ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

*INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING*
*Vol. 2, Issue 5, May 2014*

# Classification of Satellite images based on SVM classifier Using Genetic Algorithm

**Moje Ravindra K[1], Patil Chandrashekhar G[2]**

Department Of Electronics & Telecommunication, Sinhgad Academy Of Engineering, Pune, India[1,2]

**Abstract:** Support vector machine (SVM) is originally developed for linear two-class classification via constructing an optimal separating hyper plane, where the margin is maximal. In case of not linearly separable training data, SVM is by means of kernel trick to map the original input space into a high dimensional feature space to enhance the classifier generalization ability. Genetic Algorithm (GA) is a stochastic and heuristic searching algorithm that is inspired by natural evolution. In the evolution, the candidate solutions are encoded to a group of strings (called chromosomes) by some kind of encoding method. Based on Darwin's principle of 'survival of the fittest', the optimal candidate solution is obtained after a series of iterative GA computations. In each process of iteration (called generation), the GA consists of the elementary operation of Selection, Crossover and Mutation. In a GA, the fitness function is used to evaluate the quality of each individual comes out of the chromosomes. Individuals with high fitness are easier to be inherited to the next generation. By using GA along with SVM here we are trying to make classification of the objects such that it will be closer to the original image. This is simple effort to make identification easier.

**Keywords:** Support vector machine (SVM), Genetic Algorithm (GA)

## I. INTRODUCTION

With the advancement of the satellite instruments technology, the spatial resolution of earth observation data is increasing dramatically. The spatial resolution of panchromatic images acquired from World View (lunched in 2006) or GeoEye-1, 2(2008, 2010) already achieved sub-meters. High resolution remote sensing images have regarded as a cost-effective means to assist urban planning, environmental monitoring, land cover/land use etc. Compared with middle resolution remote sensing images, high resolution images provide detailed structural and textural (or contextual) information for interpretation. However, the spectral reparability degrades compared with the increasing spatial resolution. It well known the traditional "pixel-based" method, which merely the spectral characteristics are concerned, is not suitable for high resolution images classification. In the mean time, object-oriented classification accepted as an alternative to analyze high resolution remote sensing has made a great success. [1]

An image object is a group of heterogeneity pixels generated by a process of image segmentation. Numerous characteristics of image objects can be calculated, including spectral, shape, textural as well as contextual information to improve classification rate. Support vector machine (SVM) original presented by Vipnik (1995) has been proved as a promising pattern classification approach, and has recently been effectively used in the field of remote sensing. SVM uses a kernel function map the low-dimensional input features into a high-dimension, such as linear kernels, polynomial kernels and radial basis function kernels (RBF). When using SVM, the primary issue is to choose a proper kernel function and set suitable kernel parameters. RBF as the frequently used function kernel in SVM, here we just study the parameter optimization of RBF. The parameters that should be optimized include generalization C, which determines the trade-off between maximum classification rate and minimum training error, and kernel parameter γ which defines the nonlinear mapping from the original low-dimensional input space into some high-dimensional feature space. An alternative optimization approach is to estimate the generalization abilities of SVMs using a gradient descent algorithm (e.g. grid searching) over the set of parameters. The parameters with respect to the maximum generalization ability are chosen for the SVM model. However, this method usually time consuming and does not perform well, we present a GA-based parameter optimization method for SVM.

## II. GENETIC ALGORITHM

A genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover.

In a genetic algorithm, a population of strings (called chromosomes or the genotype of the genome), which encode candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem, evolves toward better solutions. Traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 5, May 2014

(recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm.

Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. If the algorithm has terminated due to a maximum number of generations, a satisfactory solution may or may not have been reached. [7]

A typical genetic algorithm requires:

1. A genetic representation of the solution domain,
2. A fitness function to evaluate the solution domain.

*Genetic algorithms manipulate a population of* potential solutions to an optimization (or search) problem. Specifically, they operate on encoded representations of the solutions, equivalent to the genetic material of individuals in nature, and not directly on the solutions themselves. Holland's genetic algorithm encodes the solutions as strings of bits from a binary alphabet. As in nature, selection provides the necessary driving mechanism for better solutions to survive. Each solution is associated with a fitness value that reflects how good it is, compared with other solutions in the population. The higher the fitness value of an individual, the higher its chances of survival and reproduction and the larger its representation in the subsequent generation. Recombination of genetic material in genetic algorithms is simulated through a crossover mechanism that exchanges portions between strings. Another operation, called mutation, causes sporadic and random alteration of the bits of strings. Mutation too has a direct analogy from nature and plays the role of regenerating lost genetic material.

Holland's genetic algorithm is commonly called the Simple Genetic Algorithm or SGA. Essential to the SGA's working is a population of binary strings. Each string of 0s and 1s is the encoded version of a solution to the optimization problem. Using genetic operators- crossover and mutation –the algorithm creates the subsequent generation from the strings of the current population.

This generational cycle is repeated until a desired termination criterion is reached (for example a predefined number of generations are processed).

## III. SUPPORT VECTOR MECHANISM (SVM)

Support vector machine (SVM) is originally developed for linear two-class classification via constructing an optimal separating hyperplane, where the margin is maximal. In case of not linearly separable training data, SVM is by means of kernel trick to map the original input space into a high dimensional feature space to enhance the classifier generalization ability. In this section, we will briefly describe the basic concepts for non-linear SVM.

Given input training samples $\{(x_1, y_1), (x_2, y_2),\ldots, (x_m, y_m)\}$, where $x_i \in R^d$ represents a training instance which belongs to a class labelled by $y_i \in \{+1,-1\}$ .For the not linearly separable training data, the SVM maps the original input space into a high-dimensional feature space via kernel function .[3] The symmetric functions which satisfy the Mercers' condition could be served as SVM kernel function. Currently, the popular kernels are used in SVM including:

Radial basis function kernel:

$$k(x_i,x_j) = \exp(-\gamma \| x_i - x_j \|^2) \quad (1)$$

Linear kernel:

$$k(x_i,x_j) = x_i^T x_j \quad (2)$$

Polynomial kernel:

$$k(x_i,x_j) = (x_i^T x_j + 1)^d \quad (3)$$

Sigmoid kernel:

$$k(x_i,x_j) = \tanh[(x_i^T x_j + b)] \quad (4)$$

The separating hyperplane is defined as

$$D(x) = w^T . x+b \quad (5)$$

where w is an m-dimension vector, b is a bias term.

To obtain the optimal hyperplane, we need to minimize

$$Q(w, b, \xi) = \tfrac{1}{2} \| w \|^2 + C \sum \xi_i \quad (6)$$

Subject to the constraints

$$Y_i (w^T x_i + b) >= 1 - \xi_I \quad \text{for } I = 1,\ldots,m \quad (7)$$

Where $\xi_I$ are nonnegative slack variables, which measure the degree of misclassification of the datum $x_i$ .The constant C is a penalty parameter, which determines the trade-off between maximum classification rate and minimum training error. We call the obtained hyperplane the soft-margin hyperplane, where w is the soft-margin. Equation (6) is a quadratic optimization problem so that it is difficult to solve because of the w.

To solve above optimization problem, we can reformulate Equation (6) through a Lagrange function:

$$\text{Min max } \{\tfrac{1}{2} \|w\|^2 + C \sum \xi_I - \sum \alpha_i [y_i (w.x_i - b) - 1 + \xi_I] - \sum \beta_i \xi_I \} \quad (8)$$

Where $\alpha_i$, $\beta_i$ are the nonnegative Lagrange multipliers. Its dual form is $\text{Max} \sum \alpha_i - \tfrac{1}{2} \sum \alpha_i \alpha_f y_i y_f k(x_i, x_j)$ (9)

Subject to the constraints

$$\sum y_i \alpha_i = 0, C >= \alpha_i >= 0 \quad \text{for } I = 1,\ldots,m \quad (10)$$

Where the penalty nonnegative C represents the upper bound here. Finally, we obtain an optimal decision hyperplane

$$D(x) = \sum \alpha_i y_i k(x_i, x_j) + b \quad (11)$$

Where $k(x_i,x_j)$ the kernel function illustrated above.

