International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 23 (2015) pp 43360-43367

Classification of SAR Images Using Support Vector Machine Learning Algorithm with Different Kernel Functions

P. Deepthi Jordhana ·
Associate Professor, Dept of ECE, Intell Engineering College, Anantapur, India. Email: p. deepthijordhana@gmail. com

K. Soundararajan
Retd Professor, Dept of ECE, JNTU College of Engineering, Anantapur, India Email: soundararajan_jntucea@yahoo. com

Abstract
This work incorporates the classification of objects as Manmade or Natural using Kernel functions based Support vector machines (SVM) on SAR Images. Support vector machine uses training images to train a classifier and this classifier is tested on the test images. Training Images are classified into positive for natural images and negative for manmade objects. Test Images are the SAR regions under consideration. The non linear image data is mapped into piecewise linear data using a kernel function. SVM Method with different kernel functions is used to categorize SAR images. The performance of each kernel function is measured by using parameters like False Alarm (FA) and Target miss (TM). They indicate the percentage of incorrectly detected objects. Simulation results based on Matlab proves the efficiency of different kernel functions for different sets of SAR Images.

Key words Support vector machines, RBF Kernel, Sigmoid Kernel, Polynomial kernel, Hellinger kernel.

I. Introduction

Synthetic aperture radars (SAR) are becoming increasingly popular in remote sensing applications. The significant properties of SAR are

- Sensitivity of back scattering coefficient to target geometry.
- Measures are independent of atmospheric conditions and solar illumination.

In this paper, we focus on the classification of manmade and natural objects based on their geometrical properties. The technique applied here is Support vector machine (SVM).
Automatic target recognition (ATR) systems incorporate SAR sensors for a wide variety of applications in the Military and defense sector. The goal of these machines is to detect and classify military targets using various signal processing techniques. In surveillance and military applications of SAR, Buildings and vehicles are the most important man-made structures, which need to be detected. Some of the previous work on detection of buildings is given at [6][7][8][9][10] on normal images. Large numbers of these techniques use aerial images for building detection by generating a hypothesis on the presence of surface on building roof top image [6].
The first step is detecting low-level image characteristics such as edges and regions. Image segmentation is a primary step in doing so. One of the methods is using Graph cut Image segmentation described in [1][2] of any SAR related image and in [3] which uses the principle of convergence of energy to a local minimum [4] to classify regions [5]. In next step either geometric feature based hypothesis [7] or a statistical models such as Markov Random Field (MRF) [8] is applied. In [11] a technique was proposed to use graph spectral partitioning for detection.
The work at [12][13] establishes method to classify the whole image as a landscape or an urban scene. Oliva and Torralba [12] obtain a low dimensional holistic representation of the scene using principal components of the power spectra. The power spectra based features to be noisy for SAR images, which contain a mixture of both the landscape and man-made regions within the same image.
One popular approach towards target classification is to a SAR’s target image’s amplitude values to generate image features. These image features can be used with common classifiers like nearest neighbor or neural networks. Principle component analysis (PCA) and independent component analysis (ICA) are popular features extraction approaches for target classification [17][18][19].
The work at [13] uses the edge coherence histograms over the whole image for the scene classification, using edge pixels at different orientations. Olmos and Trucco [14] proposed a system to detect the presence of man-made objects in underwater images using properties of the contours. The techniques discussed in [15][16] perform classification in outdoor images using color and texture features, with different classification schemes. These papers report poor performance on the classes containing man-made structures since color and texture features are not very informative for these classes [13]. However for SAR images these techniques cannot be applied. These techniques classify the whole image in a certain class assuming the image to be mainly containing either man-made or natural objects, which is not true for many real-world images. In case of SAR created images, the images are taken over wide area containing mixed real world and man-made objects.
In this paper, we propose to detect man-made structures in 2D images, formed by SAR. The proposed method uses Support vector machines with different kernel functions for classification. The section 2 illustrates introduction to Support vector machine. The section 3 explains the various kernel mapping functions and SVM algorithm simulated on the SAR images. The section 4 has algorithm, simulation results and applications.

II. Support Vector Machines

Synthetic Aperture Radars (SAR) is often employed in

Remote sensing areas because they offer better performances than Optical or Infrared applications. A SAR Image is shown in Figureure. 1. Support vector machines are applied to the problem of classification of remote sensing images where it overcomes the limitations of Fuzzy theory and neural networks.

The support vector machine is actually a training algorithm for learning and classification. It can be used to learn Multi Layer Perceptron (MLP) algorithms using kernel functions like Polynomial, Radial basis functions (RBF). The underlying concept of SVM is Statistical Learning theory shown in Figure. 2 whose goal is to model a hypothesis space which is compared with the target space. Approximation and Estimation error are found and the objective is to find the function that minimizes the risk.

