

Dimensionality Reduction using Hybrid Support Vector Machine and Discriminant Independent Component Analysis for Hyperspectral Image

Murinto¹, Nur Rochmah Dyah PA²

^{1,2}Department of Informatics Engineering
Faculty of Industrial Technology, Universitas Ahmad Dahlan Yogyakarta, Indonesia

Abstract—Hyperspectral image is an image obtain from a satellite sensor. This image has more than 100 bands with a wide spectral range and increased spatial image resolution, providing detailed information on objects or materials that exist on the ground in a specific way. The hyperspectral image is well suited for the classification of the earth's surface covering due to it. Due to the features of the hyperspectral data, then lately research related to trend hyperspectral data tend to increase. The transformation of the reduction of the dimensions of the original data into a new dimension reduction chamber is often done to overcome the problem of 'curse of dimensionality' in which its dimensions tend to increase exponentially. Data is mapped from the original data to a lower dimensionless space through a dimensional reduction procedure which must display the observation input effectively. Therefore, in this research we proposed a hyperspectral dimension hybrid reduction method which adopted Support Vector Machine (SVM) and Discriminant Independent Component Analysis (DICA) techniques to reduce original data in order to obtain better accuracy. By using SVM+DICA is used to reduction dimension hyperspectral images. In this research, we use KNN as classifier. From the experiment obtained that value of average accuracy is 0.7527, overall accuracy is 0.7901, and Kappa is 0.7608 for AVIRIS dataset.

Keywords—Classification; discriminant independent component analysis; support vector machine; hyperspectral image

I. INTRODUCTION

Experiments using Hyperspectral imagery are recently widely performed. Hyperspectral image has very high resolution, providing detailed information about objects or materials that exists on the ground in a specific way. The hyperspectral image is well suited for the classification of the earth's surface coverings due to it. Due to the features of the hyperspectral data, recent research related to hyperspectral data has increased rapidly. The dimensional data reduction transformation of observational data into a new dimension reduction chamber intended to address the dimensionality curse of dimensionality dimension tends to increase exponentially [1]. Data are transformed from original data to lower dimension space through dimensional reduction procedures which must display the observation input effectively. The effectiveness of observations and dimensional space of reduction is measured by the criteria defined in various dimensional data reduction algorithms. Many techniques to reduce dimension data are Principal Component

Analysis, Linear Discriminant Analysis [2], Independent Component Analysis (ICA) [3], and Discriminant Independent Component Analysis [4] which is an extension of the ICA.

PCA dimensional data reduction method uses Eigen value decomposition to obtain orthogonal projections, also called principal components, minimizing squared errors between original and projected data. LDA have function criteria through the covariance matrix in class and between classes such as matrix between the maximized class and the matrix in the class that is minimized to obtain better separation in the reduction chamber [5]. ICA obtains a projection vector that is statistically independent of observational data through criteria representing independence such as Kullback-Leibler (KL) divergence, mutual information (MI) and correlation. DICA in [4] represents a methodology for combining both ICA and Fisher Linear Discriminant functions in order to construct feature extraction methods through variable projection where variables are projected through maximum independence. In the previous experiment we use independent component analysis with particle swarm optimization as contrast function as a dataset feature reduction [10].

Moreover, the organization of this paper is as follows. In Section I, the background of this research is presented. In Section II it explained material and methodology are consists of input dataset, hybrid SVM + DICA, and accuracy assessment. Section III is specifically explained about result and discussion. Section IV presents conclusion of this research.

II. MATERIAL AND METHODOLOGY

This research is more deeper and study about data dimension reduction and classification especially on remote sensing hyperspectral image data, reviewing other journals or libraries relating to dimensional reduction and image classification both nationally and internationally. This research based on idea Moon [6], [11], where is SVM and ICA used to dimensionality reduction in cancer dataset. Researchers also conducted a more detailed literature review of support vector machine methods for hyperspectral image classification, conventional dimensional data reduction methods available and reviewed existing constraint optimization techniques. In this research we will look for the values of OA, AA and K of the proposed method. Then the results are compared with other existing methods such as PCA, ICA, and DICA.

