alone. Table 2 shows the classification results using DICA and SVM original methods. From table 2 it shows that in AVIRIS Indian Pines data set when using SVM obtained an OA value of 86.18%, index of K is 0.85 and AA is 85.95%, while with same data set increase in the value of OA, K and AA are 92.94%, 0.92 and 92.98%, respectively. Likewise, in Washington DC Mall data set, there is an increase in classification accuracy when using DICA approach. The values obtained are OA is 75.59%, K is 0.73 and AA is 80.86%, using SVM method. And when we were using DICA the value obtained by OA is 84.69 %, K is 0.82, and AA is 84.89%. From this result so can be concluded that by using DICA method it will be increase the accuracy of hyperspectral image classification compared with SVM original methods.
Table 2. Comparison of classification accuracy resulting from DICA and SVM methods of hyperspectral image datasets of Indian Pines and Washington DC Mall.

| Methods | Indian Pines Image | Washington DC Mall Image |
|---------|-------------------|--------------------------|
|         | SVM | DICA | SVM | DICA |
| OA      | 86.18% | 92.94% | 75.59% | 84.69% |
| K       | 0.85 | 0.92 | 0.73 | 0.82 |
| AA      | 85.95 | 92.87 | 80.86 | 84.89 |
| Class #1| 47.62 | 80.95 | 60.10 | 70.31 |
| Class #2| 80.00 | 90.00 | 75.00 | 80.48 |
| Class #3| 90.91 | 100.00 | 90.91 | 92.50 |
| Class #4| 85.00 | 95.00 | 75.00 | 80.00 |
| Class #5| 85.00 | 100.00 | 75.00 | 80.50 |
| Class #6| 100.00 | 100.00 | 100.00 | 100.00 |
| Class #7| 100.00 | 100.00 | 90.00 | 91.00 |
| Class #8| 100.00 | 100.00 | - | - |
| Class #9| 100.00 | 95.00 | - | - |
| Class #10| 95.24 | 95.24 | - | - |
| Class #11| 90.48 | 95.24 | - | - |
| Class #12| 65.00 | 90.00 | - | - |
| Class #13| 75.00 | 75.00 | - | - |
| Class #14| 100.00 | 100.00 | - | - |
| Class #15| 73.91 | 78.26 | - | - |
| Class #16| 86.96 | 91.30 | - | - |

5. Conclusions
In this research we have proposed an approach for hyperspectral image classification using Discriminant Independent Component Analysis (DICA). The proposed approach is based on the application of Independent Component Analysis (ICA) to obtain independent components of the processed data, using entropy maximization and Fisher criterion. The study used two different data sets: Aviris Indian Pines and Hydice Washington DC Mall data sets to implement the proposed approach. The experimental results show that the proposed method has better classification accuracy when compared to using one of the most commonly used classifiers in the hyperspectral image of SVM. Classification accuracy obtained by using DICA dimensionality reduction on Indian Pines Aviris dataset, the value of average accuracy (AA) is 92.87%, overall accuracy (OA) is 92.94%, Kappa index is 0.82, whereas on Washington DC mall dataset, classification result obtained by using DICA dimensionality reduction is average accuracy amounted to 84.89%, overall accuracy 84.69 % OA, Kappa index is 0.82. This classification accuracy is higher when compared to using the technique classical SVM.

Acknowledgment
This research is funded by DP2M-RISTEKDIKTI with applied products research (PPT) scheme as outlined in contract No: PPT-071/SP3/LPP-UAD/IV/2017.

References
[1] Hughes G F 1968 On the mean accuracy of statistical pattern recognizers IEEE Trans. Inform. Theory vol IT-14 pp 55–63
[2] Rodarmel C and Shan J 2002 Principal component analysis for hyperspectral image classification Surveying and Land Information Science vol 62 no 2 p 115
[3] Bandos T V, Bruzzone L and Camps-Valls G 2009 Classification of hyperspectral images with regularized linear discriminant analysis IEEE Transactions on Geoscience and Remote
Sensing vol 47 issue 3 pp 862-873

[4] Hans G and Geladi P ed. 2007 Techniques and applications of hyperspectral image analysis (John Wiley & Sons)

[5] Hyvärinen A, Karhunen J and Oja E 2002 Independent component analysis (John Wiley & Sons, Inc.)

[6] Villa A, Chanussot J, Jutten C, Benediktsson J A and Moussaoui 2009 On the use of ICA for hyperspectral image analysis Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009 Cape Town, South Africa

[7] Dhir C S and Lee S-Y 2011 Discriminant independent component analysis IEEE transactions on neural networks vol 22 issue 6 pp 845-57

[8] Melgani and Bruzzone L 2004 Classification of hyperspectral remote sensing images with support vector machine IEEE Transaction on Geoscience and Remote Sensing vol 42 issue 8

[9] Ratle F, Camps-Valls G and Weston J 2010 Semisupervised neural networks for efficient hyperspectral image classification IEEE Transactions on Geoscience and Remote Sensing vol 48 issue 5 pp 2271-82

[10] Camps-Valls G and Bruzzone L 2005 Kernel-based methods for hyperspectral image classification IEEE Transactions on Geoscience and Remote Sensing vol 43 issue 6 pp 1351-62

[11] Mollae M, Moattar M H and Mahdavi S J S 2014 Feature extraction based on modified discriminant independent component analysis via particle swarm optimization International Journal of Software Engineering and Its Applications vol 8 no 12 p 89-100

[12] Murinto and Harjoko A 2016 Dataset feature reduction using independent component analysis with contrast function of particle swarm optimization on hyperspectral image classification 2016 2nd International Conference on Science in Information Technology (ICSITech)