The two most important elements to consider while implementing SVM are mathematical programming and kernel functions. These parameters are found by solving a quadratic problem. Generally SVM's are used for a two-class classification with classes being positive P and negative N with labels +1 and-1. However it can also be extended to *k* class classification. SVM therefore searches for a hyperplane with good statistical properties equidistant from the two classes.

Figure. 1 An SAR created image

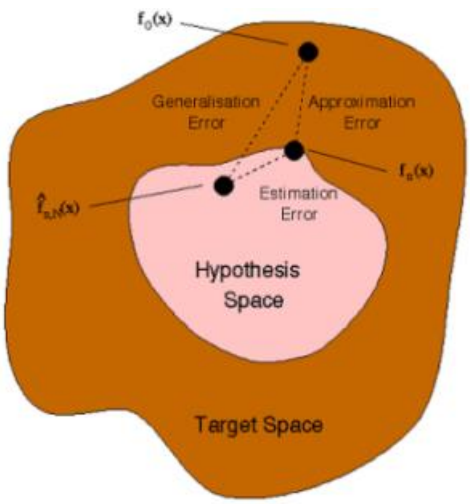


Figure. 2 Statistical Learning Theory

Maximal Margin Hyperplanes (Linearly spaced data)

If the training data are linearly separable then there exists a pair (w, b) such that

$$w^T x_i + b \geq 1, \forall x_i \in P \quad (1)$$

$$w^T x_i + b \leq 1 \ \forall x_i \in N(2)$$

With the decision rule given by

$$f_{w,b}(x) = sign(w^T x + b) \quad (3)$$

w is weight vector and b is the bias. The inequality constraints (1) and (2) can be combined to give

$$y_i(w^T x_i + b) \geq 1 \ \forall x_i \in P \cup N \quad (4)$$

The pair(w, b) can be rescaled such that

$$\min_{i=1,\ldots,l} | w^T x_i + b | = 1 \quad (5)$$

SVM always searches for the simplest solution to classify data correctly. The learning problem can thus be reformulated as: minimize $||w||^2 = w^T w$. This is equivalent to maximizing the margin i. e., distances between the 2 planes. The optimization is now a quadratic problem

$$\text{Minimize } \emptyset(w) = \frac{1}{2} ||w||^2$$

$$y_i(w^T x_i + b) \geq 1, i = 1, \ldots . l \quad (6)$$

This problem has a global optimum; thus the problem of many local optima is avoided.

Kernel induced feature spaces

A linear classifier may not be always suitable for being the most suitable hypothesis for the two classes. SVM can be used to learning non linear decision functions by first mapping the data into higher dimensional space and constructing a separate hyper plane there shown in Figure. 3.

Mapping to feature space

$$X \rightarrow H$$
$$x \rightarrow \emptyset(x)$$

The decision function is given by

$$f(x) = sign(\emptyset(x)^T w^* + b^*) \quad (7)$$

$$sign(\sum_{i=1}^{l} y_i \lambda_i * \emptyset(x)^T \emptyset(x_i) + b^*)$$

The kernel function is given by

$$K(x, z) \equiv \emptyset(x)^T \emptyset(z) \quad (8)$$

The kernel function allows us to construct a hyperplane in the feature space H without actually having to perform any calculations in the space. Instead of calculating the inner products, the value K is directly computed. The Polynomial kernel

K(x, z) = $(x^T z + 1)^d$ which corresponds to a map $\emptyset$ into the space spanned by products of up to d dimensions of $\mathcal{R}^N$. The decision function then becomes

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l} y y_i \lambda_i * K(x, x_i) + b^*\right) \quad (9)$$

where the bias is given by

$$b_i = y_i - \sum_{j=1}^{l} y_j \lambda_i * K(x_j, x_i) \quad (10)$$

for any support vector x_i.
The kernel function should therefore be easy to compute, well-defined and span a rich hypothesis space. Most commonly used kernels in this aspect are Gaussian RBF, two layer MLP or polynomial classifier. Mercer's theorem which states that any positive kernel is an inner product in feature space makes this possible.
The 'kernel trick' gives the SVM great flexibility to perform. With suitable parameters, it can separate any consistent data. Usually this flexibility causes a learner to overfit the data. Overfitting is usually the main problems of data mining in general. SVM overcomes this issue by using regularization where the data is separated by a large margin. The space of classifiers that separate the data over a large margin has much lower capacity.