The set S denotes a vector corresponding to the nonzero Lagrange multipliers I, which represents the so-called support vectors (SVs).[2]

## IV. GA-SVM BASED PARAMETER OPTIMIZATION METHOD

SVM has achieved a great success in the classification of remote sensing images. To design an effective classifier, the parameters of SVM model have to be configured properly in advance (Min & Lee, 2005). In this section, we will describe the SVM parameter optimization approach that based on Genetic algorithm.

### *1.1 Fitness function*

Fitness function is a kind of objective function that estimates the quality of each chromosome. In GA-based

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 5, May 2014

SVM parameter optimization process, the most difficult work is to design a fitness function to produce SVM parameters that are reliable and effective for SVM models. K-fold cross validation is a widely used technique to assess the generalization ability of a SVM classifier (Damian, 2002; S. Abe, 2005). Here, we apply k-fold cross-validation (CV) classification rate to the GA fitness function. The higher cross-validation classification rate represents the greater SVM classifier generalization ability on the given training data. In k-fold cross-validation, training data T are randomly divided equally into k subsets T1… Tk, a classifier is trained by k-1 subsets and tested using the remaining subset Ti (i=1,…, k).

Consequently, training is iterated k times, and the final classification rate is the average of all the k times' Classification rates. The k-fold cross-validation fitness F is obtained by:

$$F = 1/K \Sigma \text{ Ti rate} \quad (12)$$

where rate Ti is the classification rate of Ti using the remaining k-1 subsets as training set.[2]

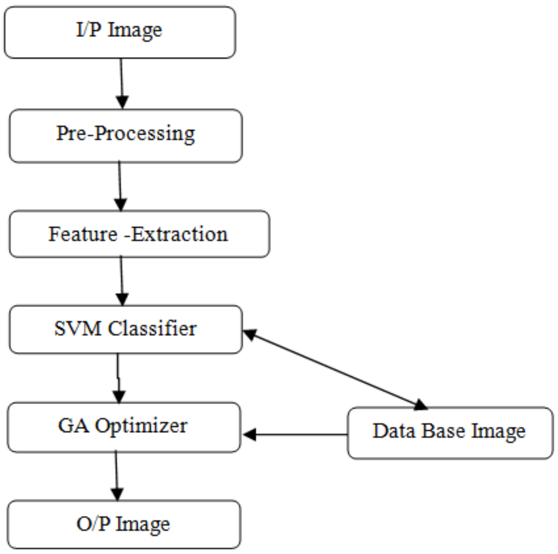



FIG 1.GA-SVM CLASSIFIER ENGINE

Experimental setup: Our idea in this project is to classify the given data set in to several categories such as Land, Water, Building, Vegetation, Road etc. A high resolution image is taken as input for GA-SVM Classifier Engine as shown in fig.1.

Pre-Processing: To improve the recognition rate of the image, a procedure of image pre-processing is performed such as 5*5 median filtering & Histogram equalization.[4],[8].

Feature-Extraction: The information about the image is extracted using features. Features are the attributes which are related to particular class. Pixel level features, Local level features, Global level features can be extracted.[4]

Support Vector Machine Classifier: The support vector machine (SVM) is superior of all machine learning algorithms. SVM employs optimization algorithms to locate the optimal boundaries between classes. The optimal boundaries should be generalized to unseen samples with least errors among all possible boundaries separating the classes, therefore minimizing the confusing between classes. Thus, by minimizing the confusion we will increase accuracy of classification.

