We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



16

Multivariate Time Series Support Vector Machine for Multispectral Remote Sensing Image Classifications

Pei-Gee Peter Ho
Naval Undersea Warfare Center, Newport Rhode Island
USA

1. Introduction

Satellite and airborne Remote Sensing for observing the earth surface, land monitoring and geographical information systems are the big issues in world's daily life as well as country defense projects. The source of information was primarily acquired by imaging sensors and spectroradiometer in remote sensing multispectral image stack format. The traditional image processing either by single picture image processing or compressing pictures stack via Principle Component Analysis (PCA) or Independent Component Analysis (ICA) into a single image component for further pixel classification or region segmentation is not enough to describe the true information extracted from multispectral satellite sensors. In an effort to significantly improve the existing classification and segmentation performance in this research, the contextual information between pixels or pixel vectors is characterized by a time series model for the remote sensing image processing.

Time Series statistical models such as Autoregressive Moving Average (ARMA) were considered useful in describing the texture and contextual information of an remote sensing image. To simplify the computation, a two-dimensional (2-D) Autoregressive (AR) model was used instead. In the first phase, the 2-D univariate time series based imaging model was derived mathematically (Ho, 2008) to extract the features for further terrain segmentations. The effectiveness of the model was demonstrated in region segmentation of a multispectral image of the Lake Mulargias region in Italy. Due to the nature of remote sensing images such as SAR (Synthetic Aperture Radar) and TM (Thermal Mapper) which are mostly in multispectral image stack format, a 2-D Multivariate Vector AR (ARV) time series model with pixel vectors of multiple elements (i.e. 15 elements in the case of TM+SAR remote sensing) are examined. The 2-D system parameter matrix and white noise error covariance matrix are estimated for further classifications in the 2nd phase of algorithm development. To compute the time series ARV system parameter matrix and estimate the error covariance matrix efficiently, a new method based on modern numerical analysis is developed by introducing the Schur complement matrix, the QR (orthogonal, upper triangular) matrix and the Cholesky factorizations in the ARV model formulation. As for pixel classification, the powerful Support Vector Machine (SVM) kernel based learning machine is applied in

conjunction with the 2-D time series ARV model. The SVM is particularly suitable for the high dimensional vector measurement as the "curse of dimensionality" problem is avoided. The performance improvement over the popular Markov random field is demonstrated. The 2-D multivariate time series model is particularly suitable to capture the rich contextual information in single and multiple images at the same time. A single composite image is constructed from the vector pixels through ARV based Support Vector Machine classifications.

2. Remote Sensing Image Data Set

The remote sensing image data of the earth surface is from either satellite or aircraft in digital multispectral or hyperspectral format. Both multispectral and hyperspectral imaging techniques are the process of capturing the same scene at different wavelengths that yield a 2-D spatial dimensions and one spectral dimension hypercube. The main properties of a remote sensing image are the wavelength bands it interprets. As far as remote sensing physical phenomena is concerned, some are the measurements of the spatial information reflected by the solar radiation in terms of visible and ultraviolet frequency range of wave (Schowengerdt, 1997). This type of remote sensing is the passive type. Some are the spatial distribution of energy emitted by the earth in the thermal infrared. Others are in the microwave band of wavelengths, measuring the energy returns from the earth that was transmitted from the vehicle which is the active type of remote sensing (Richards, 1999). The remote sensing image data is based on the concept of the "spectral signature" of an object (Prieto, 2000). Therefore, different land covers have different spectral signatures. The system produces multi-spectral images stack that each pixel is represented by the mathematical pixel vector which contains a set of brightness values for the pixels arranged in column form:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_N \end{bmatrix}$$

Table 1. (Bruzzone, 1998) below lists examples of the spectral regions on earth remote sensing.

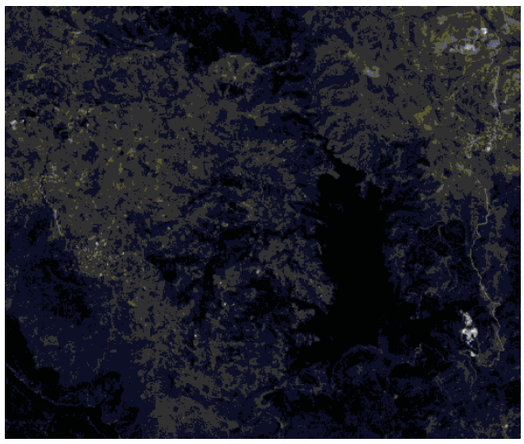
Spetral Signature	Wavelength	Radiation source	Surface
Visible	0.4-0.7 mm	Solar	Reflectance
Infrared	0.7-1.1 mm	Solar	Reflectance
Short Infrared	1.1-2.5 mm	Solar	Reflectance
Mid Infrared	3-5 mm	Solar and Thermal	Refl. and Temp.
Thermal Infrared	0.95 mm	Thermal	Temperature
Radar band Ka	0.8-1.1 cm	Thermal	Temperature

Radar band K	1.1-1.7 cm	Thermal	Temperature
Radar band Ku	1.7-2.4 cm	Thermal	Temperature
Radar band X	2.4-3.8 cm	Thermal	Temperature
Radar band C	3.8-7.5 cm	Thermal	Temperature
Radar band S	7.5-15 cm	Thermal	Temperature
Radar band L	15-30 cm	Thermal	Temperature
Radar band P	30-100 cm	Thermal	Temperature

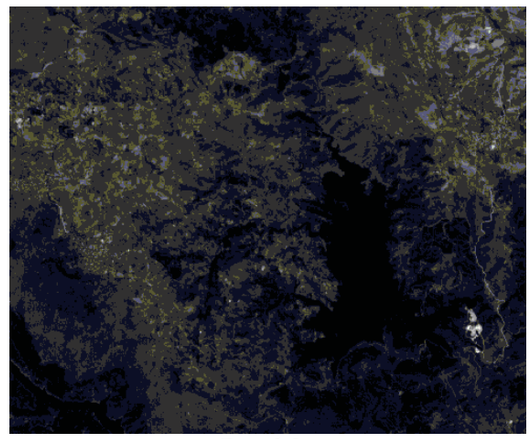
Table 1. The spectral signature of different bands used in remote sensing

Thermal Mapper (TM) Remote Sensing Images Data:

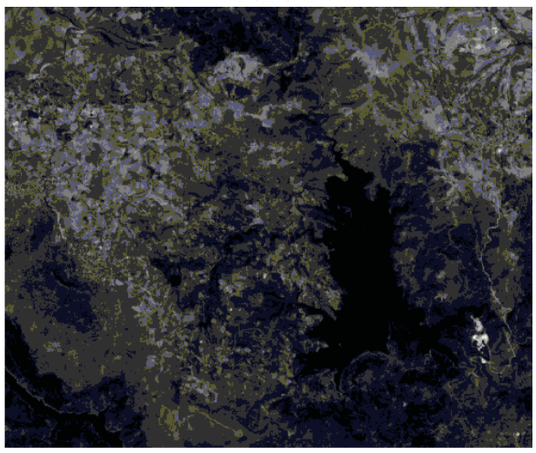
The Landsat earth resources satellite system was the first designed to provide global coverage of the earth surface via remote sensing techniques. Three imaging instruments have been used with this satellite. These are the Return Beam Vidicon (RBV), the Multi-spectral Scanner (MSS) and the Thermal Mapper (TM). Landsat uses multispectral instruments to record the remote sensing images stack. Figure 1. shows one example of 6 different bands of Thermal Mapper acquired by the Landsat 5 satellite in the area of the Mulargias lake in July 1996.