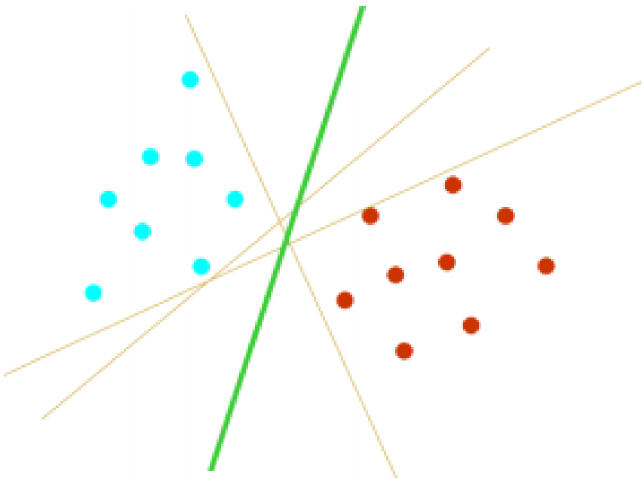


Figure. 3 Optimal Separating Hyperplane

III. SVM Algorithm and kernel functions

This section elaborates on the algorithm of Support vector machines that is applied on the images under consideration to classify whether or not it is natural or manmade.

Data Preparation

The data to be prepared is first separated in two distinct categories as Training Images and test images.
Training Images are categorized into positive images and negative Images. Positive Images are labeled as +1 and negative as-1.
The feature vectors are found for Training Images first. The feature vectors consist of SIFT features computed on a regular grid across the image. The frequency of each visual word is then recorded in a histogram for each tile of a spatial tiling as shown in Figure. 4. The final feature vector for the image is a concatenation of these histograms.

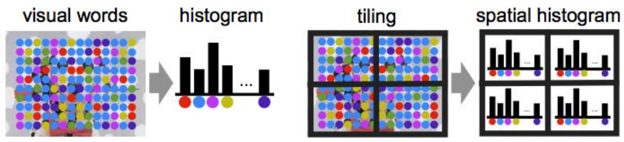


Figure. 4 Spatial tiling

Training a classifier

The natural images are all used as positives and the manmade objects are used as negatives. The classifier is a Linear Support vector machine (LSVM). The classifier can be tested by ranking all the training images qualitatively.

Classify the test images

The learnt classifier can now be applied on the test images. The retrieval performance can be quantitatively measured by computing a precision and recall curve.
Precision: It is the proportion of returned images that are positive.
Recall: It is the proportion of returned images that are returned.
The Precision-Recall curve shown in Figure. 5 is computed by varying the threshold on the classifier (from high to low) and plotting the values of precision against recall for each threshold value. In order to assess the retrieval performance by a single number (rather than a curve), the Average Precision (AP, the area under the curve) is often computed.

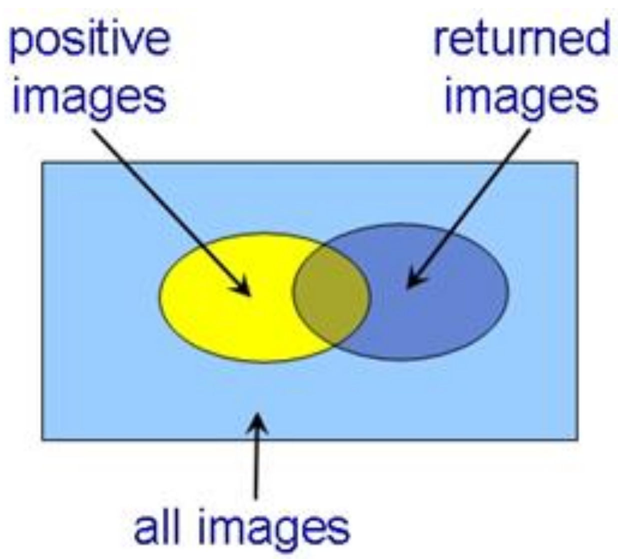


Figure. 5 Precision Recall curve

Kernel functions in Machine Learning Approach

The basic idea of kernel methods is to (φ) transform the input data points (black dots) in to a high-dimensional feature space, where they can be described by a linear model (straight solid line). The linear model found in feature space corresponds to a non-linear model in the input space (curved solid line). The

Figure. 6 has the basic illustration of mapping involved in kernel functions.

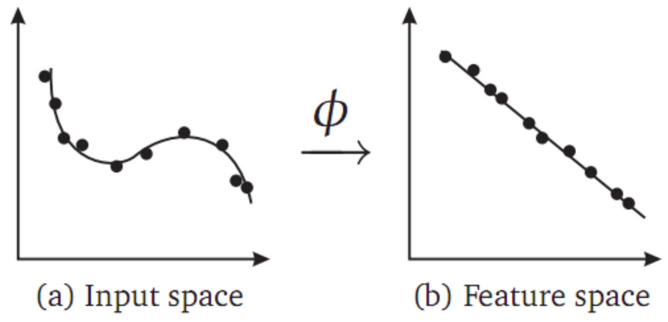


Figure. 6 Input space to linear feature space mapped by kernel function

This section describes the low level image parameters and mapped kernel functions which are considered for the ML algorithm. The edges, regions, statistical parameters are considered as input data set X. The output correct decision which is marked by operator is labeled as Y. Hence the training data set can be described as given in (10).