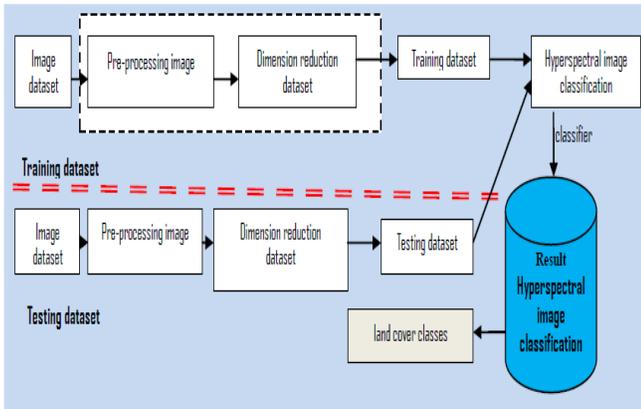


Fig. 1. Overview data dimension reduction using hybrid support vector machine and discriminant independent component analysis method.

A hybrid dimension reduction method was adopted which adopted the use of SVM as the robust dimension reduction criterion through the process of redundancy of data redundancy. Based on SVM, linear orthogonal subspace-based SVM and DICA are built here, as well as non-linear non-correlated subspace-based SVM and DICA are also built in this study. Furthermore, researchers conducted experiments to implement the proposed hybrid method in programming language by utilizing computer software. Detail diagram of the system overview to be constructed in Fig. 1.

A. Input Dataset

In this experiment we use data obtained from AVIRIS sensor, namely Indian Pines image. Image is taken in 1992 from Northwestern Indiana region. Original image has 145 x 145 pixels with 220 bands. After reduced some bands containing noise and water absorption, image has only 190 spectral bands used in this study. Wavelength between 0.4 to 2.5 This image contains 16 corresponding classes. Fig. 2 show the image of Aviris Indian Pines [7].

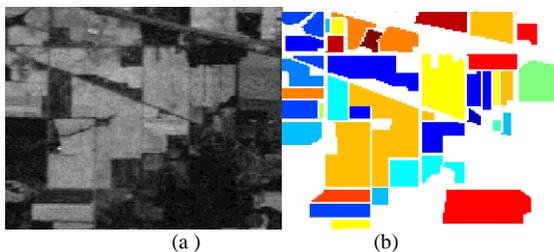


Fig. 2. Aviris Indian Pines dataset (a) composite color 3-band (R,G,B), (b) Ground truth.

B. Hybrid SVM + ICA

The purpose of this step is to apply dimension reduction data method using hybrid SVM + DICA method.

1) Discriminant independent component analysis

In the Discriminant independent component analysis (DICA) method, multivariate data with lower dimensions and independent features are obtained through Negentropy maximization [4]. In DICA, the Fisher criterion and the sum of marginal negentropy independent features are extracted by maximizing simultaneously. Therefore, DICA incorporates

representational models with Discriminant models to obtain better classification.

Negentropy is a statistical estimate of non-Gaussian random variables [11]. An approach of marginal negentropy can be written as (1).

$$N(y_i) \approx c_1(E(H^1(y_i)))^2 + c_1(E(H^1(y_i) - E(H^2(y_i) -)))^2 \quad (1)$$

In the equation (1), H^1 represented non quadratic odd function and H^2 represented non quadratic function even. The general elections for a random vector with symmetrical distribution (normal):

$$H^1(y_i) = y_i^3 \quad (2)$$

$$H^2(y_i) = \frac{1}{a_i} \log \cosh a_i y_i, 0 \leq a_i \leq 1 \quad (3)$$

$$H^2(y_i) = \exp\left(\frac{y_i^2}{2}\right) \quad (4)$$

Maximization of the marginal quantity of Negentropy with covariance unit can be obtained through the Lagrange equation in the following form:

$$L_a(W) = \sum_{i=1}^k \left[E(H(w_i^T z)) - E(H(v)) \right]^2 + \sum_{i=1}^R \beta_i (w_i^T w_i - 1) \quad (5)$$

