A Classification Technique for Panchromatic Imagery Using Independent Component Analysis Feature Extraction

Seung-Gun Byoun, Ho-Yong Lee, Min Kim and Kwae-Hi Lee

Dept. of Electronic Engineering, Sogang University

1 Shinsu-Dong Mapo-Gu, Seoul, Korea, 121-742

Tel: +82-02-704-4088, FAX: +82-02-704-4088

Email: seungun@eerobot1.sogang.ac.kr, lhy2502@eerobot1.sogang.ac.kr

Abstract **- Among effective feature extraction methods from the small-patched image set, independent component analysis (ICA) is recently well known stochastic manner to find informative basis images. The ICA simultaneously learns both basis images and independent components using high order statistic manners, because that information underlying between pixels are sensitive to high-order statistic models. The topographic ICA model is adapted in our experiment. This paper deals with an unsupervised classification strategies using learned ICA basis images. The experimental result by proposed classification technique shows superior performance than classic texture analysis techniques for the panchromatic KOMPSAT imagery.**

Keyword **- topographic ICA, unsupervised classification, texture analysis.**

I. Introduction

In the last decade, several classification techniques for remotely sensed panchromatic scene have been developed into various methodologies. Classic texture analysis based on fine resolution have tried to extract features using the co-occurrence matrix, wavelet, morphological decomposition for representing statistic properties between pixels, and then have classified a gray image [1], [10]. It is known that conventional classification algorithms have difficulty to classify high-resolution data into, such as KOPSAT (6.6 m), so that those feature extraction methods mentioned above have a certain limitation depending on which scenes are used.

In our experiment, we adapt independent component analysis; recently become the center of attention in the signal processing research field, to extract meaningful features from a patched window set. T.J. Sejnowski showed the best performance of face recognition, and T.M Lee segmented into natural objects and texts on the foreground using learned basis features derived from the ICA mixture model [2], [9]. Feasible ICA approaches for the remote sensing research have been very rare, for example separating hyper-spectral data into desire and noisy images by applying the fundamental ICA notion (e.g., blindly source separation) and classifying the SAR data into different levels of ice by learned ICA basis images using the fast ICA algorithm [4], [8].

In this paper, we propose the classification strategy for KOMPSAT gray image based on unsupervised manner. We, in advance, select texture images having similar statistic characters from the "Brodatz" texture book with corresponding to those of desired 4 classes in KOMPSAT image [3]. The algorithm for extracting basis image sets, which contain significant information, is the "Topographic Independent Component Analysis"[7]. The normalized cosine coefficients between learned basis images and patched images are dealt with like probability value on the each class pdf, then are contributed to classify test image. Our experiment result shows the possibility that a high-resolution remotely sensed panchromatic scene could be classified into desired classes.

II. Topographic Independent Component Analysis

The ICA algorithm in this paper makes good use of finding meaningful representations for the patched image set. The image data set is formed into the column structure generating the factorial code for input patched image (i.e., pixels and images are treated variables and outcomes, respectably) [2]. Typical image decomposition methods (e.g., The Principal Component Analysis) are the way to find a good basis image set for representing prototype images. The PCA generally considers the second order linear dependency by adapting uncorrelateness, so that much of information in, which exist high order relationships between pixels can not be extracted.

The ICA, that is the generalization of PCA, has been proposed with several different algorithms to separate such high order dependencies using high order statistic. Before ICA decomposition, a patched gray image set $X=[x_1(u,v),\Lambda\ ,x_n(u,v)]$ must be normalized with zero mean and unit variance called centering, where u,v are locations in the column vector and n is the total number of patched images. The centered $\widetilde{X}$ can be written as linear combination of m basis images $A=[a_1(u,v),\Lambda\ ,a_m(u,v)]$ (i.e., a mixing matrix) and the independent coefficient $S=[s_1,\Lambda\ ,s_m]$ in the ICA estimation problem [6].

$$\widetilde{x}=As \tag{1}$$

Another useful preprocessing strategy in ICA is to first whiten image set. This means that the whitened $\widetilde{X}$ become to be uncorrelated and its variance equal unity. The whitening transform is performed by the eigen value decomposition of $\widetilde{X}$'s covariance matrix.