$$(X_1, y_1), (X_2, y_2), \cdots, (X_N, y_N) \tag{11}$$

The kernel functions considered here are first order and second order partial derivatives in the x and y directions. If P being the pixel set for an edge the first order derivative can map the linear edges to constants. Similarly the second order derivative can map the circular manmade objects to constants. Let $\Phi(x)$ be the mapping from the input data space to a higher dimensional feature/mapped space. The first step in algorithm divides the image domain into multiple regions and edges (X). Each region is characterized by one identifier, Further the region characteristics are mapped to kernel function which becomes input to optimization algorithm.

$$X = (x_0, x_1, \cdots, x_d) \overset{\varphi}{\to} Z = (z_0, z_1, \cdots\cdots\cdots, z_{\tilde{d}}) \tag{12}$$

$$X_1, X_2, \cdots, X_n \overset{\varphi}{\to} Z_1, Z_2, \cdots, Z_n \tag{13}$$

The perceptron based classification problem for binary decision uses weight vectors given in (13) defined on feature space.

$$\widehat{W} = (w_0, w_1, \cdots\cdots\cdots, w_{\tilde{d}}) \tag{14}$$

$$g(x) = \text{sign}\,(\widehat{W}^T \varphi(X)) \tag{15}$$

Based on the sign the correctness of perceptron is used to update the weight vectors iteratively. The final selected hypothesis 'g' becomes the candidate for classifying on the unknown data set.

$$H = \{h\}\ g \in H \tag{16}$$

The kernel functions used for this purpose are Radial Basis Function (RBF), Polynomial, Sigmoid and Hellinger.

Radial Basis Function

A common form of Gaussian kernel is the Radial basis function (RBF) kernel. It is called radial because the weights depend upon how far a point is to y. Its performance is measured by using the width parameter σ. The smaller it is, more is the emphasis laid on nearer points. When. $\sigma \to 0$, kernel behaves like a delta function, selecting the nearest point. Therefore this parameter must be finely tuned depending upon the problem at hand. If overestimated, the exponential will behave almost linearly and the higher-dimensional projection will start to lose its non-linear power. In the other hand, if underestimated, the function will lack regularization and the decision boundary will be highly sensitive to noise in training data. The equation for RBF is given by

$$KF(x, y) = \exp\left(-||x - y||^2/\sigma^2\right) \tag{17}$$

$||x - y||^2$ is taken to be squared Euclidean distance between the two feature vectors and $\sigma > 0$ is called the width parameter is used for fine tuning. The following points are the features of RBF kernel

- The value of kernel decreases with distance and ranges between zero and one, it is referred to as Similarity measure.
- The RBF Kernel specifies an infinite dimension feature space where higher order dimensions decay faster than lower order dimensions.
- It is comparatively faster.
- It provides smoothness over the contours.

Polynomial Function

It is a non stationary kernel where all the input data is normalized. Its equation is given below

$$KF(x, y) = (x.y + c)^d \tag{18}$$

Sigmoid Kernel Function

The Sigmoid kernel, also known as Hyperbolic tangent kernel or Multilayer Perceptron (MLP) kernel. This was originally developed for neural networks field, bipolar sigmoid function is used as an activation function for artificial neurons. Its equation is given by

$$KF(x, y) = \tanh(c(x.\,y) + \theta) \tag{19}$$

An important feature of a sigmoid kernel is that a SVM model using a sigmoid kernel function is equivalent to a two layer Perceptron network making it quite popular in SVM due to its origin from neural networks field. Despite being conditionally positive, it has been found to perform well in practice also. There are two adjustable parameters which are slope value c and intercept constant θ.

Hellinger Kernel

This kernel is especially suitable for Image processing applications where data is nonlinearly separable.

$$KF(x, y) = sqrt(x. y) \tag{20}$$

IV. Algorithm, Results And Applications

Machine learning algorithms form a branch of Artificial Intelligence (AI). AI is that part of the Computer science that solves problems concerned to both humans and machines intelligence with the help of different fields like Psychology, statistics and Biology. It includes various kinds of algorithms such as genetic algorithms, language processing algorithms, algorithms for planning and searching and lastly machine learning algorithms. Machine learning is further subdivided into three categories like

- Unsupervised Learning: The computer system finds patterns or solutions without the intervention of anybody. Applications include Principal Component Analysis and clustering of data.
- Reinforcement Learning: This type of learning is a trial and error method where every solution that is generated is given a score. The goal is to maximize the score. The score reflects how close the solution achieves the goal of the problem. Thus the solution with the maximum score is considered to be the best solution.
- Supervised Learning: In this case the system learns to form a solution using the various teaching methods offered. The ultimate goal of this method is to return the correct solution for new data.

