A Survey of Image Classification Techniques for Flood Monitoring System

Richa Dandotiya, Shailendra Singh, and Shailesh Jalori

Abstract—In this paper, we presented a survey of image classification techniques for monitoring the flood affected areas on land to support flood monitoring system connected with remote sensing network. Image classification is one of the most important processes in high-resolution remote sensing information extraction. After studying various existing image classification techniques, it was found that Support Vector Machine (SVM) is the one of the most suitable techniques for analyzing natural images where high dimension data-set is used as an input for classifying the images. Remote sensing image classification using Support Vector Machines (SVMs) follows proper method to select training dataset by identifying appropriate informational class. It selects & evaluates the signatures based on various statistics analysis like mean, variance, covariance, etc for generating training dataset depending on the classifier type to be used then evaluates the results for most suitable training data-set. After getting most suitable training dataset, SVM is applied to classify the image and reports are generated to assess for next training dataset.

Keywords—Support Vector Machines (SVMs), Flood Monitoring System, Training Data-set

I. INTRODUCTION

The flood monitoring system being used all over the world not only to monitor the flood parameter to handle current situation but also to predict the flood in future and warn the people before time. Remote sensors devices being used in satellites are playing important role to send the images of flooded areas continuously to analyze and to take decisions accordingly. The image classification is one of the most important processes in high-resolution remote sensing information extraction. Image classification systems are based on the one or more features extraction, and to classify the different images based on trained data-set. The image classification techniques become much more complex when the diversity and number of images increase significantly and determined parameters are not enough to classify the images. In this scenario image classification system lacks robustness and effectiveness. There are many techniques and methods proposed by various researchers which can be studied in related work following this section, SVM found to be the most suitable to classify complex images such as flood images which are dynamic in nature. Support Vector Machines (SVMs) were developed by Vapnik [1],[2] for giving empirical performance to solve image classification problems. The formulation of the Structural Risk Minimisation (SRM) principle as approached to the Empirical Risk Minimisation (ESM) approach commonly employed with statistical learning methods and has been successfully applied to a number of real world problems such as handwritten character recognition[3], text categorization [4], and image classification [5]. SVM is one of the several kernel based techniques available in the field of machine learning. In recent years, several searching algorithms have been developed for classifying the remote sensing image. The classifying method which has been proposed during and evaluation includes SVM, Fuzzy Logic [6],[7] Artificial Neural Network [8]. Two main classification methods are supervised classification and unsupervised Classification. Supervised classification is most popular in researches and applications.

Machine learning is important concept introduced to learn the image classification based on training. Supervised learning and unsupervised learning are also applied on image classification system. One of the main critical issues in the application of supervised classification methods to the analysis of remote sensing images is the definition of a proper training set for learning of the classification algorithm. This issue can be achieved by unsupervised classification but its accuracy would be compromised.

The purpose of this survey is to study the techniques for image classification using supervised classification techniques to facilitate the search and classification of images. This paper presents various novel approaches, clustering approaches, Particle Swarm Optimization (PSO) based approaches and SVM approaches. With regard to the kernel-induced methods, such as the SVMs, recently, have been applied to remote-sensing classification. The SVMs are large-margin classifiers that define an optimal separating hyper-plane in a properly chosen kernel-induced feature space and is implemented by a convex optimization with radial basis kernel classifiers.

To address the problems of remote sensing images,
different classification approaches have been proposed in the literature. Each of them are based on one of the following four parameters:

a) Data pre-processing feature selection/extraction techniques;

b) Data analysis via our modified SVM classifier [9],[10],[11],[12],[13],[14]

c) Testing and comparison with real-world products;

d) Classification system.

This paper, presents basic concepts of completely automatic classification system that optimizes classifier accuracy in remote sensing image using SVM with Gaussian Radial Basis Kernel function. The rest of the paper is organized as follows: Section II introduces the problem background and brief of some related work that had been done in this area. Section III introduced the image classification system. The SVM technique for image classification approach described in section IV. Section V draws the conclusion of this paper and discusses the future works.