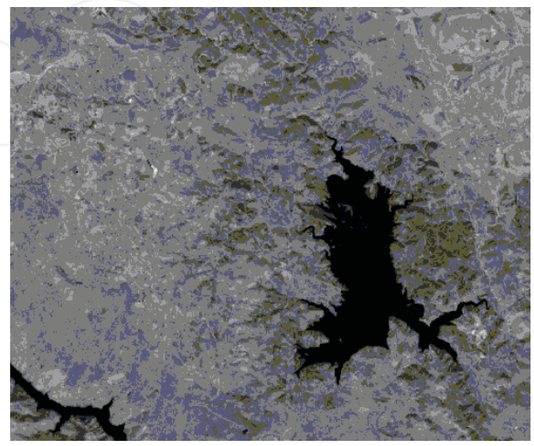
Band 1



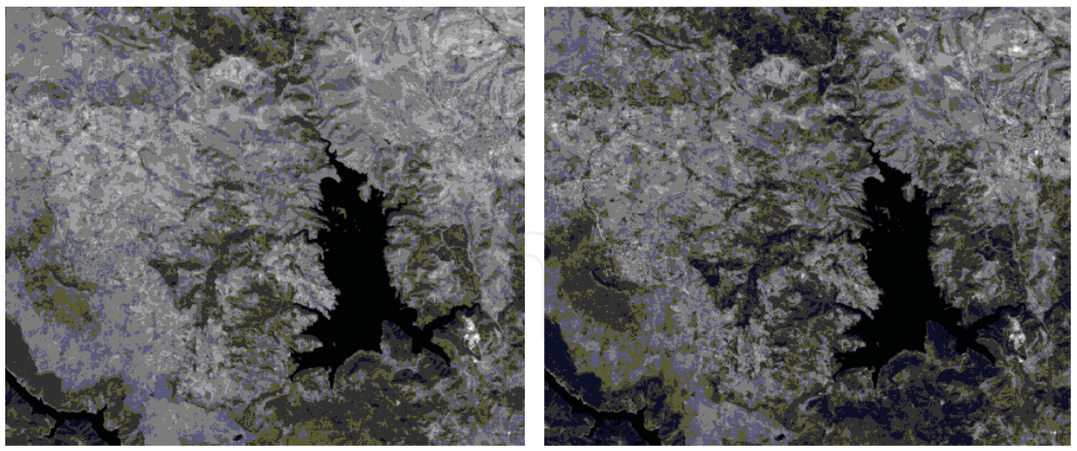
Band 2



Band 3



Band 4



Band 5 Band 7

Fig. 1. Examples of Thermal Mapper

Synthetic Aperture Radar (SAR) Remote Sensing Images Data:
Active remote sensing techniques employ an artificial source of radiation. The resulting signal scatters back to the sensor which reflect atmosphere or earth characteristics. Synthetic Aperture Radar is an imaging technology that the radiation is emitted in a beam from a moving sensor. The backscattered components returned from the ground are measured. An image of the backscatter spatial distribution is reconstructed by digital processing of the amplitude and phase of the returned signal. Samples of SAR polarization images are shown in figure 2.

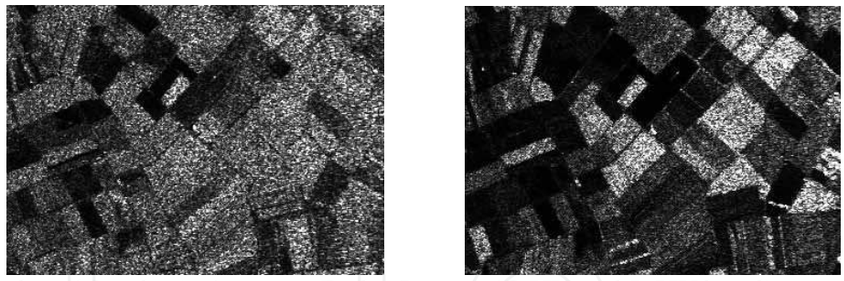


Fig. 2. Samples of Synthetic Aperture Radar Images (C-HV and L-HV band)

Though SAR images are popular due to its accessibility, the speckle noise is all over the places which degrades in the image quality.

Hyperspectral Remote Sensing Images:
The other type of multiple spectral images data are produced by spectrometers which is different from multipectral instruments. One example is the NASA's Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) optical sensor that delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral bands. The AVIRIS detector operates with a wavelength of 10 nanometers to cover the entire range. Figure 3 shows the AVIRIS hyperspectral sample images through different spectral bands.

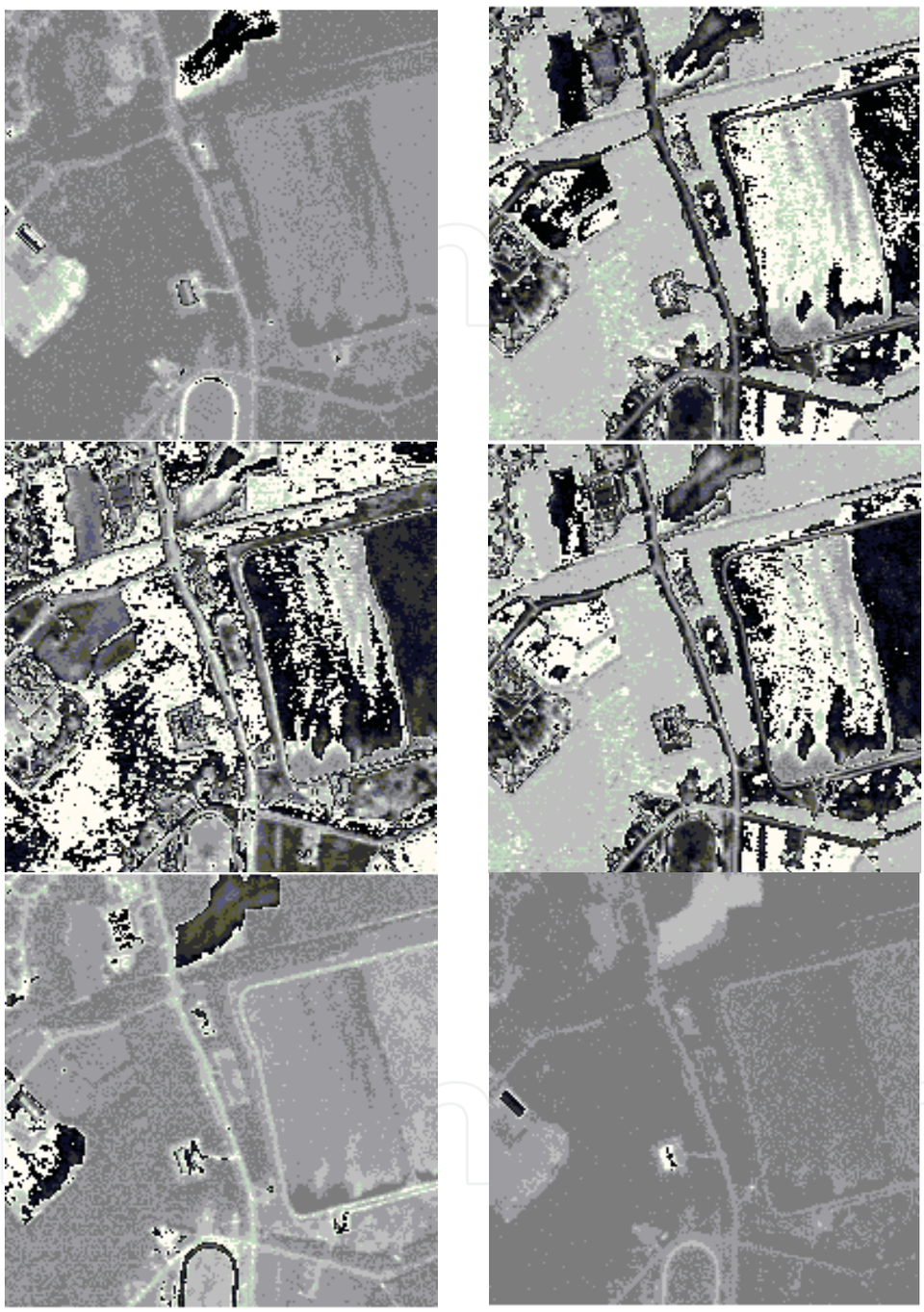


Fig. 3. AVIRIS Hyperspectral Image Examples

Either multi-spectral image stack (TM or SAR) or hyperspectral remote sensing stack are suitable for demonstrating the classification capability of the newly developed Multivariate Autoregressive image model based SVM method. Unfortunately, the MOFFET data set of AVIRIS remote sensing on NASA's website is no longer available. The remote sensing image

data for testing algorithm in this chapter is limited to the multispectral (TM and SAR) images stack for the purpose of comparisons to other classification algorithms.