This section presents the results and discusses the possible applications of the developed Support vector machine algorithm.

Algorithm

- In the first step, regions are detected using Graph cut Image Segmentation method and pixel groups are made corresponding to each region.
- In the second step the pixel coordinates are substituted in kernel mapping functions and feature space values are computed.
- To obtain region based classification, K means clustering algorithm is applied on the SAR Image to obtain clusters of the total image.
- Next on each cluster, Graph cut Image Segmentation algorithm is applied with a kernel mapping function to estimate the local minima convergence points.
- Contours are estimated over these points to finally obtain regions.
- All the regions are saved as Test Images in a separate folder.
- The Training Images are saved in one folder comprising of Natural as well as Manmade Images.
- Natural Images are considered to be positive Images with a label +1.
- Manmade Images are considered to be negative Images with label-1.
- The classifier is trained using any of the four kernel functions chosen by the user on each SAR Image using features extracted from the training Images.
- The user is allowed to browse through all SAR Images one at a time.
- The classifier is now tested using features from the test Images.
- If the score is +1 for a test image it is considered to be natural else it is manmade

The Algorithm flow of SVM based object detection is shown in Figure. 7.

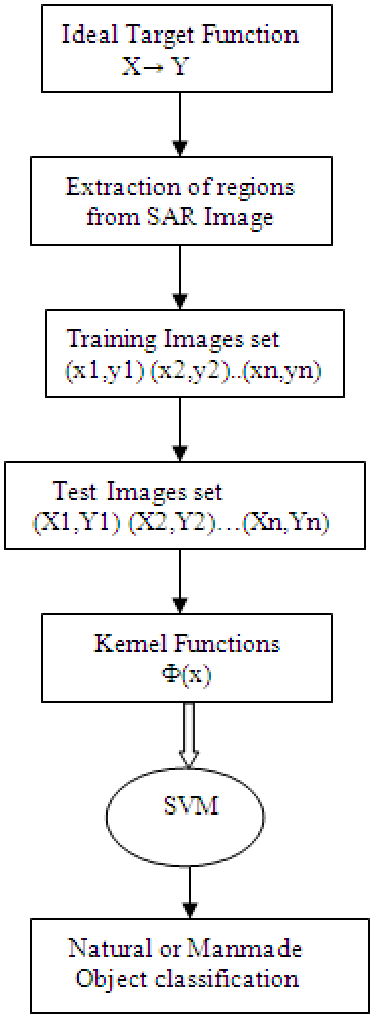
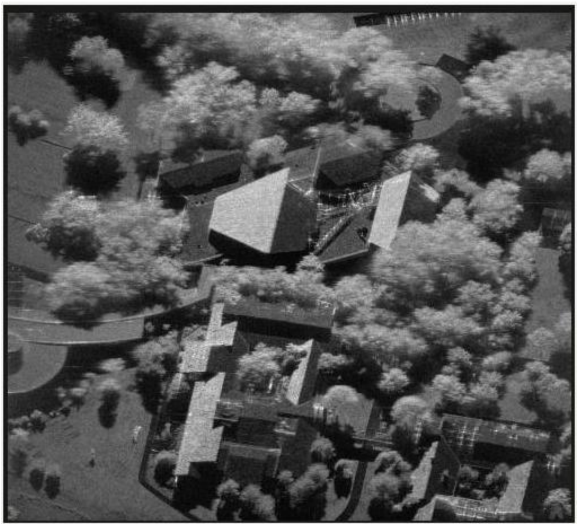


Figure 7 Algorithm flow of Object Classification

Simulation results

The SVM algorithm with Graph cut Image Segmentation is implemented in MATLAB and simulation results of the same are discussed in this section. The Figure. 8 have the region detection output for a specific SAR image.

The Natural Images consist of trees and terrains while the Manmade Images comprises of buildings as shown in Figure. 9. The training images are trained with any of the four kernel functions i. e., RBF, Polynomial, Sigmoid and Hellinger opted by the user. This classifier is tested on the test images that were saved in. jpg format as regions.



(a)

(b)

Figure. 8 (a) Input space (b) Processed image with regions

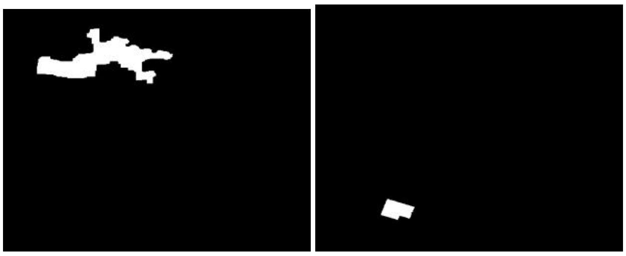


Figure. 9 (a) Natural image (b) Manmade object

Table 1 Experimental procedure parameters

Algorithm declared result	**Actual scene on SAR image**	
	Manmade object	***Natural object***
Man made object	CORRECT DECISION	FALSE ALARM
Natural object	TARGET MISS	CORRECT DECISION

MATLAB based graphic user interface (GUI) is developed to allow interactive user training for this purpose. The Figure. 10, 11, 12, 13 have the GUI's screen shot using RBF, polynomial, Sigmoid and Hellinger kernel functions. The images used in this simulation are system generated binary SAR images based on the reflections from the targets. Each iteration highlights the region of the image and the decision made by SVM algorithm. Operator declares the decision to be correct or incorrect. The statistics are displayed on the right.