The target functions in equation (5) are maximized so features can be obtained. Optimization problems to maximize criterion function for classification and negentropy performance of independent features simultaneously can be written as follows:

$$L_a = L_a(W) + c\phi(W, Z, A) \quad (6)$$

Where c is a constant, ϕ is a function to measure efficient classification from features Y given A , $L_a(W)$ same with β_i . Learning rule in the following form:

$$\Delta w_i = \eta(\gamma_i E(Z_g(w_i^T Z)) + k \frac{\partial \phi(W, Z, A)}{\partial w_i} + 2\beta_i w_i) \quad (7)$$

$$\beta_i = -\frac{1}{2} \gamma_i E(y_i g(y_i)) \quad (8)$$

$$\gamma_i = 2 \left[-\frac{\sum_{n=1}^N \exp(-\frac{y_{in}^2}{2})}{N} + \frac{1}{2} \right] \quad (9)$$

Perform a symmetric orthogonal of the matrix

$$W \leftarrow (WW^T)^{-1/2} W \quad (10)$$

Measurement of classification performance function as follows:

$$\frac{\partial \phi(W, Z, A)}{\partial w_i} = \sum_{i=1}^R \log \frac{w_i^T S_B w_i}{w_i^T S_W w_i} = \sum_{c=1}^C \log \frac{N_c(\mu_{ic} - \mu_i)^2}{\sum_{c=1}^C N_c \sigma_{ic}^2} \quad (11)$$

$$\mu_{ic} = \frac{1}{N_c} \sum_{n \in \text{Class } c} y_{in} \cdot \sigma_{ic} = \frac{1}{N_c} \sum_{n \in \text{Class } c} (y_{in} - \mu_{ic})^2 \quad (12)$$

In the gradient method, to maximize ϕ , need a descent computation from ϕ based on the vector base W , which is conducted as follow:

$$\frac{\partial \phi(W, Z, C)}{\partial w_i} = 2 \sum_{c=1}^C \sum_{n \in \text{class } c} (A_{ic} - B_{ic}) z_{in} \quad (13)$$

$$A_{ic} = \frac{(\mu_{ic} - \mu_i)}{\sum_{c=1}^C N_c (\mu_{ic} - \mu_i)^2}, B_{ic} = \frac{(y_{in} - \mu_{ic})}{\sum_{c=1}^C N_c \sigma_{ic}^2} \quad (14)$$

Furthermore, it can be written that the DICA algorithm is as follows [8]:

Algorithm 1: DICA Algorithm [8]
 1) . Centered observed data X
 2). Whiten center observed data to get orthonormal features Z
 3). Initialization $W = W_0 || W_0$
 4). Get features DICA in R space by $Y = W^T Z$
 5). Update using equation (6)
 6). Calculated Symmetric orthogonal $W = (W W^T)^{-\frac{1}{2}} W$
 7). If sum of Negentropy from y_i and Discriminant $\phi(W, Z, A)$ are konvergen then iteration stop. Else, back to step 4).

2) Hybrid SVM and DICA

Diagram of the proposed method is show in Fig. 3. In this diagram, $X = \{x_i \in R^n, \forall_i\}$ as a dataset, $Z = \{z_i \in R^n, \forall_i\}$ is data that projected based on X , $W = [w_1, w_2, \dots, w_m]$ is mapping matrix where $w_i \in R^n$ and $1 \leq m \leq n$, $W_{i,j} = [w_i, \dots, w_j]$ where $1 \leq i \leq j$ is sub matrix W and l , where $1 \leq l \leq m$, represented sum of features based on risk minimization mapping, $l = 1$ for two class dataset such that $W = [w_1]$ as vector. In this diagram, symbol \oplus indicated that two sub matrix, $W_{i,l}$ dan $W_{l+1,m}$ are concatenation matrix's $W = [w_1, \dots, w_m]$.

Sub matrix $W_{i,l}$ is adopted from risk minimization W as a part of W to directly considered supervised classification information as an intermediate step enabling optimization of risk and independence of data separately through the projected data set Z of X over w_i , where $i = \{1, 2, \dots, l\}$ for independence maximization.