$$E(\widetilde{X}\widetilde{X}^T)=EDE^T \tag{2}$$

where E is the orthogonal matrix of the eigenvector and D is the diagonal matrix of the eigenvalue. The whitening matrix is denoted as,

$$V=D^{-1/2}E^T \tag{3}$$

The input for ICA estimation (Eq. (1)) is changed to the whitened matrix $Z=D^{-1/2}E^TX$. The most important ICA assumption is that IC s_i is non-guassian and independent random variables each other as possible, so that the information of one of s_i gives little information to others. Different methods for estimating independence in ICA model have been proposed, such as the maximum likelihood estimation by minimizing mutual information and the fixed-point algorithm by maximizing negentropy. Let's define mixing matrix W as follows,

$$W=A^{-1}=[w_1,\Lambda\ ,w_m] \tag{4}$$

The maximum likelihood estimation can be formulated by [5]

$$L(W)=\prod_{j=1}^{n}\prod_{i=1}^{m}p_i(a_i^Tz(j))\left|\det W\right| \tag{5}$$

$$\frac{1}{n}\log L(W)=E\{\sum_i^m\log p_i(w_i^Tz)\}+\log\left|\det W\right| \tag{6}$$

,where the p_j denotes the densities of the independent components. The topographic ICA that is the extension of maximum likelihood estimation (Eq. (6)) is applied for learning basis images for our experiment. A neighborhood relation defines a topographic order and the model by maximizing likelihood function can be achieved by using the stochastic gradient ascent algorithm [8]. The topographic ICA is the algorithm that a basis image is learned through some high order static. Common ICA estimation algorithms are originated in perfectly statistic independence, so that can not describe the relationship between independent components. For various applications, the independence might not be always satisfied and little dependence could be remained between independent components. On a count of mentioned above, it is reasonable to use revealed dependence. The basis images using the topographic ICA model (i.e., similar in location, orientation, and frequency) are close to each other. Figure .1 shows basis image sets of topographic ICA, which are learned from 4 texture images. This model gives Gabor-like basis images for image patches [5].

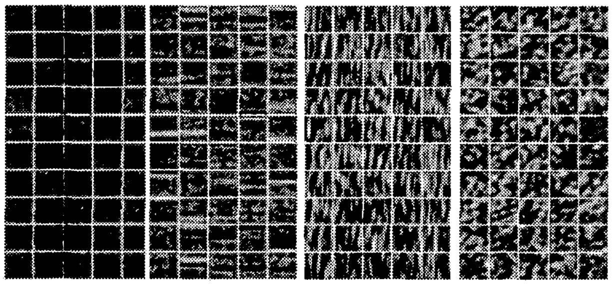

Figure.1. Topographic basis image sets of 4 texture images. Each of texture images is composed of 5×10 basis images from class 1 on the leftmost to class 4 on the rightmost.

III. Feature extraction methodology

In the case of fine resolution image (e.g., KOPSAT imagery), edges and textures in the class are visually distinguishable to other class, a directly learning basis image from remotely sensed image, however, might produce a noise mixing basis image. Some filtering method leads either to disappearance of texture, or to learning some artifact [8]. We selected 20 texture images from "Brodatz" texture book are similar to predefined classes in the KOMPSAT image, and then tried to learn basis images for unsupervised classification process. Firstly, we make texture images and KOMPSAT image to be normalized with zero mean and unit variance. To measure texture information constituting the test image, we secondly calculate the cosine value vector between basis image set A^k from the texture image and a randomly patched image from the test image, and then measure similarity.

$$S_n^k(A^k) = \left| \frac{A^k \cdot x_n^k}{\|A^k\| \|x_n^k\|} \right| \quad (7)$$

$$S_n^k(A_x^k) = \left| \frac{A^k \cdot x_n}{\|A^k\| \|x_n\|} \right| \quad (8)$$

, where A^k is the basis image set learned from k-th texture image, x_n^k is the training data to learn basis image set A^k, and x is the randomly patched image set in the test image.

$$Nearest(k) = \sum^k \sqrt{\sum_{i=1}^m (mean(S^k(A^k)) - S_i^k(A_x^k))^2} \quad (9)$$

The left in Figure.2 shows clustering groups of each $S_i^k(x^k)$ computed by using selected 4 texture images. New input x is projected into each k-th basis image set, and then $S_i^k(x)$ is calculated. If $S_i^k(x)$ has the nearest Euclidean distance between a mean vector of $S_i^k(x^k)$, it is considered that x has similar texture information. We use total 4 texture images for classification process.