II. RELATED WORK

There are many techniques available for image classification specially for flooded areas but most of them has some drawbacks. The major issues of such systems are time consumption for a large scale of data, inaccuracy, lack of trained data set, feature extraction etc. These approaches can be classified into different categories based on the techniques used, some of them are Object-Oriented Approach, GIS, Systematic Approach, Neural Network.

Multi-temporal techniques based on detection of changes can be used as a powerful tool to detect flooded area in SAR imagery. It is performed by acquiring two imageries taken before and after the flood. The object-oriented approach is preferred as an image processing method as it combines the spectral information with spatial information as shape, size, texture and neighbourhood relations to increase the classification abilities. GIS approach can be used to create, manipulate, analyze and display all types of geographically or spatially referenced data. Zhou et al. proposed a Neural Network approach to detect flooded areas with three layer structured RBFNN in which the output units from a linear combination of the kernel function (Benediktsson et al., 1990). Systematic approach to quantitatively monitor of flood monitoring using AVHRR data in flood. It uses AVHRR imagery in flood monitoring.

SAR (Synthetic Aperture Radar) data are becoming a valuable tool to analyze water related investigation (Lee & Lee, 2003).

Jian-Gang Yang et al.[16] proposed multi-class SVM based remote sensing image classification and its semi supervised improvement scheme.

Particle Swarm Optimization[17] is the technique firstly described by James Kennedy and Russell C. Eberhart in 1995. It consists of a swarm of particles. Each particle resides at a position in the search space and is based on a particle swarm optimization and SVM techniques using Corel Image Database. They found SVM, Back Propagation Neural Network and Radial Basis Function (RBF) are used to compare with PSO-SVM. This algorithm consists of these steps, which are repeated until some stopping condition. First step, evaluates the fitness of each particle according to the desired optimization, second step updates local best fitness and global best fitness and third steps updates the velocity and position of the particles.


Methods named Bagging [20] and Boosting [21] are generally used to improve classification performance in a non-parametric classification. These techniques have been used in decision tree and SVM. Experiments use two datasets namely the USPS dataset and UMIST face dataset. The experiments results show that SVM approach compared to the traditional classifying method and which achieves better accuracy. Liu Yongxue[22] gave Adaptive Boost Algorithm and Adaptive Boost-SVM best applied to classify IKONOS images(Spectral and Texture) using ensemble algorithm. Masahiro Tada and Toshikazu Kato[23] proposed analysis method of the relation between the content and the impression words and applying them to automatic image classification system based on visual impression.

Francesca Bovolo[24] used a novel context-sensitive classification technique based on SVM using remote sensing images. It has two main properties used in novel approach- i)SVM kernel by defining a context sensitive cost function; ii) the spatial-context information in the classification phase. The results provided by the Context Sensitive SVM were compared with those achieve by a standard context-insensitive SVM and found later the better technique.

Sourabh Agarwal[25] proposed a novel approach for content based color image classification using SVM with the implementation of the code k-fold cross-validation algorithm. These experimental results were based on the terms of color image classification accuracy. Wei Zheng[26] described the application of AMSR-E image classification based on support vector machine in order to monitoring flood and water logging. The experimental result indicates that classification image based on SVM over Huaihe River Basin gave accuracy 97% more than neural network method on July 6, 2003.

ENVI software is used for the classification research. Fen Chen, Zhiru Zhang[27] proposed a classification method for
remote sensing images with spectral and texture feature based on Support Vector Machine. The image is first segmented with the spectral features. Then texture features are extracted by wavelet transform. Third, the SVM is used to classify the image. The implementation of the SVM is used in LIBSVM[28]. Martha R. Quispe-Ayala[29] proposed a parameter free image classification method based on data compression techniques and calculate a measure of similarity between image based on Kolmogorov complexity. Data compression is based on data sets in repetitions that stores the data.

Yan Li, Li YAN, Jin LIU [30] explained the changes in remote sensing classification from two aspects: basic thought and new classification algorithms. The basic thought of remote sensing classification has changed from per-pixel multispectral-based approaches to multi scale object-based approaches. New categorization algorithms contain Support Vector Machine, Evolutionary Algorithm, Fuzzy Clustering algorithm, as well as Artificial Neural Networks.