Linear mapping W is generated from sequential process risk minimization to independence maximization as seen in Fig. 3.

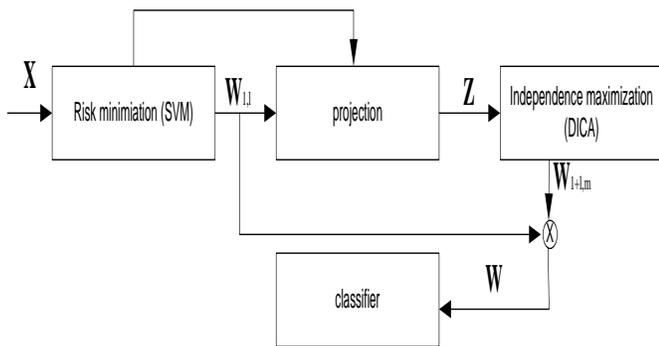


Fig. 3. Framework hybrid SVM and DICA method.

a) Risk Minimization

In risk minimization, $W_{i,j}$ has column vector representing direction of decision surface on classification process. In Support Vector Machine (SVM) there is an outcome that meets minimum structural risk by maximizing margin

separation through quadratic constraint problems with duality for binary classification problems, which can be expressed in (15).

$$\alpha^* = \text{argmin} \left\{ \frac{1}{2} \alpha^T Q \alpha - \alpha^T \mathbf{1} \right\} \quad (15)$$

$$\text{Depend on } \sum_{i=1}^N \alpha_i y_i = 0, 0 \leq \alpha_i \leq C$$

Where $\alpha = [\alpha_1, \dots, \alpha_N]^T$ and $\alpha_i, i = 1, 2, \dots, N$, that denotes multiplier Lagrange suitable to pair data i that is (x_i, y_i) with x_i as sample data and $y_i \in \{-1, 1\}$ denotes the class index for the separation problem of the two classes. N is the number of samples of the training data set. $Q = [q_{ij}]$ is a matrix $N \times N$ where $q_{ij} = y_i y_j \langle x_i, x_j \rangle$, $i, j \in \{1, 2, \dots, N\}$. While $\mathbf{1}$ represents a column vector consisting of a value of 1 N . C is the relaxation parameter of SVM to tolerate a certain level of empirical error in margin determination during training. The optimal output determination is established through:

$$w^T x + b = 0. \quad (16)$$

Where $w = \sum_i \alpha_i^* y_i x_i \in R^n$, $b \in R^1$ to take $w^T x = 0$ parallel to origin intersection in R^n . w is main information of decision process. In this paper, w is used as part of W on the proposed SVM + DICA methods.

b) Data Projection

Set of mapping vectors data obtained by structural minimization process ($W_{1,l}$) and independence maximization ($W_{l+1,m}$). Mapping vector obtained from this process without redundant information to construction of space span along ($W_{1,l}$) and ($W_{l+1,m}$). At least redundancy results from pair-wise orthogonality between w_i and w_j where $i \in \{1, 2, \dots, l\}$ and $j \in \{l+1, 2, \dots, m\}$. The pair-wise orthogonality is also depicted through $W_{1,l} \perp W_{l+1,m}$ or same with $W_{1,l}^T W_{l+1,m} = 0$.

Middle step in SVM + DICA framework is data projection component, allows for mapping vectors obtained from structural risk minimization and independence maximization to get minimum correlation. This is done through projection of data given X into a subspace that satisfies $W_{1,l}^T x = 0$, yielding the projected data, ie Z , such that subsequent independence maximization process based on Z is at least influential or correlated with process previous structural risk minimization. After data projection procedure, the projected data Z , will lose information along direction of $W_{1,l}$, which indicates that decision information through $W_{1,l}$ is no longer valid in projection subspace. However, the projection ensures that some mapping vectors of structural risk minimization, $W_{1,l}$ and independence maximization, $W_{l+1,m}$, have no correlation since $W_{1,l} \perp W_{l+1,m}$.