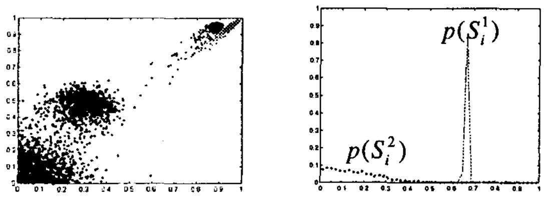

Figure.2. Clustered each $S_i^k(x^k)$ at cosine plane on the left, probability density of S_i^1 and S_i^2 on the right.

IV. Classification Algorithm

We found that hierarchical structure classification made better performance through our experimental experience (i.e., firstly segmenting two classes, and then again classifying that segmented class), Among 50 basis images of each class, the basis image being able to generate similar cosine vector with same texture information is desirable, otherwise, put the patched image on the long distance from its cluster. We should choose basis images that stands for the class and do not have interference with others at the same time. First used criterion faction is the method using variance between all classes and a element of one class [2].

$$r_1 = \frac{\sigma_{between}}{\sigma_{within}} \quad (10)$$

$$\sigma_{between} = \sum_C (\bar{a}_C - \bar{a})^2 \quad (11)$$

$$\sigma_{within} = \sum_C \sum_i (a_{iC} - \bar{a}_C)^2 \quad (12)$$

, where $\bar{a}$ is the overall mean of basis images across all basis images, $\bar{a}_C$ is the mean for class C, and i is the basis image number.

As a second criterion function, we select basis images a_i^C using modified pdf method based on cosine coefficients from [8].

$$S_i^{C1}(X^1) = \left| \frac{a_i^{C1} \cdot X^{C1}}{\|a_i^{C1}\| \|X^{C1}\|} \right| \quad (13)$$

$$S_i^{C2}(X^2) = \left| \frac{a_i^{C1} \cdot X^{C2}}{\|a_i^{C1}\| \|X^{C2}\|} \right| \quad (14)$$

$$r_2(C_1, i) = p(\alpha_1 > S_i^{C1}(X^1) > \alpha_2) \quad (15)$$

$$r_2(C_2, i) = p(\alpha_1 > S_i^{C2}(X^2) > \alpha_2) \quad (16)$$

$$p(\alpha_1) \cong p(\alpha_2) \cong (peak\ p(S_i^{C1}(X^1)))/T \quad (17)$$

,where $S_i^1(X^1)$ is the cosine coefficient vector between i-th basis image in the first class and the sample set in the first class and α is the each cosine value from 0 to 1. We fix threshold T=100 that divides peak probability $p(S_i^{C1}(X^1)$ from our experimental experience. If $r_2(C_1, i) - r_2(C_2, i)$ is large enough as seen at the right in Figure.2, it is regarded basis image a_i^1 in the first class as good as identifier, then used for classification. If an input with similar texture information has high probability by projected into the basis image in the case of different texture information, it has low probability by the same basis image. These basis images are applied for image classification because of faithful discriminators. We selected 20 basis images for each class trough above process.

Our classification algorithm is that test image is classified based on probability density of chosen basis image. The cosine value of input image x can be obtained by $S_i^C(x)$ and dealt with on the previously calculated pdf $p(S_i^C)$. This way gives us very fast operating time.

$$p^C(S_i^C(x)) = \prod_{i=1}^{n} p(S_i^C(x)) \quad (18)$$

When we, in practice, assign an input x to certain class having maximum probability product, just one zero value makes result to be zero as seen at the right in Figure. 2. We adapt the summation of probability instead of just product [8].

$$p^C(S_i^C(x)) = \sum_{i=1}^{n} w_i p(S_i^C(x)) \quad (19)$$

$$x^C = \max p^C(S_i^C(x)) \quad (20)$$

, where w_i is the weighting value.

The class C that makes maximum summation of probability for the input patched image is assigned.

V. Experimental Result

The KOMSAT has been successively launched at December 21, 1999 providing panchromatic 6.6m resolution and the test image is Chung-Ju area where the middle of KOREA on first, March 2000. The KOMPSAT image contains fine edges and texture information on a count of relatively high-resolution image, there, however, exist thin scanning lines and typical noises. If one tries to directly pick up samples from the KOMPSAT image for supervised classification approach, he and she would not obtain good performance. Although the goal of our experiment was the unsupervised classification without any previous knowledge, we properly assumed 4 classes in the test image to select texture images having similar statistic properties with respect to those of defined 4classes. We considered the test image composed of mountain (class 1), land & cultivated land (class 2), cultivated land (class 3), residence zone (class 4), and selected 20 texture images resembling to defined 4 classes from "Brodatz" texture book. The ICA basis image sets are learned from given 20 images, normalized cosine vectors between randomly patched images and 20 basis image sets by (Eq. (9)). We finally chose 4 texture images having nearest Euclidean distance between $mean(S^k(A^k))$ and assumed 4 classes (i.e., arbitrary classes in case of not defined class before).