3. Methodologies on Time Series Remote Sensing Image Analysis

A general scheme of Time Series Remote Sensing Image Processing as described in this chapter is shown in figure 4 and 5. below:

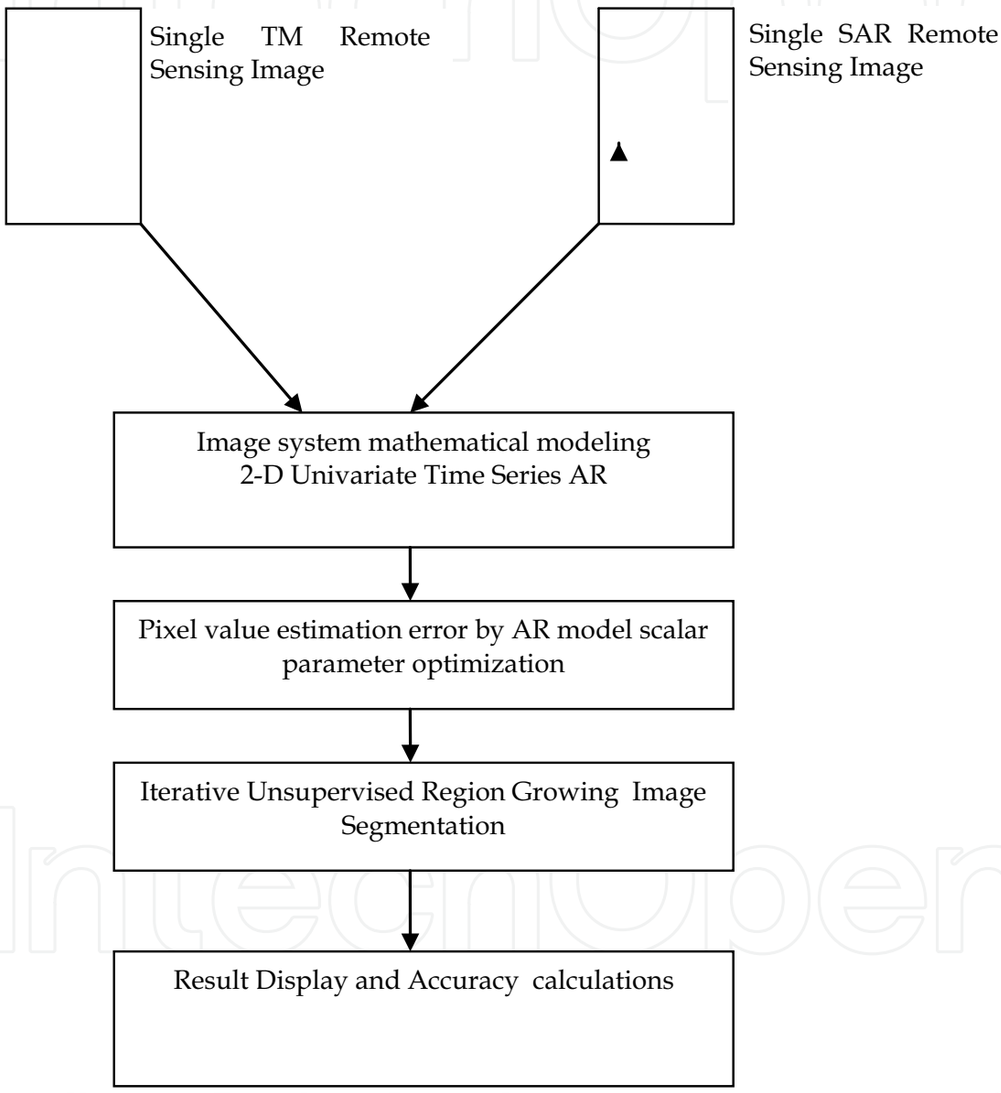


Fig. 4. Univariate Time Series Region Growing image processing system scheme

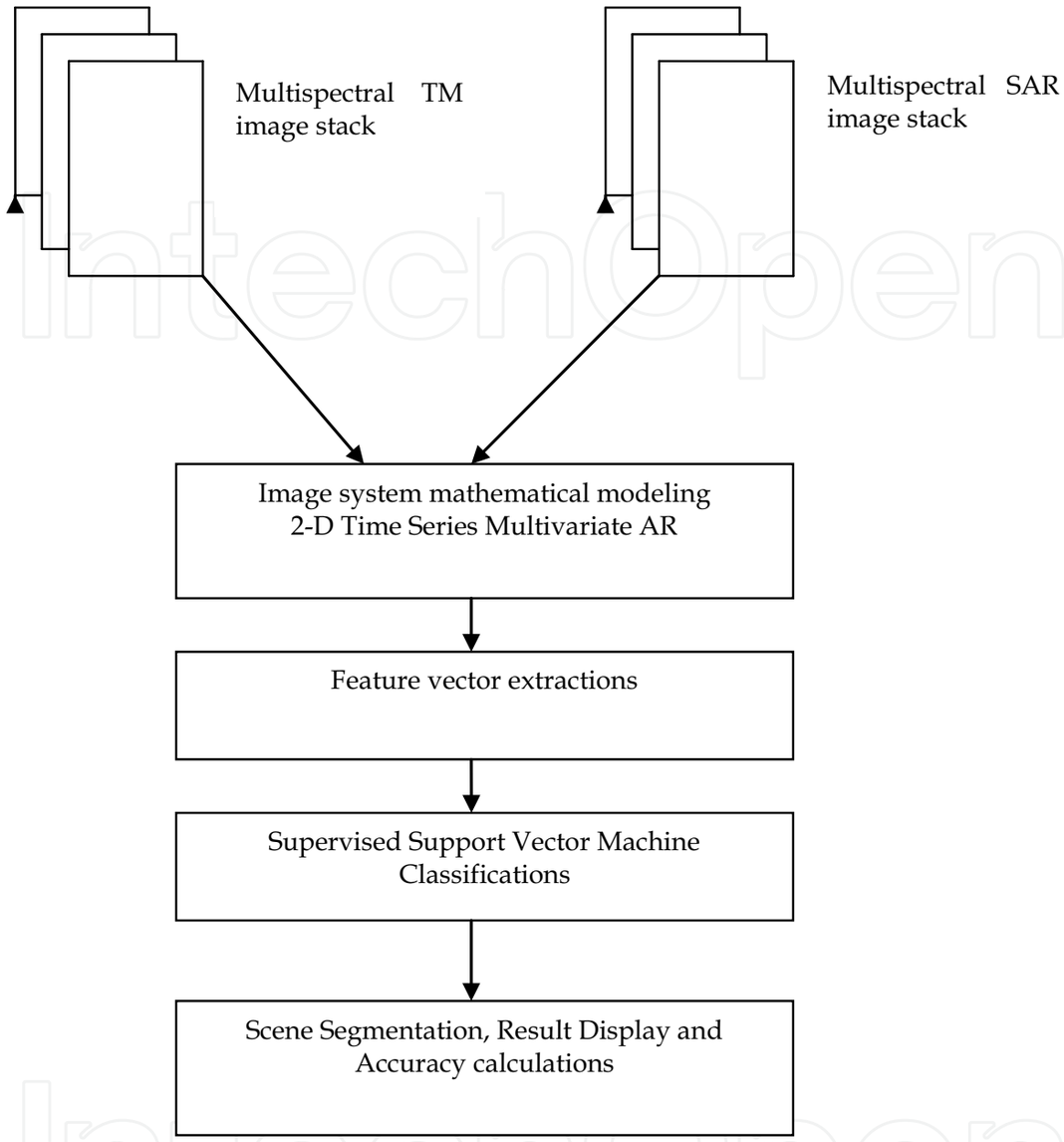


Fig. 5. Multivariate Time Series AR parameter matrix estimation SVM system scheme

4. Univariate Time Series Region Growing

We assume airborne and satellite's remote sensing can be defined as a collection of random variables indexed according to the order they are obtained in time-space. For example, we consider a sequence of random variables $x_1, x_2, x_3, ...$, in general, $\{x_t\}$ indexed by t is the stochastic process. The adjacent points in time are correlated. Therefore, the value of series x_t at time t depends in some fashion on the past values $x_{t-1}, x_{t-2}, ...$ Suppose that we let the value of the time series at some point of time t to be denoted by x_t. A stationary time series is the one for which the probabilistic behavior of $x_{t_1}, x_{t_2}, ...x_{t_k}$ is

identical to that of the shifted set $x_{t_1+h}, x_{t_2+h}, ..., x_{t_k+h}$. In our remote sensing application, the 2-D image was scanned from left upper corner to right bottom as a sequence of time series pixel values. Further, to simplify the numerical calculations, we model each class of surface textures by 1st order and 2nd order Autoregressive stationary time series models. In another way of thinking, the two-dimensional Markov model is a similar mathematical model to describe an image area per remote sensing texture class. By using time series model, when the within-object interpixel correlation varies significantly from object to object, we can build effective classifiers. The unsupervised Region Growing is a powerful image segmentation method for use in shape classification and analysis. The LANDSAT 5 database in the area of Italy's Lake Mulargias remote sensing image data acquired in July 1996 to be used for the computing experiments with satisfactory results. The advanced statistical techniques, such as Gaussian distributed white noise error confidence interval calculations, sampling statistics based on mean and variance properties are adopted for automatic threshold finding during Region Growing iterations. The linear regression analysis with least mean squares error estimation is implemented as a time series system optimization scheme (Chen, 2003). The classification and segmentation results are shown in Figure 6,7,8.