When data projection into subspace, orthogonal to decision hyper planes of structural minimization, $W_{1,l}$, is written as a constraint of optimization problems as follows:

$$z^* = \text{argmin}_z \|x - z\|^2 \text{ Subject to } W_{1,l}^T z \quad (17)$$

Where z represents data projected into an orthogonal subspace $W_{1,l}$, and parallel to decision hyper planes. Due to the orthogonality between $W_{1,l}$ and some components in decision plane, the structural risk minimization and

independence minimization are isolated and displayed individually by independence between multiple pairs w_i and w_j , $i \in \{1, 2, \dots, l\}$, $j \in \{1 + 1, 2, \dots, m\}$.

Furthermore, equation (6) solved by using Lagrange optimization, $\lambda \in \mathbb{R}^1$ as follows:

$$L(z, \lambda) = \|x - z\|^2 + \lambda^T W_{1,1}^T z \quad (18)$$

By taking partial derivatives of L against z and λ , we find equation:

$$\frac{\partial L(z, \lambda)}{\partial z} = -2(x - z^*) + W_{1,1} \lambda = 0 \quad (19)$$

$$\frac{\partial L}{\partial \lambda} = W_{1,1}^T z^* = 0 \quad (20)$$

From sum of equation (19) and (20) can be obtain matrix:

$$\begin{bmatrix} 2I_n & W_{1,1} \lambda \\ W_{1,1}^T & 0 \end{bmatrix} \begin{bmatrix} z^* \\ \lambda \end{bmatrix} = \begin{bmatrix} 2x \\ 0 \end{bmatrix} \quad (21)$$

Where I_n identity matrix is n dimension. z^* is form of the projected Z dataset.

c) Data independence maximization

The search intelligence maximization for a linear non-orthogonal coordinate system having axes is established through both the first statistical order and higher than the original data. As an unsupervised feature extraction method in the proposed framework, independence maximization is applied to the projected data set Z when the data representation capability through independence maximization. It's likely to result in better unsupervised classification accuracy compared with other conventional unsupervised feature extraction methods. Independence maximized by adopted approximated negative entropy criterion, because it is efficiently both error and computational, as one of the variants of several ICA. The negative entropy approach can be written as:

$$w_i^+ = E\{zg(w_i^T z)\} - E\{g'(w_i^T z)\} \quad (22)$$

Where is w_i^+ is temporal approximation independent component, $i \in \{1 + 1, 2, \dots, m\}$. g is a derivative of a non-quadratic function introduced in [6], and $g(u) = \tanh(au)$, g' is a derivative of g , and $g'(u) = \text{asech}^2(au)$ can be written as:

$$W_{l+1,m} = W_{l+1,m}^+ [(W_{l+1,m}^{+T} W_{l+1,m}^+)^{-1/2}]^T \quad (23)$$

Where is $W_{l+1,m}$ is a representation of decorrelated mapping based on $W_{l+1,m}^+ = (w_{l+1,1}^+, \dots, w_m^+)$ maximization.

C. Accuracy Measurement

Evaluation of classification accuracy of y will be done by classification accuracy assessment which is looking for Kappa (K), overall accuracy (OA) and average accuracy (AA) value [9].

III. RESULT AND DISCUSSION

Aviris Indian Pines dataset is a multiclass in the nature that used in this experiment dimensionality reduction purpose. The number of features or dimensions is reduced to 95% due

to hybrid method. The k nearest neighbor algorithm is used for the classification accuracy which is our performance metric. Some other metrics' are also used for this which are Kappa (K), overall accuracy (OA) and average accuracy (AA) value. Above all the parameters are calculated by the use of confusion matrix. From experimental result we can show that proposed hybrid technique are performs extremely well. The result of classification accuracy shows better than other methods such as PCA, ICA, DICA, and SVM+DICA. Result of Classification show in Fig. 4 and Table 1. We can see that value AA is 0.7527, OA is 0.7901, K is 0.7608 when using SVM +DICA as reduction dimensionally methods. Similarly we can see when using PCA methods only, value of AA) is 0.7501, OA is 0.7450, and value of K is 0.7410.