Genetic Algorithms: The techniques of image classification ranging from maximum likelihood to neural networks depend on feature vectors formed by the intensity values in each spectral channel for each pixel. But the spectral information alone is not sufficient to exactly identify a pixel. The features of its neighborhood, like texture, or the average value of nearby pixels are necessary to get good spectral information. The different kinds of spatial content information could also be added into the pixel feature vector as additional feature dimensions.

So there are a large number of choices for additional feature vectors that could make classification better than just having the raw spectral values as feature vectors. Hence to choose these features automatically a new evolutionary hybrid genetic algorithm is used.

Data Base: It contains all class information, features as data base. Our input image's features are compared with this data base and then final labeling is done.
Finally we get the classified output image.

## IV. METHODOLOGY

The implementation of the GA-SVM parameter optimization process can be summarized as follows:

### *1.2 Data pre-process: Attributes scaling.*

To avoid attributes in greater numeric ranges dominating those in smaller ranges, generally, each feature should be linearly scaled to the range [0, 1] or [-1, 1].

### *1.3 Initializing population*

In this step, the parameters (i.e. phenotype), C and _, are transformed into chromosomes (i.e. genotype), which are represented by bit strings using binary coding scheme. And generate randomly a population of N chromosomes m C (m=1,..,N).

### *1.4 Decoding*

Convert genotype to phenotype: This is an inverse process of encoding, which converts the binary chromosomes into the numeric values.

### *1.5 K-fold cross-validation*

In this paper, k-fold crosses validation is used to estimate the generalization ability of SVM classifier on the given training data. The detailed description of cross-validation, please refer to above section.

### *1.6 Training SVM model*

SVM classification model is obtained by training, and the model parameters are set according to the GA decoding results.

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 2, Issue 5, May 2014

*1.7 Calculating classification rate*

The classification rate on testing data is calculated using the trained SVM model obtained by previous execution.

*1.8 Fitness evaluation*

The cross-validation accuracy is used as GA fitness function. The chromosomes with respect to greater fitness are easier to be selected to produce next generation.

*1.9 Termination*

When the terminating condition is satisfied, the evolution ends, otherwise the process executes next operations. In this research, the termination criterion is that the number of evolutionary generation reaches the maximum generation.

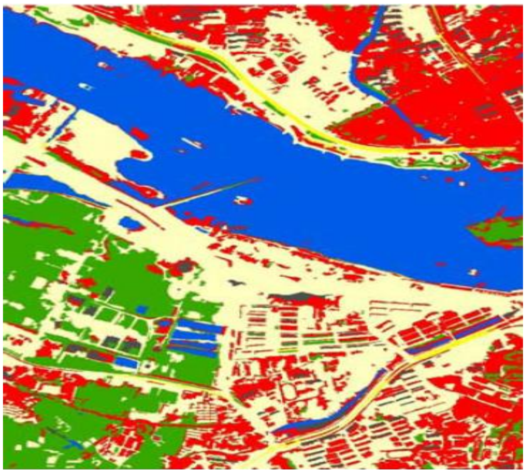
1.10 *Genetic operation*:

In one course of evolution, the new population is produced after a series of genetic operation, such as selection, mutation and crossover.

## V. RESULT

In order to evaluate the effectiveness of the proposed method for high resolution remote sensing images object based classification, an experiment is carried out on a subset image with 2859×3806 pixels chosen from Quickbird of Fuzhou, acquired in March 2002. To improve the recognition rate of the image, a procedure of image pre-processing is performed such as Pansharping fusion, 5×5 median filtering and histogram equalization. In the first step, the image objects are generated by means of FNEA segmentation, which implemented in the eCognition5.0 software tool (Definiens Imaging, 2006). In the literature, it is demonstrated that image characteristics such as texture, shape and pattern are helpful to the high resolution remote sensing images

Interpretation (Lillesand, 2004). Secondly, objects features, involve spectral, textural and shape, are calculated after segmentation. In this research, we extract the widely used objects features including:1) spectral features : layer means, brightness, standard deviation and maximum difference; 2)texture features based on GLCM: mean, standard deviation, contrast, dissimilarity, homogeneity, entropy and angular second moment; 3)shape features: shape index and length/width; 4)NDVI and NDWI (Frieke,2007; Wu, 2009).