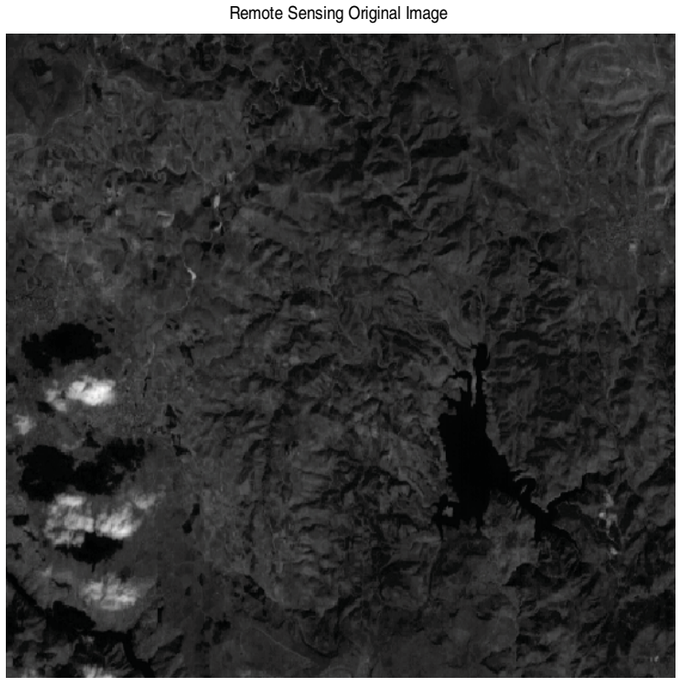


Fig. 6. Original Lake Region Remote Sensing band 5 image



Fig. 7. Ground Truth Information

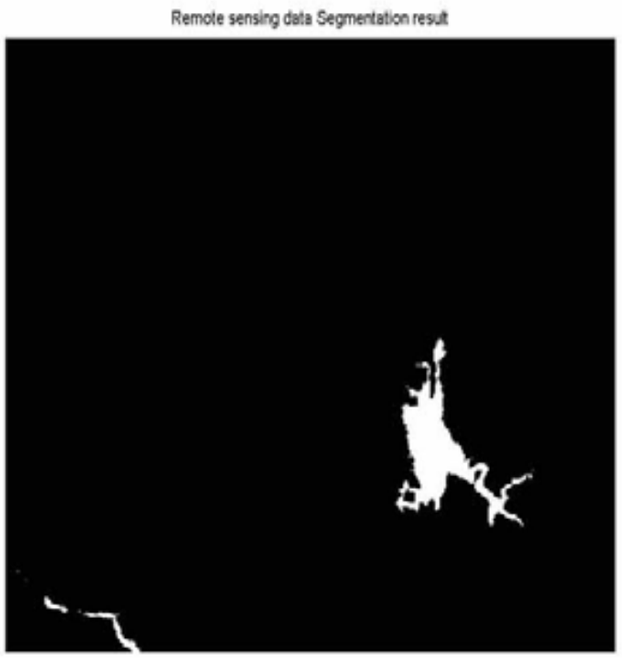


Fig. 8. Segmentation Result After Region Growing Based On univariate AR Model

5. Multivariate AR Model and Error Covariance Matrix Estimation

The remote sensing image data widely used in military, geographic survey and scientific research such as SAR (Synthetic Aperture Radar), TM (Thermal Mapper) are in multi-spectral format. As shown in figure 9, it consists of a stack of images. There are correlations between pixels in a single image as well as correlations among image slices. The Autoregressive (AR) model described in univariate time series model was based on Box-Jenkins system (Box, 1970) which is not enough to describe information extracted from multi-spectral satellite sensors. The innovative 2-D Multivariate Vector AR form (ARV) time series model described in this section is aimed to solve the problem.

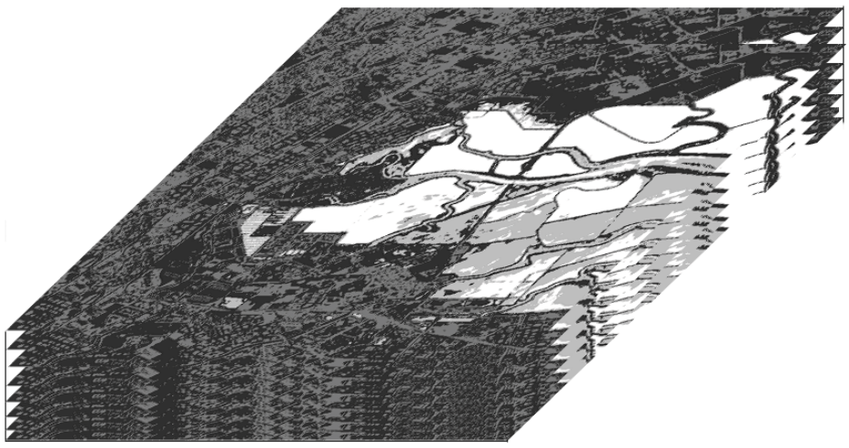


Fig. 9. Multi-spectral remote sensing image stack

In order to preserve most of the stacking remote sensing information for further 2-D image processing, there are methods such as Classification using Markov Random Fields and PCA method (Chen, 2002), Independent Component Analysis (Chen, 2002), …etc which require data compression before using 2-D single image classification technique. Both PCA and ICA methods are really time consuming. Lengthy computation time plus undetermined components to select make them unfeasible. On the contrary, the computation involved with the multivariate ARV method is the simplest matrix operations which are faster (roughly 2 times faster in computations). Besides, the classification accuracy results have shown that this new Multivariate Vector AR method is feasible and superior.
The Multivariate Time Series data analysis model is a generalized vector form of the Box-Jenkins' univariate time series model. It is also called the ARMAV (AutoRegressive Moving Average Vector) model. The fact is that each time series observation is a vector containing multiple factors.

Let $X_i = [x_{1i}, x_{2i}, x_{3i}, ...x_{mi}]^T \qquad -\infty \le i \le \infty$

$$X_i = W + \phi_1 X_{i-1} + \phi_2 X_{i-2} + ... + \phi_p X_{i-p} + \varepsilon_i - \theta_1 \varepsilon_{i-1} - \theta_2 \varepsilon_{i-2} - ..... - \theta_q \varepsilon_{i-q} \quad (1)$$

X_i : m-by-1 column vector, time series state variable

ε_i : m-by-1 column vector, multivariate white noise

$\phi_k : k = 1,2,3,.....p$ m-by-m autoregressive parameter matrix

$\theta_k : k = 1,2,3,.....q$ m-by-m moving average parameter matrix

W : m-by-1 Constant Vector (deterministic DC term)

AutoRegressive Vector (ARV) is reduced to:

$$X_i = W + \phi_1 X_{i-1} + \phi_2 X_{i-2} + ... + \phi_p X_{i-p} + \varepsilon_i \tag{2}$$

Example: m=2, p=2 , W=0 of ARV model

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \phi_{1,11} \phi_{1,12} \\ \phi_{1,21} \phi_{1,22} \end{bmatrix} \begin{bmatrix} x_{i-1} \\ y_{i-1} \end{bmatrix} + \begin{bmatrix} \phi_{2,11} \phi_{2,12} \\ \phi_{2,21} \phi_{2,22} \end{bmatrix} \begin{bmatrix} x_{i-2} \\ y_{i-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \tag{3}$$