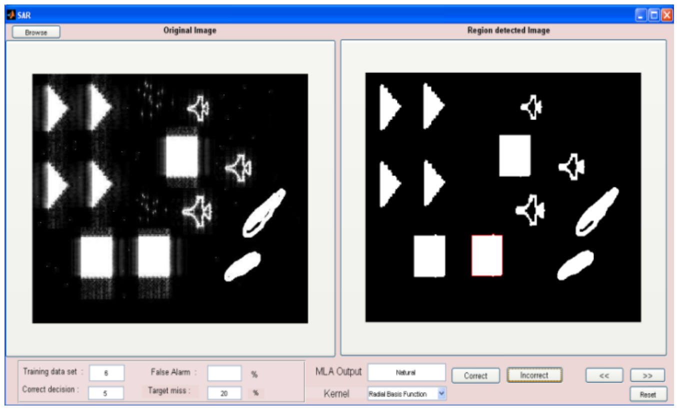


Figure. 10 GUI for interactive SVM Algorithm using RBF

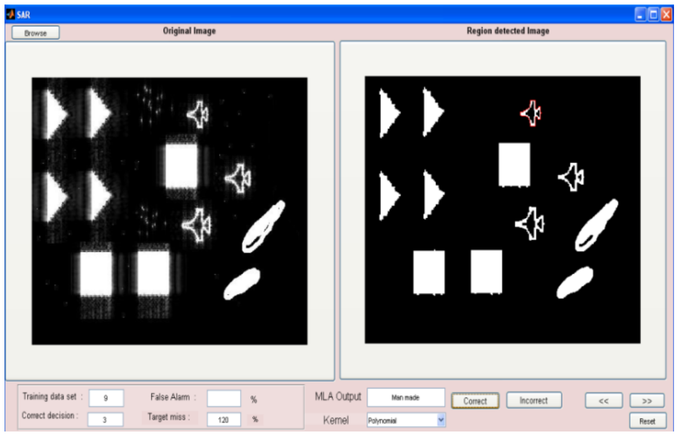


Figure 11 using polynomial kernel function

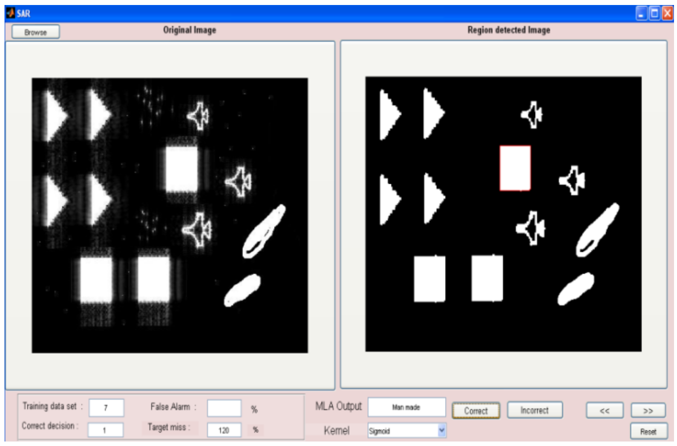


Figure 12 using Sigmoid kernel function

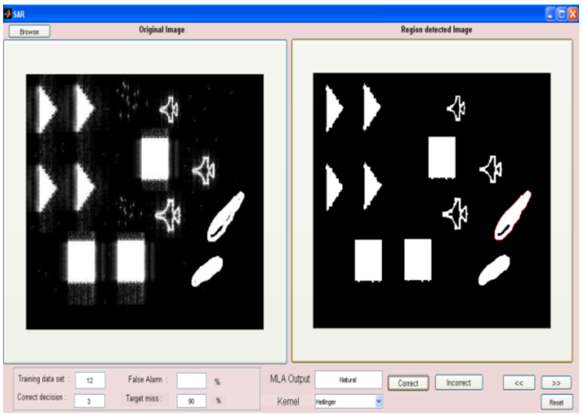


Figure 13 using Helinger kernel function

The two parameters measured for every incorrect decision made are False alarm and Target miss. They are given by the below formulae:

$$False\ Alarm = \frac{False\ count\ of\ Natural\ objects}{Total\ count} \quad (21)$$

$$Target\ Miss = \frac{False\ count\ of\ Man\ made\ objects}{Total\ count} \quad (22)$$

Table 2 displays significant improvement i. e., decreases in false alarm % and target miss % as the training data set size increases.