III. DEVELOPMENT OF IMAGE CLASSIFICATION SYSTEM

Image classification is one of the important and complex processes in image processing. There are several image classification methods. Basically, classification systems are either supervised or unsupervised, depending on the finite number of discrete supervised classes and unsupervised categories. Most of classifiers, such as maximum likelihood, minimum distance, neural network, decision tree, and support vector machine, are making a definitive decision about the land cover class and require a training sample. On the contrary, clustering based algorithm, e.g. K-Means, K-NN or ISODATA, are unsupervised classifier, and fuzzy-set classifier are soft classification providing more information and potentially a more accurate result. This study uses supervised classification technique using SVM to classify images because of high accuracy to classify remote sensing data in huge size and dynamic in nature, which was a statistic classification method proposed by Cortes and Vapnik in 1995. It was originally designed for binary classification but because of structural risk minimization from computational learning theory, it tries to find the separating hyper-plane with maximum margin to separate positive and negative samples from the training set.

SVM have been many times found to be the best classification technique for higher accuracies than other widely used pattern recognition techniques, such as the maximum likelihood and the multilayer perceptron neural network classifiers.

Classification accuracy is one of major parameters for classifying images for better results. SVM-based multiple classifiers use fusion method for remote sensing image classification. Wavelet texture feature and spectral feature to can be used to construct SVM classifier separately. The most reliable classification result is thus the one that gives the largest distance in hyper plane in SVM.

Texture analysis is mostly used to classify remote sensing images using efficient filters in extracting textured image. SVM classifier separates the classes with a decision surface called optimal hyper-plane that maximizes the margin between the classes and the data points closest to the hyper plane are called support vectors. SVM have much potential for the supervised classification of remotely sensed images. Comparative studies shows that the classification by a SVM can be more accurate than popular contemporary techniques. The usefulness of SVM shows that it requires small set of training data set. Figure 1 shows the common classification process after studying many SVM classifiers for image classification applied to the remote sensing dataset for flood monitoring.

In general, image classification techniques include input data called remote sensing image's dataset. This input dataset is continuous data stream to the flood monitoring system. Second steps require feature extraction or feature selection method to define the proper input to the image classifier. Third steps is the classification process which involves trained dataset to classify the image under supervised classification. Every classification process helps the whole system to learn new thing from supervised learning mechanism. Finally, Graphical User Interface creates the reports in required format for flood monitoring and control.

Supervised Classification

The supervised classification is the essential tool use for extracting quantitative information from remotely sensed image data[31]. Commonly used supervised approaches are - parallel piped classification, minimum distance classification, maximum likelihood classification etc. For this reason, sampling of training data from clearly identified training areas, corresponding to define classes is made for the population statistics. This is called supervised classification.

The steps in supervised classification approach are:

1. Proper method to select training dataset by identified informational class
2. Select & evaluate the signatures based on various statistics analysis like mean, variance, covariance, etc for generating training dataset depending on the classifier type to be used.
3. Evaluate the results for most suitable training data
4. Apply the supervised classification technique on remote sensing image
5. Report the image found after classification
6. An interesting accuracy assessment is an indispensable step in remote sensing imagery classification process, which does not depend on developing a testing set of pixels.

IV. SUPPORT VECTOR MACHINE

1. Support Vector Machine(SVM)

Support Vector Machines (SVMs) [32] were first introduced in the mid of 1990s and have since been established as one of standard tools for machine learning and data mining. This study uses the SVM to classify images, which was a statistical classification system proposed by Cortes and Vapnik in 1995 [1],[2]. The simplest SVM was a binary classifier, which was mapping to a class and just can identify an input image data belongs to the class or not.

Applications of a wide range of pattern recognition problems, image classification, financial time series prediction, face detection, biomedical signal analysis, medical diagnostics, and data mining uses SVM now a days. Standard SVM classifies objects into two classes by calculating the maximum margin hyper-plane between the training objects of both given classes. Structural risk minimization is associated with such a scheme to show a good trade off between low empirical risk and small capacity.

The use of soft margins by specifying a cost factor c and kernel functions (e.g. Radial Basis Kernel or Polynomial) which enables to classify any kind of data.