Estimation of the Multivariate AR model Parameter Matrix:
Optimization Method: Least Squares
Let's assume the system is a m-dimensional time series
An AR(p) time series model can be expressed as the following regression model:

$$X_v = BU_v + \varepsilon_v \tag{4}$$

where ε_v = noise vector with covariance matrix C $v = 1,2,.....n$ and n is the total number of samples.
Therefore, $C = E(\varepsilon_v \varepsilon_v^T)$
The Parameter Matrix $B = \lfloor W \; \phi_1 \quad \phi_2 \quad .... \quad \phi_p \rfloor$

$$\text{Define} \quad U_v = \begin{bmatrix} 1 \\ X_{v-1} \\ X_{v-2} \\ \vdots \\ X_{v-p} \end{bmatrix} \qquad V = \begin{bmatrix} X_1^T \\ X_2^T \\ \cdot \\ \cdot \\ \cdot \\ X_n^T \end{bmatrix} \tag{5}$$

Let's also define:

$$T = \sum_{v=1}^{n} U_v U_v^T \tag{6}$$

$$X = \sum_{v=1}^{n} X_v X_v^T \tag{7}$$

$$S = \sum_{v=1}^{n} X_v U_v^T \tag{8}$$

The least squares estimation of the B matrix can be found as

$$X_v = BU_v + \varepsilon_v \tag{9}$$

$$X_v U_v^T = BU_v U_v^T + \varepsilon_v U_v^T \tag{10}$$

$$\sum_{v=1}^{n} X_v U_v^T = \sum_{v=1}^{n} \hat{B} U_v U_v^T \tag{11}$$

$$S = \hat{B}T \tag{12}$$

Therefore, the estimated parameter matrix is

$$\hat{B} = ST^{-1} \tag{13}$$

The error covariance matrix is

$$\hat{C} = \frac{1}{n - nf} \sum_{v=1}^{n} \hat{\varepsilon}_v \hat{\varepsilon}_v^T \tag{14}$$

Where n is the total number of samples and nf is the degree of freedom (Glantz,2002). In normal cases, this parameter can be ignored.
We have

$$\hat{\varepsilon}_v = X_v - \hat{B}U_v \tag{15}$$

$$\hat{\varepsilon}_v^T = X_v^T - U_v^T \hat{B}^T \tag{16}$$

$$\hat{\varepsilon}_v \hat{\varepsilon}_v^T = X_v X_v^T - \hat{B}U_v X_v^T - X_v U_v^T \hat{B}^T + \hat{B}U_v U_v^T \hat{B}^T \tag{17}$$

$$\hat{\varepsilon}_v \hat{\varepsilon}_v^T = X_v X_v^T - ST^{-1} U_v X_v^T - X_v U_v^T T^{-1} S^T + ST^{-1} U_v U_v^T T^{-1} S^T \tag{18}$$

$$\sum_{v=1}^{n} \hat{\varepsilon}_v \hat{\varepsilon}_v^T = \sum_{v=1}^{n} X_v X_v^T - ST^{-1} \sum_{v=1}^{n} U_v X_v^T - \sum_{v=1}^{n} X_v U_v^T T^{-1} S^T + ST^{-1} \sum_{v=1}^{n} U_v U_v^T T^{-1} S^T \tag{19}$$

$$\sum_{v=1}^{n} \hat{\varepsilon}_v \hat{\varepsilon}_v^T = \sum_{v=1}^{n} X_v X_v^T - ST^{-1} S^T - ST^{-1} S^T + ST^{-1} TT^{-1} S^T \tag{20}$$

$$\sum_{v=1}^{n} \hat{\varepsilon}_v \hat{\varepsilon}_v^T = X - ST^{-1} S^T \tag{21}$$

Therefore,

$$\hat{C} = \frac{1}{n - nf} (X - ST^{-1} S^T) \tag{22}$$

The error covariance matrix $\hat{C}$ is similar to a Schur complement matrix
Let's define the Schur complement matrix as

$$M = \begin{bmatrix} T & S^T \\ S & X \end{bmatrix} = \sum_{v=1}^{n} \begin{bmatrix} U_v \\ X_v \end{bmatrix} \begin{bmatrix} U_v^T & X_v^T \end{bmatrix} \tag{23}$$

which is the moment matrix

$$M = K^T K \tag{24}$$

where

$$K = \begin{bmatrix} U_1^T & X_1^T \\ U_2^T & X_2^T \\ . & . \\ U_n^T & X_n^T \end{bmatrix} \tag{25}$$

The least squares estimate can be computed from a QR factorization of the data matrix K. According to QR decomposition theorem (Anton, 2000), (Moler, 2004), (Cheney, 1999) and (Chapra 2002), if K is n by $(1+(p+1)*m)$ matrix with linearly independent column vectors, then K can be factored as

$$K = QR \tag{26}$$

where Q is an nxn matrix with orthonormal column vectors, and R is an n by $(1+(p+1)*m)$ upper triangular matrix.

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \\ 0 & 0 \end{bmatrix} \tag{27}$$

$$R^T = \begin{bmatrix} R_{11}^T & 0 & 0 \\ R_{12}^T & R_{22}^T & 0 \end{bmatrix} \tag{28}$$

The submatrices R_{11} and R_{22} are upper triangular square matrices of order $p \times m + 1$ and m, respectively. The submatrix R_{12} has a dimension of $p \times m + 1$ by m.
According to matrix numerical analysis, the QR factorization of the data matrix K leads to the Cholesky factorization

$$M = K^T K = R^T Q^T Q R = R^T I R = R^T R \tag{29}$$

of the moment matrix.
Therefore,

$$M = \begin{bmatrix} T & S^T \\ S & X \end{bmatrix} = \sum_{v=1}^{n} \begin{bmatrix} U_v \\ X_v \end{bmatrix} \begin{bmatrix} U_v^T & X_v^T \end{bmatrix} = R^T R = \begin{bmatrix} R_{11}^T R_{11} & R_{11}^T R_{12} \\ R_{12}^T R_{11} & R_{12}^T R_{12} + R_{22}^T R_{22} \end{bmatrix} \tag{30}$$

Therefore,

$$S = R_{12}^T R_{11} \tag{31}$$

$$T = R_{11}^T R_{11} \tag{32}$$

From equation (13) $\hat{B} = ST^{-1}$ and equation (22) $\hat{C} = \frac{1}{n-nf}(X - ST^{-1}S^T)$
we can then derive the estimated parameter matrix $\hat{B}$ and error covariance matrix $\hat{C}$:

$$\hat{B} = (R_{11}^{-1} R_{12})^T \tag{33}$$

$$\hat{C} = \frac{1}{n-nf}(R_{22}^T R_{22}) \tag{34}$$

where again n is the sample size and nf is the degree of freedom based on statistical sampling theory.

The error covariance matrix $\hat{C}$ is further factorized by Cholesky decomposition. The diagonal terms of this factorized Cholesky matrix are used as feature vector inputs for Support Vector Machine (SVM).

$$\hat{C} = E^T E \tag{35}$$

Where E is the upper triangular matrix.

Suppose E is an m-by-m upper triangular, lower triangular or diagonal matrix , the eigenvalues of E are the entries on the main diagonal of E. The eigenvalues of E matrix are the characteristics that E contains. Based on theorem, the E matrix contains the major information on the estimated error covariance matrix $\hat{C}$.

Note that the theory behind QR decomposition, Cholesky factorization and the Schur Complement form will be detailed in section 5.2, section 5.3 and section 5.4 respectively.

5.1. Numerical Simulation

The following 2 simulated experimental examples will explain and verify the multivariate ARV model algorithm described above.