Table 2 Improvement in SVM based object classification with larger data set for figures 10, 11, 12, 13

Kernel	Analysis of SVM Algorithm			
	Training data set size	***False alarm***	***Target miss***	***Efficiency***
RBF	123	16. 7%	0%	83. 3%
Polynomial	123	8. 33%	50%	41. 66%
Sigmoid	123	0%	60. 66%	33. 33%
Hellinger	123	0%	75%	25%

Figures 14 15 16 17 shown below are another set of GUI screen shots. Table 3 displays the significant improvement.

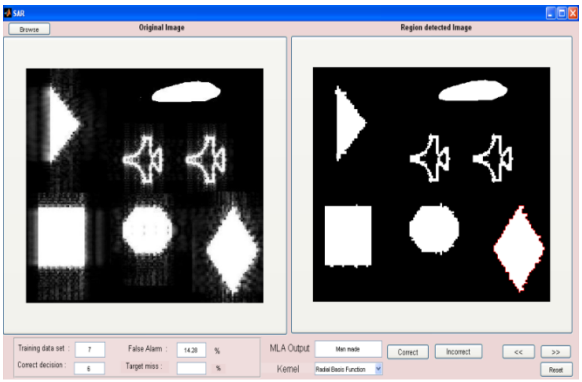


Figure. 14 GUI for interactive SVM Algorithm using RBF

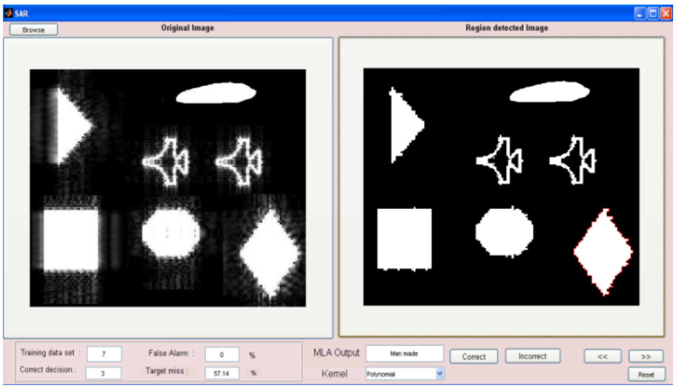


Figure. 15 GUI for interactive SVM Algorithm using Polynomial

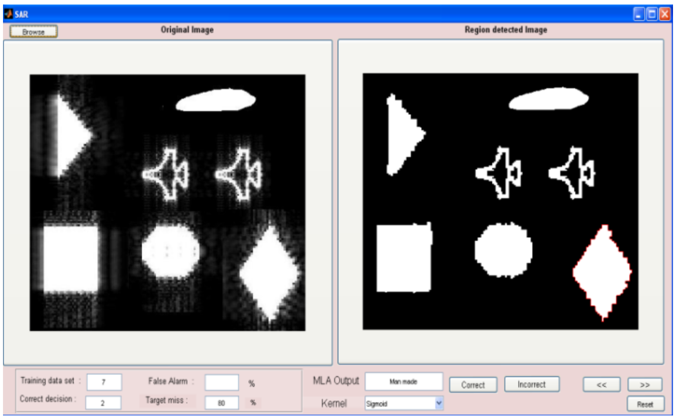


Figure. 16 GUI for interactive SVM Algorithm using signmoid

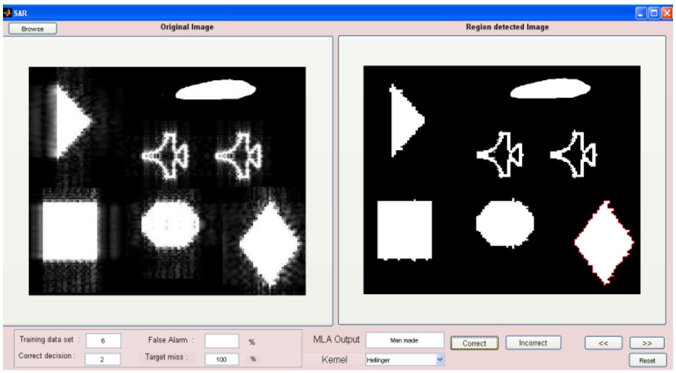


Figure. 17 GUI for interactive SVM Algorithm using Hellinger

Table 3 Improvement in SVM based object classification with larger data set for figures 14, 15, 16, 17.