We made sure that hierarchically classification strategy produced better result. To classify test image using hierarchical manner, Class1 and class 3 are put together and class2 and class 4 are also put together for first level classification because of similar texture statistic under laid. When first level classified image is again divided into four classes, weighting value w_i mentioned in (Eq. (19)) is used. The right in Figure.3 is the proposed classification result for test image. There are some misclassified regions because statistic properties of class2 (land & cultivated land) and Class3 (cultivated land) are very similar each other.

The operating time of this algorithm based on pdf is pretty short. It took about 45 minutes to learned 20 ICA basis image sets and took about 12 minute to classify 400 ×400 test image on the right in Figure.3 using IBM PC Pentium-3 800MHz.

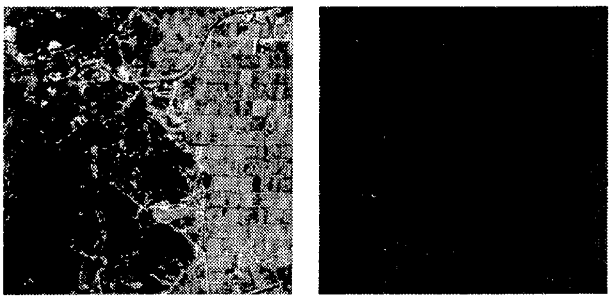

Figure. 3. The original KOMPSAT Image (400×400 size) on the left, classified image using 4 ICA basis image sets; black (class1: mountain), blue (class2: land & cultivated land), green (class3: cultivated land), red (class4: residence zone) on the right.

VI. Conclusion

It has been very hard to obtain desire result from the KOMPSAT panchromatic scene using classic texture analysis. The classification strategy of ICA basis image shows the reasonable performance. Contained texture information of predefined 4classes is not unique each other. 3 classes (e.g., class1, class2, and class3) have similar statistic component. We might have good performance if the higher resolution scene including detail texture information is used.

From several experiments, we experience that learned basis images, more or less, act differently to visually similar texture. It is very difficult to define appropriate textures corresponding to desire classes. We are able to correctly choose texture images if we measure the texture information of class according to our purpose.

Though topographic ICA is used for trying to compensate rotated orientation of the patched image, that finally obtained basis images are rotated by certain angles may improve classification result. It is also necessary to apply the circle shape mask, but the rectangular mask(21×21) is used in our experiment.

References

[1] J.A. Benediktsson, M. Pesaresi, 2001, Feature extraction and classification of urban high-resolution satellite imagery based on morphological preprocessing, Geo-science and Remote Sensing Symposium, IGARSS '01. IEEE 2001 International , 1(1): 534 –536.

[2] M.S. Bartlett, J.R. Movellan, & T.J. Sejnowski, 2001, Face recognition by independent component analysis, IEEE Transactions on Neural Networks(in press).

[3] Phil Brodatz, 1966, A Photographic Album for Artists and Designers, New York, Dover Inc.

[4] Chein-I. Chang Shao-Shan. Chiang, 2002, Linear spectral random mixture analysis for hyperspectral imagery, Geo-science and Remote Sensing IEEE

Transactions, 40(2): 375 -392.
[5] Mark Girolanmi, 2000, Advance in independent Component Analysis, London, Springer Inc, 45-62
[6] Aapo Hyvarinen, Juha Karhunen, 2001, Independent Component Analysis, John wiley &Sons Inc,
[7] A. Hyvarinen and P.Hoyer, 2000, Emergence of topographic and complex cell properties from natural images using extension ICA, In Advanced in Neural Information Processing System, 12: 827-833.
[8] J. Karvonen, M. Simila, 2001, Independent component analysis for sea ice SAR image classification, Geoscience and Remote Sensing Symposium IGARSS '01, 3(3): 1255 -1257.
[9] Te-Won. Lee, 2002, Unsupervised image classification, segmentation, and enhancement using ICA mixture models, Image Processing IEEE Transactions, 11(3): 270 -279.
[10] P.C. Smits, A. Annoni, 1999, Updating land-cover maps by using texture information from very high-resolution space-borne imagery, Geo-science and Remote Sensing IEEE Transactions, 37(3): 1244 -1254.