The SVM algorithm provides a choice of four kernel types: (i) Linear, (ii) Polynomial, (iii) Radial Basis Function, and (iv) Sigmoid.

A. Optimal Separating Hyper planes

Let training set \((x_i, y_i)\) be separated by a hyper plane with margin \(\rho\).

\[
W^T x_i + b \leq -\rho/2 \quad \text{if} \quad y_i = -1
\]

\[
W^T x_i + b \geq \rho/2 \quad \text{if} \quad y_i = 1
\]

\[
y_i(W^T x_i + b) \geq \rho/2
\]

Then,

We can formulate the quadratic optimization problem.

Maximum

\[
\rho = \frac{2}{\|w\|}
\]

Subject to

\[
y_i(W^T x_i + b) \geq 1
\]

Quadratic optimization problems are well known class of mathematical programming problem for which several algorithms exist in order to solve the optimization problem, given a solution \(\alpha_i = \cdots \alpha_n\).

For the dual problem, solution to the optimal is

\[
w = \sum\alpha_i y_i x_i \\
\]

\[
b = y_k - \sum\alpha_i y_i x_i x_k
\]

Finally the classification function is written as follows:

\[
f(x) = \sum\alpha_i y_i x_i^T x + b
\]

where \(w\) is a vector coefficient, \(b\) is constant, \(i\) is index label, \(x_i\) is the independent variables.

B. Linearly non-separable case

\[
y_i(W^T x_i + b) \geq 1 - \xi
\]

The generalized Optimal Separating Hyper-plane is then regarded as the solution of the SVM type-1 classification (also known as C-SVM classification).

\[
\frac{1}{2}w \cdot w + c \sum_{i=1}^{N} \xi_i
\]

Subject to constraint

\[
y_i(W \cdot x_i + b) \geq 1 - \xi, \xi_i > 0
\]

Where \(c\) is a capacity constant, \(w\) is a vector coefficient, \(b\) is a constant, \(\xi_i\) is a parameter for handling non separable data inputs, \(i\) is index label, \(N\) is training case, \(x_i\) is independent variables.

C. Non-Linear SVM

A kernel functions is a function that is equivalent to an inner product in some feature space [12].

\[
f(x) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i k(x_i, x) + b\right)
\]

Here, kernel \(k(x_i, x_j)\) is the kernel function. The most widely used kernel function is the Gaussian function \(e^{-\frac{|x_i - x|}{\sigma^2}}\), where \(\sigma^2\) is the width of the Gaussian kernel, \(\alpha_i\) Lagrange multiplier to be optimized.
A large number of software implementations of SVM with distinctive features have been developed by researchers in the machine learning area.

2. K-Means Clustering

Clustering has been applied in a wide variety of fields, ranging from engineering (machine learning, artificial intelligence, pattern recognition, mechanical engineering, electrical engineering), computer sciences (web mining, spatial database analysis, textual document collection, image segmentation), life and medical sciences (genetics, biology, microbiology), to earth sciences (geography, geology, remote sensing). The K-means algorithm is the best-known squared error-based clustering algorithm. It is very simple and can be easily implemented for solving many practical problems. It can work very well for compact and hyper spherical clusters. The time complexity of K-Means is \(O(NKd)\). An interesting technique, called ISODATA developed by Ball and Hall, deals with the estimation of \(K\).

V. CONCLUSION

This paper attempts to study and provides a brief knowledge about the different image classification approaches used to classify the remote sensing image for flood monitoring. Most of classifiers, such as maximum likelihood, minimum distance, parallelepiped algorithm, neural network, decision tree, and support vector machine, are making a definitive decision about the land cover class and require a training sample. This study uses supervised classification technique using SVM to classify images because of high accuracy to classify remote sensing data in huge size and dynamic in nature. This paper also explains the common classification process used in classification of flood images with steps required in supervised classification. This survey gives theoretical knowledge about different classification methods and describes the main features of SVM techniques for supervised classification of the remote sensing images used in flood monitoring system.

REFERENCES

[27] FenChen,Zhiru Zhang.”Image classification with spectral and texture features based on SVM.