Kernel	Analysis of SVM Algorithm			
	Training data set size	***False alarm***	***Target miss***	***Efficiency***
RBF	123	14. 28%	0%	85. 714%
Polynomial	123	14. 28%	42. 85%	42. 85%
Sigmoid	123	0%	71. 42%	28. 57%
Hellinger	123	0%	71. 42%	28. 57%

Applications

Synthetic Aperture Radar imagery is the only imagery that can be acquired at any time of the day or night and during any adverse weather conditions. SAR products that are available are full resolution (25 m) and high resolution (240 m) images; Complex format raw SAR data that retain amplitude and phase information; geo coded images etc.

SAR's ability to pass relatively unaffected through clouds, illuminate the earth's surface with its own signals and measure the distance precisely make it useful for the applications like Lake and river ice monitoring, Cartography, surface deformation detection, Glacier monitoring, forest cover mapping, Urban planning and monitoring disasters such as forest fires, floods, volcanic eruptions and oil spills.

V. CONCLUSIONS

The experiment is performed on a total set of 15 SAR Images. The image under consideration is first segmented into regions via Multi graph Image cut Segmentation. SVM technique is then applied on each region to classify it as either Manmade or Natural object. Various kernel functions like RBF, Polynomial, Sigmoid and Hellinger have been used. Upon the completion of classification process it has been observed that RBF kernel gives a better performance measure of False alarm and Target miss when compared to other kernel functions.

References

[1]. Mohammed Ben Saleh, Multiregion Image Segmentation by Parametric Kernel Graph cuts, IEEE Transaction on Image processing, vol 20 no 2, pp 545-556 (2011).

[2]. J. shi and J. Malik, Normalized cuts and Image Segmentation, IEEE Trans Pattern Anal. Mach, Intell, vol. 22 no 8, pp 888-905 (2000).

[3]. I. B. Ayed, A. Mitchie and Z. Belhadi, Multiregion level set partioning of Synthetic Aperture Radars, IEEE Trans Pattern Anal. Mach, Intell., vol. 27 no 5, pp 793-800 (2005).

[4]. Y. Boykov, O. Veksler and R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans Pattern Anal. Mach, Intell., vol. 23 no 11, pp 1222-1239 (2001).

[5]. A. Jain, P. Duin and J. Mao Statistical pattern recognition,, IEEE Trans Pattern Anal. Mach, Intell., 22, pp 4-37 (2000).

[6]. H. Mayer, Automatic Object extraction from aerial imagery survey focusing on buildings, Computer Vision and Image Understanding, vol. 74 no 2, pp 138-149 (1999).

[7]. C. Lin and R. Nevatia Building detection and description from a single intensity image, Computer Vision and Image Understanding, Vol 72, pp 101-121 (1998).

[8]. S. Krishnamachari and R. Chellapa, Delineating buildings by grouping lines with mrfs, IEEE Trans. on Pat. Anal. Mach. Intell., vol 5 no 1, pp 164-168 (1996).

[9]. Q. Iqbal and J. K. Aggarwal, Applying perceptual grouping to content based image retrieval, In Proc. IEEE Int. Conf. on CVPR, vol 1, pp 42-48 (1999).

[10]. B. Bradshaw, B. Scholkopf and J. C. Platt, Kernel methods for extracting local image semantics (2001).

[11]. Sarkar and P. Soundararajan, Supervised Learning of large perceptual organization: Graph spectral partitioning and learning automata. IEEE Trans. on Pat. Anal. Mach. Intell., vol 22 no 5, pp 504-525 (2000).

[12]. A. Oliva and A. Torralba The shape of the scene: a holistic representation of the spatial envelope, International Journal of Computer vision, vol. 42 no 3, pp 145-175, (2001).

[13]. A. Vailaya, A. K. Jain and H. J. Zhang On image classification: City images vs landscapes, vol 31, pp 1921-1936, (1998).

[14]. A. Olmos and E. Trucco, Detecting manmade objects in unconstrained subsea videos, In Proc. British Machine Vision Conference, pp 517-526 (2002).

[15]. X. Feng, C. K. I. Williams and S. N. Felderhof, Combining belief networks and neural networks for scene segmentation, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24 no 4, pp 467-484 (2002).

[16]. S. Konishi and A. L. Yuille Statistical cues for domain specific image segmentation and performance analysis, In Proc. IEEE Int. Conf. CVPR, pp 125-132, (2000).

[17]. Lee, K. –C., J.-S. Ou and M.-C. Fang Application of SVD noise reduction technique to PCA based radar target recognition, vol. 81, pp 447-459, (2008).

[18]. P. Felzenszwalb and D. Huttenlocher Efficient graph based image segmentation, IEEE Intl Conf Comp. vision, vol. 59 no 2, pp 167-181, (2004).

[19]. Yang. Y. Allen, J. Wright, Y. Ma and S. S. Sastry Unsupervised segmentation of natural images via lossy data compression, Comput Vis. Image. Understand, vol. 110 no 2, pp 212-225, (2008).