Example 1:

From Equation (3)

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \phi_{1,11}\,\phi_{1,12} \\ \phi_{1,21}\,\phi_{1,22} \end{bmatrix} \begin{bmatrix} x_{i-1} \\ y_{i-1} \end{bmatrix} + \begin{bmatrix} \phi_{2,11}\,\phi_{2,12} \\ \phi_{2,21}\,\phi_{2,22} \end{bmatrix} \begin{bmatrix} x_{i-2} \\ y_{i-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \tag{36}$$

$$B_1 = \begin{bmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.28 \\ 1.3 & 1 \end{bmatrix} \tag{37}$$

$$B_2 = \begin{bmatrix} \phi_{2,11} & \phi_{2,12} \\ \phi_{2,21} & \phi_{2,22} \end{bmatrix} = \begin{bmatrix} 0.35 & -0.4 \\ -0.3 & -0.5 \end{bmatrix} \tag{38}$$

$$\text{Let} \qquad B = \begin{bmatrix} B_1 & B_2 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.5 & 0.28 & 0.35 & -0.4 \\ 1.3 & 1 & -0.3 & -0.5 \end{bmatrix} \tag{39}$$

$$C = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{bmatrix} \tag{40}$$

The 200 simulated 2 elements vector of Multivariate Time Series random process data set V_{200x2} was generated by defining this 2nd order ARV parameter matrix:

$$V = \begin{bmatrix} 0.8778 & -2.1116 \\ 2.717 & 4.0011 \\ 1.8159 & 6.3244 \\ 1.5357 & 6.9345 \\ 1.2415 & 4.9586 \\ 0.5507 & 2.3339 \\ . & . \\ . & . \\ 3.2515 & 12.5413 \\ 2.8186 & 10.1881 \end{bmatrix} \quad \text{where } V \text{ was defined in (5)} \tag{41}$$

After multivariate ARV time series model estimation process, we get the estimated:

$$\hat{B} = \begin{bmatrix} 0.5985 & 0.2055 & 0.3347 & -0.3394 \\ 1.4657 & 1.0258 & -0.5929 & -0.4753 \end{bmatrix} \tag{42}$$

$$\hat{C} = \begin{bmatrix} 1.1896 & 0.6626 \\ 0.6626 & 1.7277 \end{bmatrix} \tag{43}$$

We can see that both the estimated parameter matrix and error covariance matrix by the above least squares algorithm is accurate by measuring a small percentage of matrix norm of matrices difference:

$$\hat{B} \approx B \qquad \hat{C} \approx C$$

Example 2:

$$B = \begin{bmatrix} 0.5 & 1.3 & 0.35 & -0.3 \\ 0.28 & 1 & -0.4 & -0.5 \end{bmatrix} \tag{44}$$

$$C = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{bmatrix} \tag{45}$$

Estimation result:

$$\hat{B} = \begin{bmatrix} 0.35437 & 1.3809 & 0.4864 & -0.27223 \\ 0.25398 & 1.0096 & -0.36946 & -0.47929 \end{bmatrix} \tag{46}$$

The estimated parameter matrix is also accurate in this 2nd example.

5.2 QR Algorithm

To find the eigenvalue decomposition is to find a diagonal matrix Λ and a nonsingular matrix S such that

$$A = S\Lambda S^{-1} \tag{47}$$

Two problems might occur. First is that the decomposition may not exist. Secondly, even if the decomposition exists, it might not be robust enough. The way to solve the problems is to get as close to diagonal as possible. Ideally we would like to find the Jordan canonical form of the matrix, however, this is not practical to do in finite precision arithmetic. Instead we compute the Schur decomposition of the matrix $A = TBT^H$, where B is the upper triangular and T is unitary. Every square matrix has a Schur decomposition which can be computed using an iterative algorithm. The algorithm is called the QR algorithm since it performs a QR factorization in each iteration. The eigenvalues of A are on the diagonal of its Schur form of B. Since the unitary transformations are perfectly well conditioned, they do not magnify errors. The diagonal elements of B are the eigenvalues of A if A is symmetric and B is diagonal. In this case, the column vectors of T are orthonormal eigenvectors of A. In general, the large off-diagonal elements of B measure the lack of symmetry in A. In the non-symmetric case, the eigenvectors of A are the column vectors of $G = TX$, where X is a matrix contains the eigenvectors of the upper triangular matrix B. The QR algorithm computes the eigenvalues of real symmetric matrices, real nonsymmetric matrices and complex matrices. The singular values of general matrices are computed using the Golub-Reinsch algorithm which is based on the QR algorithm.

5.3 Cholesky Factorization

In mathematics, the Cholesky factorization (or decomposition) (Kreyszig, 1999) is named after Andre-Louis Cholesky, who found that a symmetric positive-definite matrix can be decomposed into a lower triangular matrix and the transpose of the lower triangular matrix. The lower triangular matrix is the Cholesky triangle of the original, positive-definite matrix. Cholesky's result has since been extended to matrices with complex entries.

Any square matrix A with non-zero pivots can be written as the product of a lower triangular matrix L and an upper triangular matrix U; this is called the LU decomposition. However, if A is symmetric and positive definite, we can choose the factors such that U is the transpose of L, and this is called the Cholesky decomposition. Both the LU and the Cholesky decomposition are used to solve systems of linear equations. When it is applicable, the Cholesky decomposition is twice as efficient as the LU decomposition.

5.4 The Schur Complement Form

Let's understand what is Schur complement form. It is a block of a matrix within the larger matrix which is defined as follows. Suppose A, B, C, D are respectively $p\times p$, $p\times q$, $q\times p$ and $q\times q$ matrices, and D is invertible. Let

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{48}$$

so that M is a $(p+q)\times(p+q)$ matrix.

6. Support Vector Machine classifier

Support Vector Machine (SVM), a new class of machine learning with superior classification capability over other learning algorithms . It can be analyzed theoretically using concepts from statistical learning theory (Vapnik, 1998) and at the same time to achieve good performance when applied to real problems. We have implemented software on top of SVM-KM matlab toolbox developed by Dr. A. Rakotomamonjy of INSA, France for remote sensing image classification and region segmentations. The results on both Multivariate AR model (ARV) and non-AR pixel-by-pixel based methods are reasonably good. Of course, the ARV's additional contextual information gives a better performance. The main idea of this classification is to construct a hyperplane as a decision surface in a way that the margin of separation between positive and negative classes is maximized. The support vector machine can provide a good generalization performance on pattern classification problems. The SVM constructs models that are complex , it contains a large class of neural networks, radial basis function and polynomial classifiers as the special cases. Nevertheless, it is also simple enough to be analyzed mathematically due to the fact that it can be shown to correspond to a linear method in a high dimensional feature space nonlinearly related to input space. By use of kernels, all needed computations are performed directly in the input space. There are different cases of SVM, the first case is the linear separable data to be trained on linear machine. The 2nd case is the nonlinear SVM trained on non-separable data, this will result in a quadratic programming problem. The SVM can work with high dimensional data as the method is less dependent on the statistical distribution of the data. The SVM avoids the "Curse of Dimensionality" problems typically experienced in statistical pattern classifications (Chen, 1999).

SVM were specifically designed for binary classification. The multiclass applications on SVM is still an on-going popular research topics. A few straight forward methods have been proposed in such a way that multiclass classifier can be constructed by combining several binary classifiers. Some other researchers are proposing to create multi-classifiers that process multi-class data at once but it is most likely less accurate. The multi-class remote sensing SVM classifier we developed that describes in this chapter is based on One-Again-One pairwise with majority votes.

Support Vector Machine Concepts

The SVM in most cases is competitive among the existing classification methods. It is relatively easy to use. We assume the data vectors $\mathbf{x} = [x_1, x_2, ...x_n]^T$

$x_j \quad j = 1,.....n$ where n is the number of features in the feature space

Let's also consider simple case of two classes data set :

Define classification output indicator vector $Y = [y_1, y_2, ..., y_l]^T$

$$y_i = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ in class 1} \\ -1 & \text{if } \mathbf{x}_i \text{ in class 2} \end{cases}$$

A hyperplane $\mathbf{w}^T\mathbf{x} + b = 0$ is to be found to separate all data.

Where $\mathbf{w}$ is an weight vector and b is a bias from the origin.

A separable hyperplane is shown in figure 10.