For the segmented 300 image objects, we select 100 objects as samples, which randomly partitioned into two parts: 70 objects are used for SVM training while the remaining served as validation data, to assess the performance of our proposed GA-based approach. In this research, we classified the objects into 6 categories: bare land, building, road, shade, vegetation and water.

The experiments were conducted on the MATLAB7.7 development environment using the SVM toolbox of LIBSVM 3.0(Chang & Lin, 2001).

## VI. CONCLUSION

Image features are important to be considered since the images are retrieved based on these features. The extraction of suitable features from the image is the basic step by which the query image and the database images can be compared. The commonest features in an image are color, shape and texture. Generally, in the classification of object-based high resolution remote sensing images, numerous object features, such as spectral, texture, shape and contextual, are calculated after segmentation. However, the determination of the most appropriate feature subsets not only degrades the computational complexity, also can obtain a higher classification rate. By using SVM classifier we can classify the objects more accurately when we use a Non-parametric classifiers.

A feature extraction from invisible bands for hyper-spectral image analysis, where there is no input to the human perception system at all. The perception-centered approach has the obvious advantages in serving the "master" or "user" of the image retrieval system human, given that matching human performance is the ultimate goal of the system. We conclude that, whenever there is confusion occurs near the boundary, genetic parameter for image classification is best.

## ACKNOWLEDGMENTS

I take this opportunity to express my profound gratitude and deep regards to my guide Professor Chandrashekhr Patil for his exemplary guidance, monitoring and constant encouragement throughout the course of this paper.

This paper would not have been possible without the guidance and the help of him.

I would also like to convey my sincere regards to people who knowingly or unknowingly help me during my project. Last but not the least, I would like to express my love and gratitude to my beloved family for its understanding & endless love, through the duration of my studies.

## REFERENCES

[1] E. Saravana Kumar, A. Sumathi, K. Latha “Feature Selection and Extraction for Content-Based Image Retrieval” International Journal of Mathematics Trends and Technology- Volume3 Issue2- 2012, ISSN: 2231-5373, 70-73.

[2] Mengmeng Li1, Xiaocheng Zhou, Xiaoqin Wang, Bo Wu “Genetic Algorithm Optimized SVM in Object-based Classification of Quickbird Imagery” Spatial Information Research Center of Fujian Province, Fuzhou University Fuzhou 350002, China, 2011 IEEE**,** 348-352.

[3] DevisTuia, GustavoCamps-Valls, GionaMatasci, andMikhailKanevski, “Learning Relevant Image Features with Multiple-Kernel Classification”IEEE VOL. 48, NO. 10, OCTOBER 2010, IEEE, 3780-3791.

[4] Xiang Sean Zhou, Ira Cohen, Qi Tian, Thomas S. Huang Beckman Institute for Advanced Science and Technology” Feature Extraction and Selection for Image Retrieval” University of Illinois at Urbana ChampaignUrbana, IL 61801 VOL. 12, NO. 6, DECEMBER 2008, 765-779.

[5] Luca Bertelliy Tianli Yuy Diem Vu Burak Gokturk,” Kernelized Structural SVM Learning for Supervised Object Segmentation”, flbertelli, tianliyu, diemvu, burakgg@google.com, China, 2007, IEEE 2153-2160.

[6] D. LU, Q. WENG,“ A survey of image classification methods and techniques for improving classification performance”, International Journal of Remote Sensing,Vol. 28, No. 5, 10 March 2007, 823–870.

[7] Genetic Algorithms: A Survey M. Srinivas, Motorola India Electronics Ltd. Lalit M. Patnaik, Indiani nstitute of Science,June 1994,IEEE 17-26.

[8] “Object-Oriented Feature Extraction Method for Image Data Compaction “Hassan Ghassemian and David A. Landgrebe, 1988, IEEE 42-48.