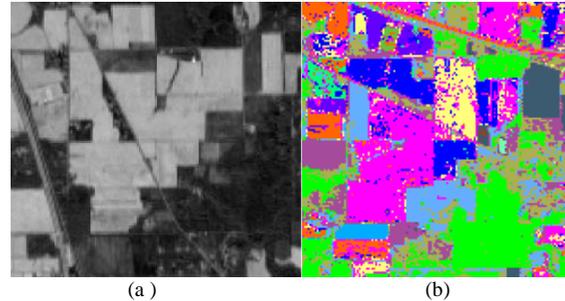


Fig. 4. Aviris Indian pines (a) 10 independent component (IC) using SVM+DICA method (b) Result classification.

TABLE I. ACCURACY CLASSIFICATION USING THE PROPOSED HYBRID METHOD

Methods	OA	AA	K
PCA	0.7510	0.7450	0.7410
ICA	0.7316	0.7203	0.7356
DICA	0.7820	0.74.10	0.7580
SVM + DICA	0.7901	0.7527	0.7608

IV. CONCLUSION

Based on the results of research that has been done, it can be concluded things as follow: We have reviewed and implemented a hybrid method for reducing dimension of hyperspectral image data using support vector machine (SVM) and discriminant independent component analysis (DICA). KNN is used to be classifier in this experiment. The value of AA is 0.7527, OA is 0.7901, and K is 0.7608 when classification done by using SVM+DICA dimensionality reduction on AVIRIS dataset. Therefore, this is required formation of appropriate method for reduction of image data dimension for classification process so that obtained higher accuracy compared to previous method such as only use PCA, ICA, DICA. Furthermore the researcher wants to use this proposed method for image segmentation based on bio-inspired algorithm for hyperspectral image.

ACKNOWLEDGMENT

The researchers expressed gratitude to DP2M-RISTEKDIKTI who has funded this research with applied products research (PPT) scheme as outlined in the research contract PPT-071/SP3/LPP-UAD/IV/2017.

REFERENCES

- [1] Hughes, Gordon. "On the mean accuracy of statistical pattern recognizer", IEEE Transaction of information theory 14.1 (1968):55-63.
- [2] Rodarmel, Craig, and Jie Shan. "Principal component analysis for hyperspectral image classification", Surveying and Information Science 62.2 (2002): 115.
- [3] Hyvarinen, Aapo, Juha Karhunen, and Erkki Oja. Independent component analysis. Vol.46. John Wiley&Sons, 2004.
- [4] Dhir, Chandra Shekhar, and Soo-Young Lee. "Discriminant independent component analysis". IEEE transactions on neural networks 22.6 (2011):845-857.
- [5] Bandos, Tatyana V., Lorenzo Bruzzone, and Gustavo Camps-Valls. "Classification of hyperspectral images with regularized linear discriminant analysis". IEEE Transactions on Geoscience and Remote Sensing 47.3 (2009):862-873.
- [6] Moon, Sangwoo and Hairong Qi. "Hybrid Dimensionality Reduction Method Based on Support Vector Machine and Independent Component Analysis" *IEEE Transaction on Neural Network and Learning Systems*, Vol.23, No.5, May 2012.
- [7] Langrebe, D and Biehl, L. AVIRIS Dataset. (2002). School of Electrical and Computer Engineering Purdue University
- [8] Dhir, Chandra Shekhar, Jahyung Lee, and Soo-Young Lee. "Extraction of independent discriminant features for data with asymmetric distribution", *Knowledge and information systems* 30.2 (2012):359-375.
- [9] Congalton, Russel G., Kass Gree. *Assesing the accuracy of remotely sensed data: principle and practice*. CRC press, 2008.
- [10] Murinto, Agus harjoko. "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), 2016.
- [11] Sangwoo Moon. "A hybrid feature extraction framework based on risk minimization and independence maximization", 2009 International Joint Conference on Neural Network, 2009.