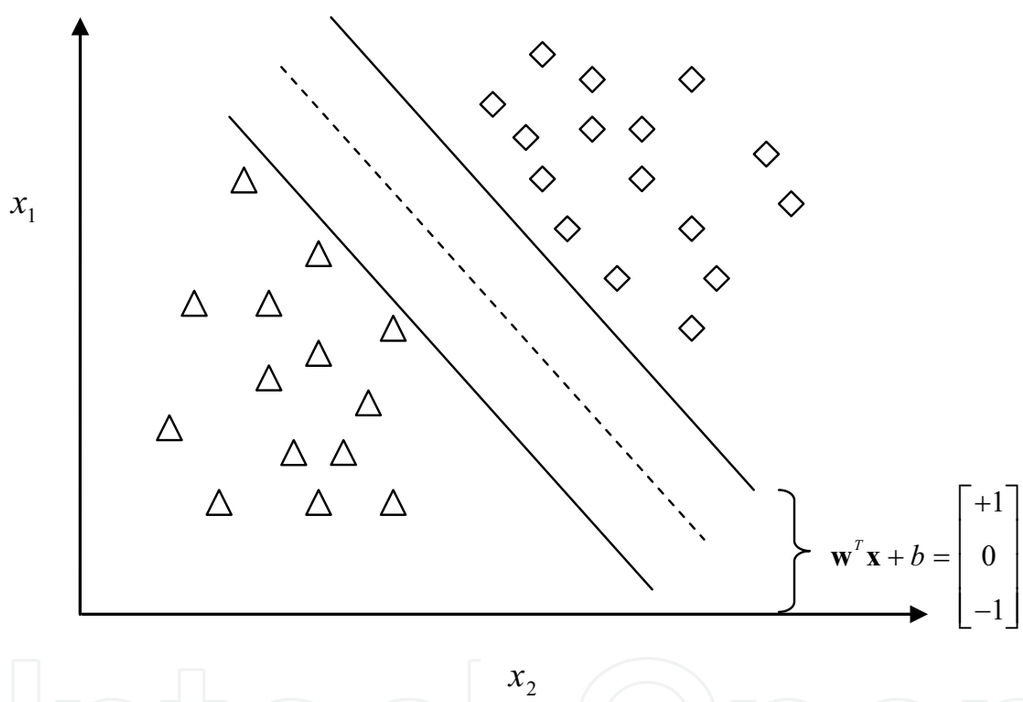


Fig. 10. Optimal hyperplane separator (in the case of 2 dimension feature space)

To illustrate the concept of SVM kernel mapping, we show in figure 11 how to map the input space to a feature space such that the non-separable input space can be separable in the feature space. This is done by nonlinear mapping to higher dimension and constructing a separating hyperplane with a maximum margin.

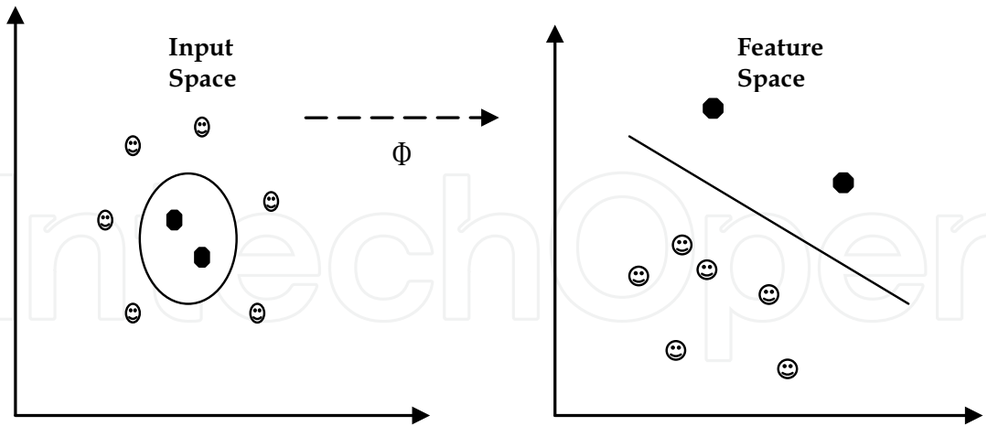


Fig. 11. Map training data nonlinearly into a higher dimensional feature space

Feature space kernel mapping example:

Figure 11 shows the basic idea of SVM which maps data into dot product space (feature space) F by a nonlinear mapping:

$$\Phi : R^N \rightarrow F \tag{49}$$

The dot product is

$$k(\mathbf{x},\mathbf{y}) := (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})) \tag{50}$$

If F is in high dimension, the right hand side of equation will be expensive to compute.

7. Experimental Results

The UK village remote sensing data set were used and the Multivariate Time Series SVM performance is compared with other existing classification techniques. This image data set was kindly offered by Italian Remote Sensing Research group (Serpico 1995, Bruzzone 1999). It consists of a set of 250x350 pixel images acquired by two imaging sensors installed on a 1268 Airborne Thematic Mapper (ATM) scanner and a PLC band, full polarimetric NASA JPL SAR sensor. For performance comparison, Table 2 shows the experimenting results of the classification accuracies by several remote sensing popular classifiers (Ho, 2008):

Method 1: Structured Neural Networks (TLN) by University of Genoa, Italy (Serpico 1995)
Method 2: K-mean + PCA (Ho, 2008)
Method 3: K-mean + ICA (Ho, 2008)
Method 4: K-mean + (PCA+MRF) (Ho, 2008)
Method 5: FCM + (PCA+MRF) (Ho, 2008)
Method 6: Multivariate ARV Support Vector Machine (Ho, 2008)

Method 1: **Structured Neural Networks (TLN) method:**
This method uses the architecture of structured multilayer feedforward networks. The networks are trained to solve the problem by the error backpropagation algorithm. They are transformed into equivalent networks to obtain a simplified representation. It is considered to be a hierarchical committee that accomplishes the classification task by checking on a set of explicit constraints on input data.
Method 2: **K-mean + PCA:**
Principal Component Analysis (PCA) is a technique to reduce multidimensional data to a lower dimension for analysis. By constructing the transformation matrix which consists of eigenvectors and ordering the eigenvalues, the statistically uncorrelated components can be extracted. It is an unsupervised approach that finds the best features from data. K-means clustering is a form of stochastic hill climbing in the log-likelihood function. The contours in the feature space represents equal log-likelihood values. It iteratively calculates the mean vectors in the selected classes. As more data are inputted, it dynamically adjusts until there is no change on mean vectors.
Method 3: **K-mean + ICA**:
Independent Component Analysis (ICA) is a computational method to separate a multivariate signal into additive subcomponents which are statistically independent and at least one of which is a non-Gaussian source signal. Method 3 is the same as Method 2 except that PCA is replaced by ICA.
Method 4: K-mean + (PCA+MRF):
Markov Random Field (MRF) theory provides a basis for modeling contextual constraints in image processing. It is a model of the joint distribution of a set of random variables. The estimated pixels based on MRF model form a new image. PCA is applied to each pixel vector of the new image stack. This is followed by K-mean operation.
Method 5: FCM + (PCA+MRF):
Fuzzy-C-Mean (FCM) is an iterative clustering method. In every iteration of the classical K-Means procedure, each data point is assumed to belong to exactly one cluster. But in FCM, we relax this condition and assume that each sample has some graded or fuzzy membership in a cluster. Method 5 is the same as Method 4 except that K-mean is replaced by FCM operation.
Method 6: Multivariate ARV Support Vector Machine
This method was newly developed as described in the sections earlier.

Figure 12,13,14,15 show the results of region segmentation by this novel multivariate Time Series model based SVM classification method as described in this chapter.
Figure 12. is the original th-c-hh (one example of 15 remote sensing input data). Figure 13. is the Multivariate AR – SVM without smooth post-processing result. Figure 14. is the Multivariate AR – SVM with post-processing result and Figure 15. is the Multivariate AR – SVM with post-processing and tone re-scaling result.

	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6
Accuracy	86.49%	64.93%	65.79%	72.46%	75.27%	87.11%

Table 2. Classification performance comparison

The followings are the color bar and crop identifications in UK village remote sensing image data set:

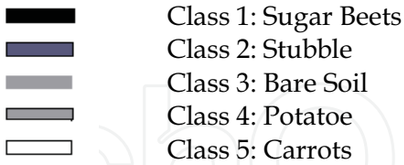


Figure 12. Original th-c-hh (one example of 15 remote sensing input data)
Figure 13. Multivariate AR – SVM without smooth post-processing result
Figure 14. Multivariate AR – SVM with post-processing result
Figure 15. Multivariate AR – SVM with post-processing and tone re-scaling result


Fig.12


Fig. 13

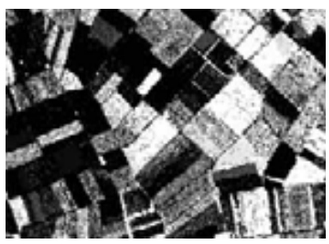

Fig. 14


Fig. 15

8. Conclusion

This chapter has focused on contexture information in order to improve remote sensing pixel classification accuracy. The study of time series is primarily concerned with time or spatial correlation structure. The time series model has its world wide applications in the area of finance and oceanography but not much in remote sensing image processing. We took the big challenge to develop this new idea for remote sensing information processing. As we mentioned earlier, the time series models were started from univariate system to move gradually toward complicated multivariate matrix structures in order to solve the multiple image pixel stack problems. The system optimization and image pixel estimation solutions were changed from simple derivative, linear system to complex matrix decompositions by numerical analysis. It does open up a new era away from traditional approaches as far as remote sensing image processing is concerned. Numerical methods are the arithmetic operations to solve the mathematically formulated problems. Although they are involved with large numbers of tedious calculations, the most recent fast and efficient

digital computer can solve them quickly. The role of numerical analysis in solving the engineering problem has increased dramatically. They are capable of handling large systems of equations, nonlinearities and geometries that are often impossible to solve analytically. The remote sensing data mining is huge. The system is initially gained by empirical means through observations and experiments. As new measurements are taken, the generalizations might be modified or newly developed. On the other hand, generalizations can have a strong influence on the observations. Hopefully the conclusions can be drawn eventually. From an electrical and computer engineering problem solving perspective, a system's mathematical model has to be usefully expressed. We are hoping that system mathematical modeling and the powerful numerical methods that take full advantage of fast computing tools can resolve most of the remote sensing issues. Take one example in this chapter: the Multivariate Time Series model with system parameter matrix estimation and error covariance matrix (solved by Cholesky decomposition and the stable QR algorithm) might be able to capture remote sensing research attentions for their complex unsolvable image stacking problems. Though originality of this method exists, the algorithm stability does require years of testing by researchers and interested parties. The novel Estimation of Parameters Matrix of Multivariate Time Series techniques we developed in this chapter can be useful not only for image classification but also good for other research areas such as accurate stock market predictions, video prediction for wireless multimedia applications, adaptive frame prediction for scalable video coding …etc. The reasons are that multivariate time series can handle multiple data factors at every single time stamp. It can also expand the traditional Digital Signal Processing capability which is mostly in univariate time sequences. Advanced methods for automatic analysis of multisensor and multitemporal remote sensing images is still on-going in the years to come. Image processing and pattern recognition researches have been going on for over half century since the digital computer was invented. Many excellent and useful tools such as Maple, Matlab, IDS, Labview …etc have been created accordingly. Millions of image processing algorithms and great research results were also published. Remote sensing, image processing and pattern recognition related community have generated many journal papers and held developer conferences each year around the world to exchange ideas. Still, none of the "final" universal optimal algorithm has been done successfully yet. Take an example of the remote sensing texture classifications, it is difficult to obtain a good texture representation and to have the true adaptive function for segmentation for all kinds of remote sensing data.

9. Future Work

Though the multivariate ARV model was developed, to select the optimal order of an ARV model is one of complicated problem to solve. The error covariance matrix $\hat{C}$ has to be computed and the order selection criterion such as famous Akaike's Final Prediction Error (FPE) (Akaike, 1971) criterion, Schwarz's Bayesian Criterion (SBC) (Schwarz, 1978) and Lutkepohl's improved order criterion (Lutkepohl, 1985) has to be determined in order to decide the optimal ARV system order. As far as today's most decent and generalized data classification method –Support Vector Machine, there are still many holes to be filled and explored. To mention a few, improving system model formulation for extracting more useful feature vectors for SVM , new kernel functions developments for better SVM

classification performance, convex and non-convex optimization theory, intelligent numeric computational methods …etc are the open research area for the near future.

10. References

Akaike, H. Autoregressive Model Fitting for Control. *Ann. Inst. Statistical Mathematics* 23, 1971, pp. 163-180.

Anton, H. Rorres, C. *Elementary Linear Algebra, Applications,* John Wiley and Sons 2000.

Bruzzone, L. and Prieto, D. A Technique for The Selection Of Kernel Function Parameters in RBF neural Networks for Classification of Remote Sensing Images, *IEEE Trans. Geosci. Remote Sensing,* Vol. 37, No.2, pp. 1179-1184, March 1999.

Bruzzone, L. *Advanced Methods for the Analysis of Multisensor and Multitemporal Remote-Sensing Images*, PhD dissertation 1998, Electronics and Computer Science, Univeristy of Genova, Italy.

Chapra, S. Canale, R. Numerical Methods for Engineers with Software and Programming Applications, 4th edition, McGraw Hill 2002.

Chen, C. editor, Information Processing for Remote Sensing, World Scientific 1999, pp. 12-13

Chen, C. and Ho, P. On the ARMA model based Region Growing method for extracting lake region in a Remote Sensing image, on *Proceedings of SPIE*, Vol. 5238, Paper No. 5238-13, September 8-12, 2003, Barcelona, Spain.

Chen, H. Remote Sensing Data Unsupervised Classification Using Gauss-Markov Random Fields and PCA, MS thesis 2002 ECE University of Massachusetts at Dartmouth.

Cheney, W. and Kincaid, D. *Numerical Mathematics and Computing*, 4th edition, Publisher: Gary W. Ostedt Book 1999.

Glantz, S. *Primer of Biostatistics*, 5th edition, McGraw-Hill 2002.

Box, G. and Jenkins, G. *Time Series Analysis: Forecast and Control*, San Francisco, Holden-Day, 1970.

Ho, P. *Multivariate time series model based support vector machine for multiclass remote sensing image classification and region segmentation,* Ph.D dissertation, Univ. of Massachusetts Dartmouth, Jan. 2008.

Kreyszig, E. *Advanced Engineering Mathematics*, 8th edition, John Wiley and Sons 1999.

Lutkepohl, H. Comparison of Criteria for Estimating the Order of a Vector Autoregressive Process. *Journal of Time Series*, Vol 6, 1985, pp. 35-52.

Moler, C. *Numerical Computating with Matlab*, Chapter 10, SIAM 2004 .

Prieto, D. *Automatic Analysis of Multisource and Multitemporal Remote-Sensing Data for Land-Cover Monitoring Application*, Ph.D Thesis, University of Genova, 2000.

Richards, J. and Jia, X. *Remote Sensing Digital Image Analysis*, Springer-Verlag 1999.

Schowengerdt, R. Remote Sensing, Models and Methods for Image Processing. 2nd edition, Academic Press, 1997.

Serpico S. and Roli, F. Classification of Multisensor Remote Sensing Images by Structured Neural Netorks, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 33, No.3, pp. 562-578, May 1995.

Schwarz, G. Estimating the Dimension of a Model. *Ann. Statistics* Vol. 6, 1978, pp. 461-464.



Geoscience and Remote Sensing

Edited by Pei-Gee Peter Ho

ISBN 978-953-307-003-2
Hard cover, 598 pages
Publisher InTech
Published online 01, October, 2009
Published in print edition October, 2009

Remote Sensing is collecting and interpreting information on targets without being in physical contact with the objects. Aircraft, satellites ...etc are the major platforms for remote sensing observations. Unlike electrical, magnetic and gravity surveys that measure force fields, remote sensing technology is commonly referred to methods that employ electromagnetic energy as radio waves, light and heat as the means of detecting and measuring target characteristics. Geoscience is a study of nature world from the core of the earth, to the depths of oceans and to the outer space. This branch of study can help mitigate volcanic eruptions, floods, landslides ... etc terrible human life disaster and help develop ground water, mineral ores, fossil fuels and construction materials. Also, it studies physical, chemical reactions to understand the distribution of the nature resources. Therefore, the geoscience encompass earth, atmospheric, oceanography, pedology, petrology, mineralogy, hydrology and geology. This book covers latest and futuristic developments in remote sensing novel theory and applications by numerous scholars, researchers and experts. It is organized into 26 excellent chapters which include optical and infrared modeling, microwave scattering propagation, forests and vegetation, soils, ocean temperature, geographic information , object classification, data mining, image processing, passive optical sensor, multispectral and hyperspectral sensing, lidar, radiometer instruments, calibration, active microwave and SAR processing. Last but not the least, this book presented chapters that highlight frontier works in remote sensing information processing. I am very pleased to have leaders in the field to prepare and contribute their most current research and development work. Although no attempt is made to cover every topic in remote sensing and geoscience, these entire 26 remote sensing technology chapters shall give readers a good insight. All topics listed are equal important and significant.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Pei-Gee Peter Ho (2009). Multivariate Time Series Support Vector Machine for Multispectral Remote Sensing Image Classifications, Geoscience and Remote Sensing, Pei-Gee Peter Ho (Ed.), ISBN: 978-953-307-003-2, InTech, Available from: http://www.intechopen.com/books/geoscience-and-remote-sensing/multivariate-time-series-support-vector-machine-for-multispectral-remote-sensing-image-classificatio

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China

www.intechopen.com

51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen

IntechOpen

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen