An Efficient Framework for Object Detection and Classification in Remote Sensing Images based on BOW and Unsupervised Classification Models

N. Bharatha Devi[1*] and Dr. A. Celine Kavida[2]
[1]Research Scholar, Anna University, Chennai, Tamilnadu, India.
[2]Associate Professor, Department of Physics, Vel Tech Multi Tech Dr. RR Dr. SR Engineering College, Avadi, Tamilnadu, India.
[*]Correspondence: bharathi.actech@gmail.com

Abstract: *Enthused by the current development of satellite and remote sensing images have attracted extensive attention. Nowadays, large number of research areas are focusing on developing applications It is one of the most significant challenges in real-world applications. The remote sensor collects data by detecting the energy that reflected from the earth in various location information and its store, retrieve, manage, display, and analyze all types of spatial data even though the accuracy is not satisfactory. This paper considers the problem of object detection and recognition as the main problem, and it is motivated to provide a better solution by designing and implementing an Efficient Framework for Object Detection and Classification (EFODC) on remote sensing images. The efficiency is improved by applying various image processing stages such as Image Acquisition and preprocessing, Image Enhancement, Object Detection, Bag-of-Words creation, and Training - Testing process. The bag-of-words method enables the user to maintain ground truth values for classifying the objects and improves the accuracy of classification. EFODC is experimented. The performance is evaluated by comparing with the state-of-the-art methods. Comparing with the existing approaches the proposed framework obtained 97.88% of precision and 97.47% of recall over 3000 images.*

Keywords: *Image Processing, Remote Sensing, Geo Images, Bag of Words, Object Classification.*

1. Introduction

Remote sensing is sensing the information about the object from the remote and not getting touch with the sensing object. The remote sensors are used to detect and gather information from the energy depleted from earth. These sensors may be satellites, or the sensors can fitted in an airplane. There are two types of sensors that are passive or active sensors. Passive sensors need external stimulators to react. It gathers the energy that depleted from the Earth's surface. The natural source of radiation used by these sensors is reflected sunlight. But the active sensors use the internal stimulators to gather information from the Earth. For example, a laser beam projected to the Earth's surface and the time can be measured to reflect the source.

Coastal applications that are to track the transport and monitor the shoreline, Ocean applications like to watch the ocean circulation, temperature, wave heights and follow sea ice. Also hazard assessments like to track hurricanes, earthquakes, erosion, and flooding. And also, natural resource management is to monitor wildlife habitat, to monitor wetlands. These are the some of the applications in various fields using remote sensing. Remote sensing provides the most necessary data in the monitoring of different applications like image fusion, detection, and land cover classification.

Object identification is an ability of the machine to recognize an object from an image to obtain information from the image data. This process of object recognition in an image might be a straightforward task to a human brain but not for a machine. A device needs a computer vision which is the ability of the machine to process the data, and it requires a large memory and many graphics capabilities. To recognize objects in a picture, there are several states of the art technologies proposed earlier for the process of object identification process with the method of implementations. Since there are several potential applications for object detection using computer vision technique this area of research is growing faster and the different technologies proposed every day are getting better regarding efficiency and accuracy. Object identification is one of the applications of computer vision.

There are several important applications for object identification using computer vision as said earlier, one among those is Satellite image processing application which is one of the most complicated forms of object detection. When compared to terrestrial images the satellite images are less in resolution, noisy and blur. Also, the number of objects in a satellite image is more than that of a typical terrestrial picture which makes the process of object identification in satellite images a much more complicated process even with several advancements in technology. The texture of objects present in a satellite image file is cumbersome and overlaps with other layers. When the layers or objects in an image overlap with each other there will be a loss of contour parts in any of the overlapped objects. There are also illumination variations in the satellite images which make object parts look like they belong to different regions. Due to this illumination variation in the object parts, it causes inhomogeneity in objects and causes overlapping in objects identified by the machine. The grayscale satellite image is less in size, cheaper and eventually has only minor information values and they are obtained just in conditions where there is no cloud. So, from the previously discussed factors, it is essential to process the image data for object identification using computer vision using the object contour information based on the level of brightness as well as the illumination information in the image to avoid objects overlapping.

The research works on satellite image examination conducted earlier are mostly focused on segregation of object areas and classification problems. Also, in previous investigations, the minimum target has also been restricted up to significant objects such as buildings or highway roads in high-resolution panchromatic satellite images. Meanwhile, the terrestrial image analyzing has processed up to sensing smaller things possible from the image data using computer vision. From all the previous works that discussed object identification in satellite images, it is clear that SoHee Jeon et al. are the first group of authors who talked about the identification of smaller things such as automobiles in his research work took place in 2005. In their work SoHee Jeon et al. used template matching method to identify smaller objects in the satellite image. The pattern matching requires high detail in the pixel. Since the pixel quality of satellite images reduced, the rate of recognition using pattern matching in satellite image is very less. Several research people discussed and provide various solutions for object detection and image processing regarding accurate object detection given in literature survey.

The key objective of this paper is to design and implement a novel framework for object detection and classification over remote sensing images with improved accuracy. A sequence of image processing steps is carried out and applied to the input image (see Figure-1) for improving the accuracy. The entire contribution of the paper is listed as:

- Learning the Objects in the Image
- Object Detection and Segmentation
- Matching the segmented object with the learned object and bag-of-words indexed objects
- Retrieve 60% of the database image for training process and 20% for the testing process.
- Apply pre-processing image methods for enhancing the image quality.
- Create, maintain and update a visual bag-of-words model for fast and accurate classification.
- Finally, the results are compared with the existing systems for performance evaluation.

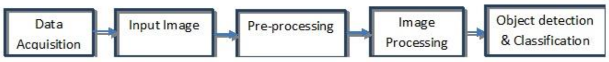


Figure-1: Semantic Procedure for Image Processing

Initially, sensors and sensor-based devices are used for sensing the remote information, and the data acquisition is applied to the monitored data to fetch the input image. The monitored data may be a video where it comprises of images. A set of all preprocessing steps are applied to the input image for enhancing the quality of the image where it increases the accuracy of image processing output. Finally, the appropriate objects are identified, detected, extracted and classified as trees, road, building, land, etc. The overall process is shown in Figure-2.

2. Background Study

Before designing the proposed framework, it is essential to understand the issues and challenges faced by earlier research works. Hence this section discussed some of the previous research methods. The process of extracting the information from image data is a multistep process among which the object detection process is the most important one. Computer vision is a way of understanding by which the machine understands the image or any other high dimensional data. It is mostly cast-off in the area of image processing. According to the concept discussed in [1], the process of extracting the specific content of an image using computer vision technique is the machine does not understand a challenging problem as the information in the image like a human brain does. The tool recognizes the term texture or pattern from an image can be used for several applications such as detection of roads, the positioning of an object in the image, etc. Among many applications, the identification of paths is a significant application which can be useful for routing of people and goods with better efficiency [2]. In [3], [5] Hough transform method is applied for feature extraction process which is used in image analysis, image processing, and computer vision processes. This technique extracts the features from the image data using the mismatch in the distances between the detected objects in the image based on a voting procedure. In [4], this voting is carried out based on a parameter from which the local maxima of the object obtained in an accumulator space considered as the reference space in the algorithm built for computing the Hough transform. There are several methods proposed earlier for the process of object identification in an image. The edge detection techniques are used to detect significant edges in the image. The edges in the image detected by the difference in the color of the adjacent pixels. In [6], canny edge detection a similar technique is used to find the edges in the image. Region growing method is another method used to detect objects in an image. When compared to edge detection methods these methods produce better results in noisy images. In this method, the similarity of the regions is used to segregate an image into objects. The similarity can be calculated using several factors such as color, size, shape, color levels and texture. In [7] split and merge technique is used to find regions and their boundaries.

Statistical methods use statistical analysis on the data obtained from techniques such as thresholding, component labelling, adaptive thresholding, amplitude projection and clustering techniques for identifying objects in an image. Kohonen maps [8] also known as self-organizing maps (SOM) used as a statistical method of object identification. This mapping has two modes of operation, a training mode where the map is built using input samples and mapping mode where the input is classified based on the constructed map. In [9], a model called Active Appearance Model (AAM) proposed which is a knowledge-based method. Even though the knowledge-based techniques are affected by object variability, they are the most efficient methods when the image objects are similar to each other. There are also other methods proposed combining any two of the methods discussed above. In [10] a hybrid method called watershed transformation method is introduced where the image characteristics obtained from the results of techniques discussed above. In [11], a neural network-based hybrid method is proposed. In this method, the segmentation algorithm draws boxes from a set of pixels named seeds to find the boundary of the region. Other than methods discussed above [12], [13], [14] and [15] authors have presented various target detection techniques based on visual saliency. One of the significant methodologies is feature extraction and feature comparison for classifying the images.

Some of the applications are particularly important and essential for detecting and recognizing the objects such as image retrieval [16], automatic vehicle detection and tracking [17], sign recognition [18] and mobile robot localization [19]. Thinking of various images processing stage, enhancing the image quality is the most critical step. To improving the image quality, multiple filters are used like linear filters [20] are used for obtaining theoretical foundation. Some of the non-linear filters like stack filter [21], medial filters [22], order-static filters [23] and weighted median filters [24] are used for increasing the image quality, which used in the earlier researches. Regarding feature extraction, it leads to improving the accuracy of object classification. Some of the methods like SIFT [25], SURF and U-SURF [26] methods are considered as pixel level feature extraction methods. Some of the research works [27-32] discussed various texture feature information for object detection and classification. It involves first and second order statistical features, GLCM feature, Gabor features, and histogram and fractal information of the objects available in the RS images.

From an in-depth study of earlier research works, some of the issues and challenges found such as indecision in image classification, influence on spatial resolution and choosing the best variables. The above-said factors influence the classification accuracy since uncertainty occurs in various stages of image classification process. It is necessary to understand the various stages and the contribution of risk to quality image classification. Another critical factor which affects the classification accuracy is the spatial resolution. Hence it is essential to choose the spatial information of the image for classification. To do that different reduction method are used for improving the classification accuracy.

Finally, to improve the classification accuracy, the feature selection and extraction is the more important process for improving the classification accuracy. And it is decided from the background study, that still the classification accuracy should be developed by concentrating on pixel level information, extract and select features based on the spatial and temporal information. The entire architecture of the proposed work figuratively illustrated in Figure-2. Also, the step by step procedures is described in detail below.

3. Proposed System

The architecture flows from training phase to testing phase.

Image registration

Image registration determines the relative orientation between two images. Process of transforming different sets of data into one coordinate system that called as image alignment. Data may be multiple photographs, data from various sensors, times, depths or viewpoints.

In Image registration techniques, the control points locating is not as simple as with moderate resolution images, manual selection is tedious, time consuming and high data volume affect the processing speed. It is suitable for computer vision, medical imaging, military recognition and analyzing images from satellites. According to the transformation models, the target image space relate to reference image space that includes rotation, scaling, translation and elastic, nonrigid transformations. Translation of a full image characterized by a translation vector called as parametric model.

Image Pre-Processing

Process to access a digital image from image database that involves extraction, analysis, recognition of image coding, filtering, normalization, segmentation, and object identification.

Image Noise Removal

Image noise is random variation of brightness or color information produced by the sensor, scanner or digital camera. The noise is divided into Gaussian noise, balanced noise and impulse noise. Linear or nonlinear filter methods used to reduce noises. The median filter is a nonlinear filter used in digital image processing and is a rank-order filter depend on size and shape of filtering mask. The algorithm complexity depends on how to get the median value.

(1) The mask may be adaptively resized according to noise levels.

(2)To find the median value of filtering mask for all the pixels and statistical histogram is introduced to speed up searching process.

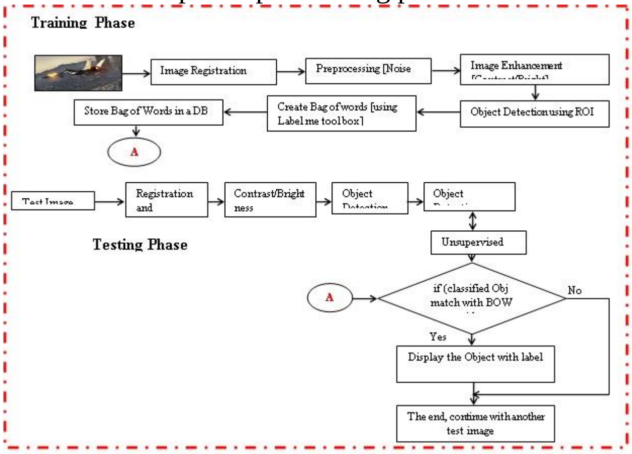


Figure-2: Overall diagram of proposed work

The noise occurred in the image $\boldsymbol{I}$ is removed by the filter parameter "$\boldsymbol{h}$".

$$\mathbf{DIhv + n(DIh, v)}$$

Where the denoised image is represented as DIh, and $n(DIh, v)$ is noise predicted, and v is the decomposed image.

Image Enhancement

To improve the quality of the image by manipulating software. Image enhancement or image editors used to remove noise, correct the image density, contrast and help to store and retrieve the image for visual interpretation. The range of digital value is 8 bits or 256 levels.

The different image enhancement techniques used in linear contrast stretch, histogram-equalized stretch, spatial filtering.

Figure-3: Image Enhancement

4. ROI Based Region Detection

A region of interest is a group of data sets used in many applications either feature-based or object-based approaches. Clustering points of interest determines the object location. A blob is a large binary object stored in database management system that is not universal. Blob extraction or region detection or labeling isolate the BLOBs (objects) in a binary image that consists of a group of connected pixels.

The image is scanned and every pixel is labeled with an identifier. Blob detection is performed on binary image by SUSAN (Smallest Univalve Segment Assimilating Nucleus) edge detector. The blob extraction process extract single image by different filtering techniques.[9] [23] location-related pixel information is used to count persons in images that are not consistent with the defined size. Blob's length and width values are used to find the area and centroid of sub image scaled to a fixed size to obtain with high probability.

Zhang et al. evaluated the effectiveness of different color spaces in face and lip regions by plotting the histograms of RGB. Hue-based transforms for lip-skin segmentation achieving greatest accuracy 94%.The hue component is rotated about 0 by 0.0333 within the range [0, 1] performing a thresholding operation then all non-red hue components can be removed from the image. Once the thresholding operations are completed all remaining pixels are designated as lip pixels. The result is a binary image:

$$T(i,j) = \begin{cases} 1, for\ H(i,j) > H_0\ \ and\ S(i,j) > S_0 \\ \qquad\qquad 0, otherwise \end{cases}$$

Where H is hue, S is saturation, i and j are the pixel locations. An accumulated difference image (ADI) is calculated to remove non-lip red hue pixels in binary image.

$$ADI_k(i,\ j) = \sum_{q\,=\,k\,-\,98}^{k} \Delta R_q(i,j)$$

$$\Delta R_q(i,j) = \left|R_q(i,j) - R_{q-1}(i,j)\right|$$

Where k is the frame number and R is the red component of RGB. Lip segmentation utilized for lip reading, speech recognition, facial expression, emotional state, pain recognition. The color transformations are compared based on histogram intersection and Otsu's method. AND operation is performed on the threshold binary ADI and hue images. A bounding box drawn around the mouth region. Full ROI refers centered mouth and chin and partial ROI refers only centered mouth. Poor ROI refers other region is selected than the mouth.

Bag-of Words Method

In the Bag of words (BOW) representation as an unordered collection of words and a text document is encoded as a histogram in image. It divided into local region detection, words encoding and histogram computation. Bag of visual words (BOV) computed by attention-based local descriptors that evaluating new features and classification performance. Combinations of feature descriptors, classifiers, datasets produced an average accuracy around 90% over global representation.

BoW model done by feature detection and description with codebook generation used in information retrieval (IR), document classification. BOV is a vector of local image features that objects are recognized by predefined shape and edge detection approach. BOW created using online labelling software. GIS and G-MAP labels are stored in a bag as an index.

BoW model-histogram representation based on content based image indexing and retrieval (CBIR)[3].Each image is abstracted by several local patches as numerical vectors. Scale-invariant feature transform (SIFT) converts each patch to 128-dimensional vector.

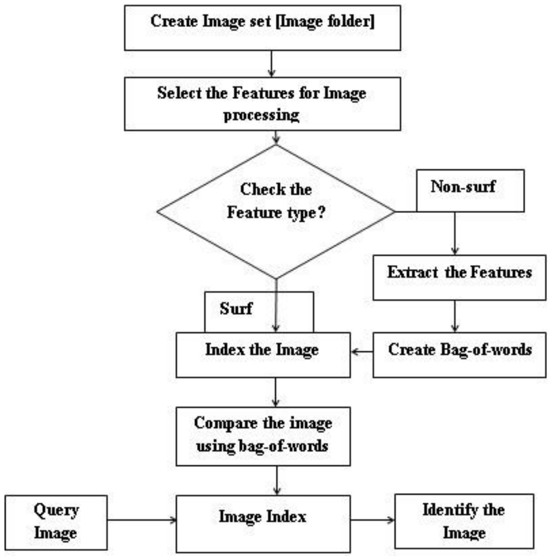


Figure-4: Bag of Words Based Image Comparison

Vector Quantization (VQ) is lossy image compression techniques performed by Codebook Generation, Image Encoding and Decoding. In codebook generation, an image is split up into blocks of size 4 x 4 pixels and converted into vectors of K dimension. K-means clustering over the vectors called training vectors. The codebook and codevector size are directly proportional to the value of the locally global codebook. BoW model is to convert vector- patches to code words. Each patch in image is mapped to a particular codeword through the clustering process. BoW model for image divided into generative and discriminative models.

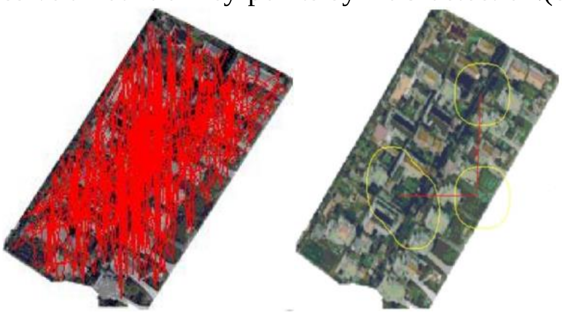
Learning the Objects

Supervised learning methods used to learn the objects internally and externally in remote sensing images. Object detection done by the amount of spatial resolution. Unsupervised learning methods for learning and classifying the objects in the remote sensing images. Texture and spectral features for describing ROI detected region in grid by calculating means and standard deviations from 11×11 pixels and form a patch. GLCM is a tabulation to generate the matrix at a fixed relative position.

Feature	Description
GLCM_Contrast	$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$
GLCM_Correlation	$\sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right]$
GLCM_ Energy	$\sum_{i,j=0}^{N-1} (P_{i,j})^2$
GLCM_Homogeneity	$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$

Table-1: Features Used in BOW Model

Total features are calculated in each patch by applying region-based feature extraction. (Table 1).Each object is obtained using histogram values calculated by euclidean distance. The entire patch is assigned to similar visible words by calculating its occurrence. Figure-5 describes the complexity reduction in patch detection. Image extracts a massive amount of key-points by DoG detection.(b).



(a). (b).

Figure-5: (a) DoG key points
(b) salient regions in saliency model

Radial Basis Function-Support Vector Machine (RBF-SVM) is applied to objects using low-level descriptors and all the features are stored in the form of vectors. The classification function is expressed as:

$$f(x) = sign(\sum y_i \alpha_i K(x_i, x_i) + b\ )$$

Where xi values extracted from X and Yi label assigned to xi. Objects (oi) are binned histograms created by visual words in vocabulary. The set of classes is used to train SVM denoted by I.

Datasets

Researches developed several methods and algorithms with benchmarking datasets for identifying object in remote sensing images. Some available websites such as:

1.http://pan.baidu.com/s/1hqwzXeG - provided by the German Society for Photogrammetry, Remote Sensing, and Geoinformation (DGPF).
2.http://www.ifp.uni-stuttgart.de/dgpf/DKEPAllg.html- SZTAKI-INRIA building detection dataset. This dataset is an outstanding resource for benchmarking building extraction algorithms.
3.http://web.eee.sztaki.hu/remotesensing/building_benchmark.html- This dataset contains 30 images extracted from Google Earth and a total of 1319 manually labeled cars with axis-aligned bounding boxes used for ground truth.
4.http://ai.stanford.edu/~gaheitz/Research/TAS/tas.v0.tgz,http://sourceforge.net/apps/mediawiki/oirds - Overhead imagery research dataset (OIRDS) - The OIRDS is designed for vehicle detection algorithms.
5 http://www.cse.iitm.ac.in/~vplab/satellite.html- this dataset was made from satellite images with a 1-m spatial resolution from Wikimapia (http://www.wikimapia.org/).

Dataset was separated into four types with 50 images per category and a ground truth roadmap obtained using a human operator. NWPU VHR-10 dataset used for both single and multi-class objects detection. The 10 classes of objects are an airplane, ship, storage tank, baseball diamond, tennis court, basketball court, ground track field, harbor, bridge, and vehicle. It contains 800 VHR optical RSIs, from which 757 airplanes, 302 ships, 655 storage tanks, 390 baseball diamonds, 524 tennis courts, 159 basketball courts, 163 ground track fields, 224 harbors, 124 bridges, and 477 vehicles annotated with axis-aligned bounding boxes used for ground truth.715 color images were acquired from Google Earth with the spatial resolution ranging from 0.5-m to 2-m, and 85 pans sharpened color infrared (CIR) images were obtained from Vaihingen data with a spatial resolution of 0.08-m.

6. Experimental Results and Discussion

Sequence of image processing by MATLAB software and road based, land based, residential based and agriculture based dataset collected from the website (Fig.2). Objects are trained and label is created by Label toolbox that named as codes words are stored in Table-2. Column numbers 1, 3 and 5 shows the objects and 2, 4 and 6 shows the appropriate codes (labelled) created in the toolbox in table-1.The first two characters of each codes indicate the name of the objects to improve the classification accuracy.

The entire dataset is divided into 60**%** of training data set and **40%** of testing data. Each object is labelled in training process and comparing the images in testing process. Input testing images are chosen randomly such as roads, agriculture lands and some buildings. Object considered as patches and all the patches are assigned by constructing a set of vocabulary (500 visual words). Finally, SVM classifier is used to classify the objects based on color, textures and valid feature set and comparing the objects based on the GLCM features. In Figure-7, building, trees and roads are segmented and extracted.Figure-7(a) shows the input image,(b) shows the gray scale converted and noise removed image,(c) shows the color space model for differentiating different objects. Red color for representing buildings, green color for trees/lands and blue color for representing roads. The input image is shown in Figure-7.

Figure-8 shows the result obtained from the experiment over image-2.Image-2 is applied as input image in the framework and results are verified.Figure-8 (a) is the input image, (b) is the buildings segmented from the input image and (c) shows the roads in the objects. From image-3 based results are shown in Figure-9.The segmentation of road obtained more accurate. Finally, image-4 is used verifying the action of the framework. Input image (a) is processed when the size is small and the efficiency is verified when the image is enlarged. Object detection is experimented from different dataset-based images in Figure-11.

Same process is repeated on the entire set of database images that classified by proposed framework and their efficiency is verified by calculating TP, TN, FP and FP with Sensitivity, specificity and accuracy. The performance is evaluated by comparing classification accuracy with existing approaches in Table-3.Among the total number(3000) of images,800 images are suitable for road extraction, 900 images for building extraction and 900 images for trees extraction and 400 images for extracting building, trees and roads in order to compute the correct and incorrect classification. From the experiment the classified results are given in Table-3. TP, FP, TN and FN values are calculated with Precision and Recall to evaluate the performance. The obtained precision and recall values are compared with the existing research works in [36, 37] is given in Table-4.

Table-3: Comparison of Objects Classification

Images	Objects Clearly available	Objects not Clearly available
Data Base Images	2600	400
PF: Correctly Classified	2545	334
PF: Incorrectly Classified	55	66

***PF – Proposed Framework**

From the above table, the TP, TN, FP and FN values are calculated and given here to evaluate the performance.

TP = 2545
TN = 334
FP = 55
FN = 66
Precision = 97.88%
Recall = 97.47%

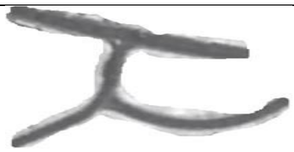
Tabel-4: Performance Comparison in terms of Precision and Recall

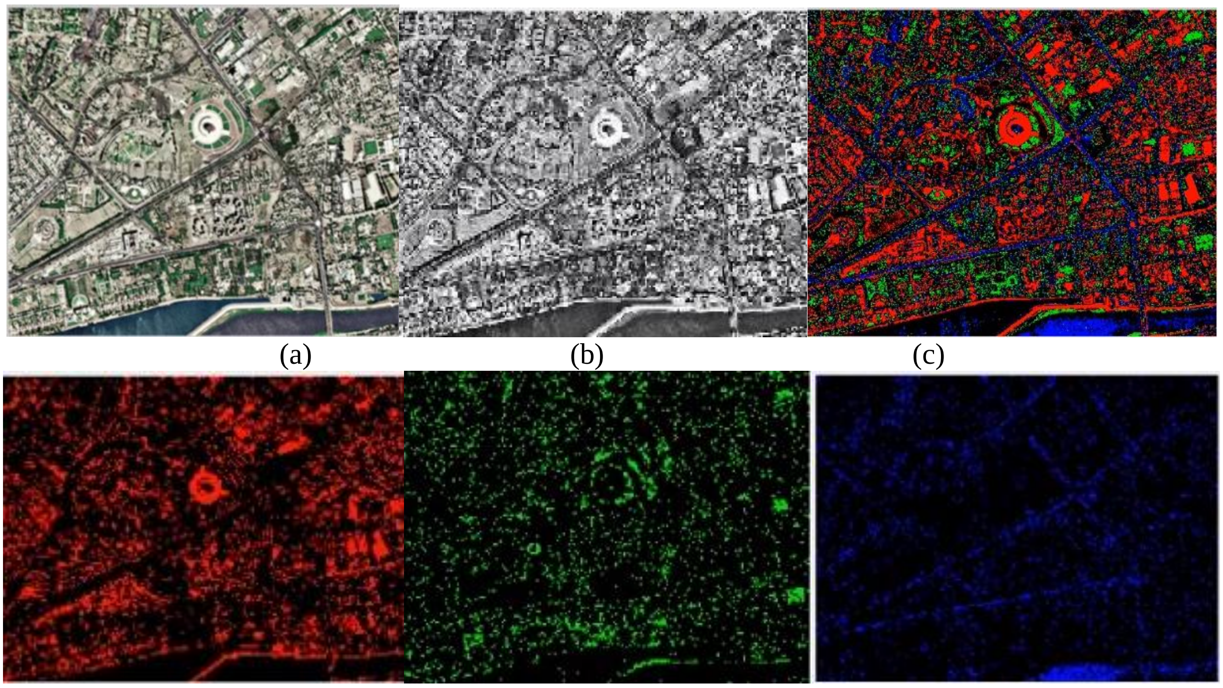
Methods	Precision (%)	Recall (%)	Time(s)
DSIFT+BOW+SVM+GRABCUT **[Ref-37]**	89.22	87.80	515.1
DSIFT+BOW+SVM **[Ref-37]**	86.10	79.77	419.6
DSURF+BOW+SVM+GRABCUT **[Ref-37]**	88.56	80.34	683.6
HOG + SVM + GRABCUT	76.43	87.11	205.5
Proposed Framework	**97.88**	**97.47**	**188.78**

From table-4, it is very clear and identified that the obtained precision and recall values using proposed framework is comparatively high than the existing approaches discussed in [37]. Hence, it is concluded that the proposed framework is suitable for object detection and classification in remote satellite images.

Figure-6: Set of all Input Images

Table-2: Code Word Created and Stored in Codebook

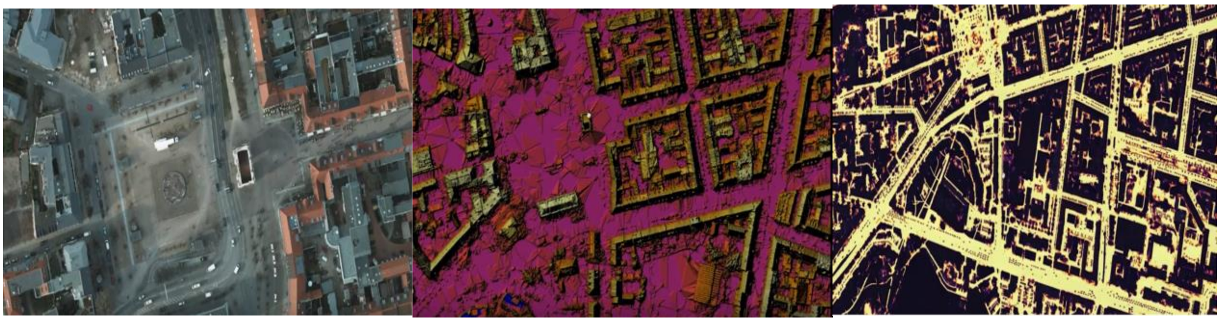
Object	Code	Image	Code	Image	Code
	Tr0001		Bu0001		Rd0001
	Tr0002		Bu0002		Rd0002
	Tr0003		Bu0003		Rd0003
	Tr0004		Bu0004		Rd0004
	Tr0005		Bu0005		Rd0005



(a) (b) (c)

(d) (e) (f)

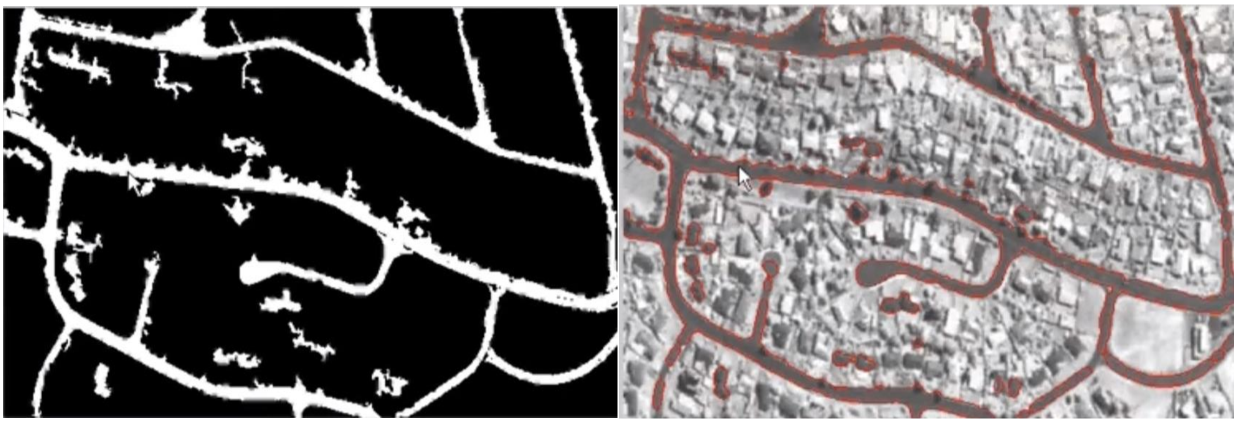
Figure-7: Results Obtained from Image-1. (a). Input Image, (b). Noise Removed Image, (c). Color Enhanced Image, (d). Buildings, (e). Trees, (f). Road.



(a) (b) (c)

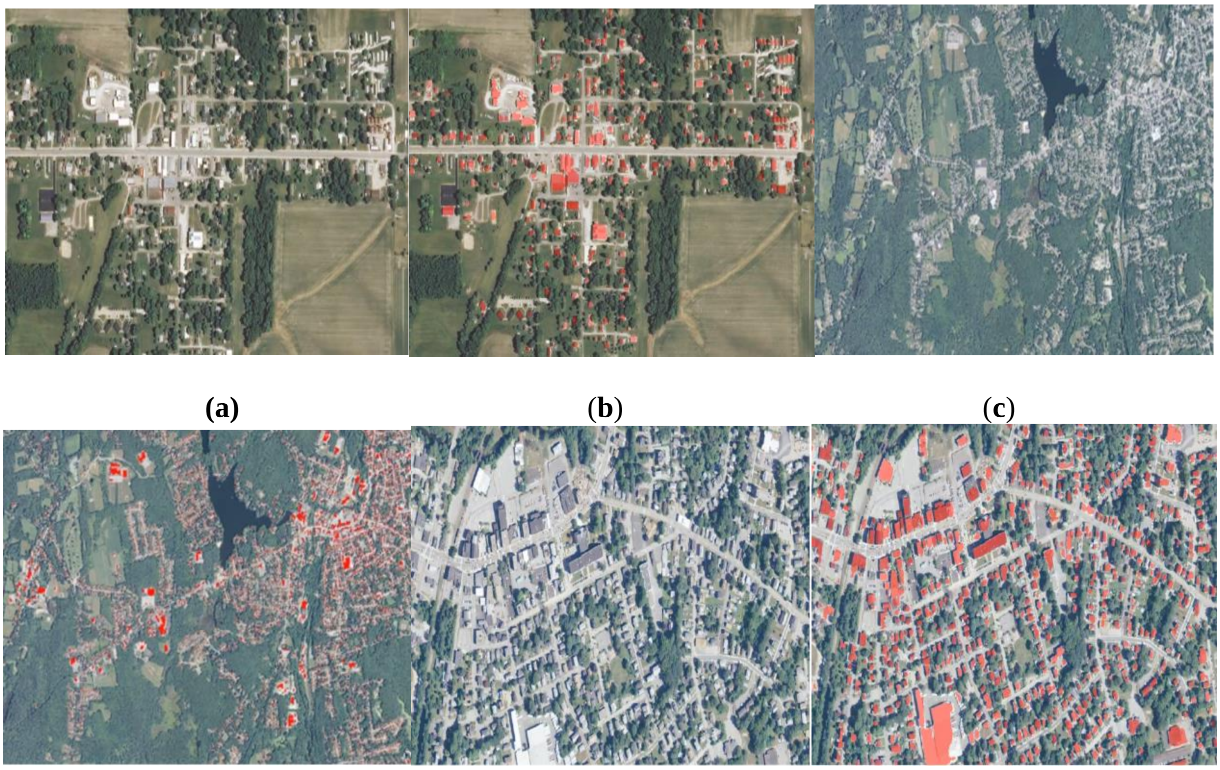
Figure-8: Results Obtained from Image-2. (a). Input Image, (b). Buildings, (c). Roads

(a) (b)



(c) (d)

Figure-9: Results Obtained from Image-3. (a). Input Image, (b). Noise Removed and Contrast Enhanced Image, (c). Road Extracted, (d). Road Segmented



(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g)

Figure-10: Results Obtained from Image-4. (a). Input Image, (b). Exploring the Building, (c). Enlarged Image, (d). Segmenting Building, (e). Exploring only Building Portion, (f). Segmenting Buildings, (g). Building Segmentation Applied on the whole Image.

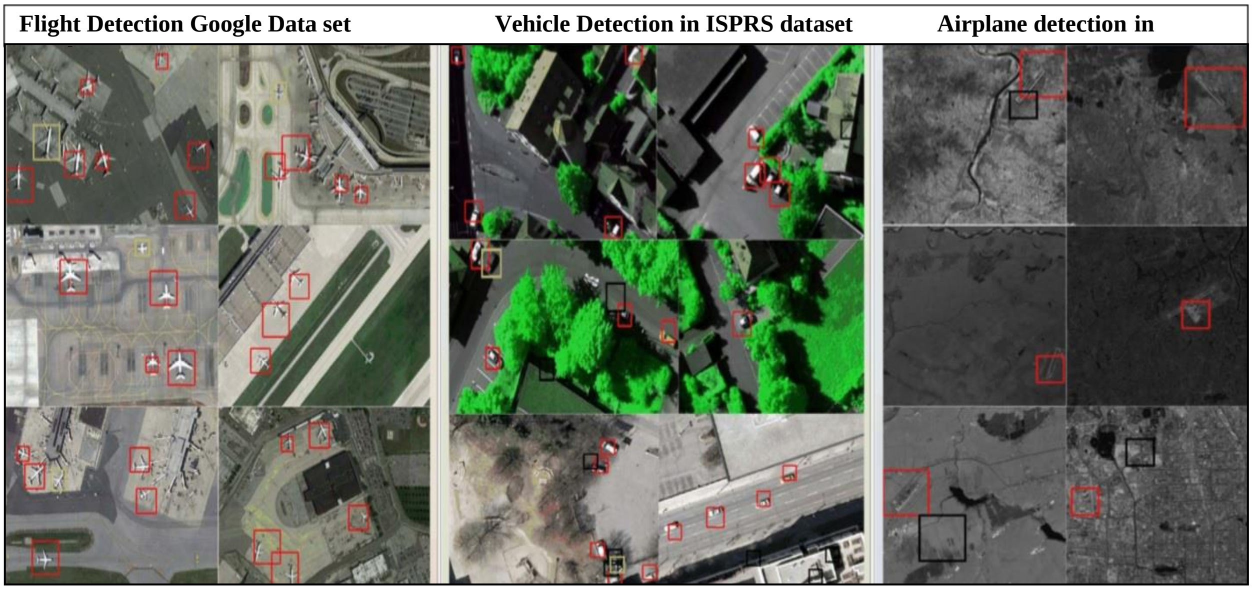


Figure-11. Different Object Detection from Various Dataset

7. Conclusion

Object detection in the remote sensing images is the challenging task, and there are various methods applied by multiple researchers to identify the objects. In this paper, we proposed a framework for the Object Detection and Classification in Remote Sensing Images based on BOW and Unsupervised Classification Models. We used the different stages of work, to improve the accuracy in object detection. We have experimented with the system and verified the results with the other modern methods and thus proved that our proposed framework gives more accuracy in detecting the object. In-depth learning-based feature representation and weakly supervised learning-based geospatial object detection, to further advance the development of object detection task. It still needs many efforts to develop more effective methods to improve the detection accuracy. The obtained precision and recall using the proposed framework is 97.88% and 97.47% respectively and it is better than the other existing approaches.

References

[1]. D. ping Tian, "A Review on Image Feature Extraction and Representation Techniques," Int. J. Multimedia. Ubiquitous Eng., 2013.

[2]. V. Sivaraman, “Rural Road Feature Extraction from Aerial Images Using Anisotropic Diffusion and Dynamic Snakes,” University of Florida, 2004.
[3]. D. M. Alex, "Robust and Efficient Method to Extract Roads from Satellite Images."
[4]. T. M. Talal, M. I. Dessouky, A. El-Sayed, M. Hebaishy, and F. A. El-Samie,"Road Extraction from High-Resolution Satellite Images by Morphological Direction Filtering and Length Filtering,"Proc ICCTA08, pp. 137–141, 2008.
[5]. Schneiderman, Henry, and Takeo Kanade. "A statistical method for 3D object detection applied to faces and cars. "In Computer Vision and Pattern Recognition, 2000.Proceedings. IEEE Conference on, vol. 1, pp. 746-751. IEEE,2000.
[6].Lienhart, Rainer, and JochenMaydt. "An extended set of Haar-like features for rapid object detection." In Image Processing. 2002. Proceedings. 2002 International Conference on, vol. 1, pp. I-900. IEEE, 2002.
[7].Felzenszwalb, Pedro F., Ross B. Girshick, David McAllister, and Deva Ramanan. "Object detection with discriminatively trained part-based models."Pattern Analysis and Machine Intelligence, IEEE Transactions on 32, no. 9 (2010): 1627-1645.
[8].Lienhart, Rainer, Alexander Kuranov, and Vadim Pisarevsky. "Empirical analysis of detection cascades of boosted classifiers for rapid object detection." In Pattern Recognition, pp. 297-304. Springer Berlin Heidelberg, 2003.
[9]. Gall, Juergen, and Victor Lempitsky. "Class-specific Hough forests for object detection." In Decision Forests for Computer Vision and Medical Image Analysis, pp. 143-157. Springer London, 2013.
[10]. Gall, Juergen, Angela Yao, Nima Razavi, Luc Van Gool, andVictor Lempitsky. "Hough forests for object detection, tracking, and action recognition." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33, no. 11 (2011): 21882202.
[11].Liebelt, Joerg, and Cordelia Schmid."Multi-view object class detection with a 3d geometric model." In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp.1688-1695. IEEE, 2010.
[12]. Laurent Itti, Christof Koch. A Saliency-based Search Mechanism for Overt and Covert Shifts of Visual Attention [J]. Vision Research 2000(40):1489–1506.
[13].Itti, L. Visual Attention and Target Detection in Cluttered Natural Scenes[J]. Optical Engineering, 2001, 40(9):17841793.
[14].Hae Jong Seo, Peyman Milanfar. Visual Saliency for Automatic Target Detection, Boundary Detection and Image Quality Assessment[C]. ICASSP 2010.5578-5581.
[15]. Wei Li, Chunhong Pan, Li-Xiong Liu. Saliency-based Automatic Target Detection in forward Looking Infrared Images [C]. ICIP, 2009:957-960.
[16] L. Ledwich and S. Williams. Reduced SIFT features for image retrieval and indoor localization. In Australian Conference on Robotics and Automation, 2004.
[17] K. Primdahl, I. Katz, O. Feinstein, Y. Mok, H. Dahlkamp, D. Stavens, M. Montemerlo, and S. Thrun. Change detection from multiple camera images extended to non-stationary cameras. In Proceedings of Field and Service Robotics, Port Douglas, Australia, 2005.
[18] C. Roduner and M. Rohs. Practical issues in physical sign recognition with mobile devices. In Thomas Strang, Vinny Cahill, and Aaron Quigley, editors, Pervasive 2006 Workshop Proceedings (Workshop on Pervasive Mobile Interaction Devices, PERMID 2006), pages 297– 304, Dublin, Ireland, May 2006.
[19] A. Gil, O. Reinoso, A. Vicente, C. Fern´andez, and L. Pay´a. Monte Carlo localization using SIFT features. Lecture Notes in Computer Science, 3522:623–630, 2005.
[20] A. C. Bovik and S. T. Acton, “Basic linear filtering with application to image enhancement,” in Handbook of Image and Video Processing. Academic Press, 2000, pp. 71–80.
[21] M. Gabbouj, E. J. Coyle, and N. C. Gallagher, Jr., “An overview of median and stack filtering,” Circuits, Systems, and Signal Processing, vol. 11, no. 1, pp. 7–45, Jan. 1992.
[22] B. I. Justusson, "Median filtering: Statistical properties," in Two-Dimensional Digital Signal Processing II, ser. Topics in Applied Physics. Springer Berlin/Heidelberg, 1981, vol. 43, pp. 161–196.
[23] A. C. Bovik, T. S. Huang, and D. C. Munson, Jr., “A generalization of median filtering using linear combinations of order statistics,” IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 31, no. 6, pp. 1342–1350, Dec. 1983.
[24] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, “Weighted median filters: a tutorial,” IEEE Transactions On Circuits And Systems II: Analog And Digital Signal Processing, vol. 43, no. 3, pp. 157–192, Mar. 1996.
[25] D. G. Lowe. Distinctive image features from scale-invariant key points. International Journal of Computer Vision, 60(2):91–100, 2004
[26] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In Proceedings of the 9th European Conference on Computer Vision, 2006.
[27] NYOUNGUI, A., TONYE, E., and AKONO, A., 2002, Evaluation of speckle filtering and texture analysis methods for land cover classification from SAR images. International Journal of Remote Sensing, 23, pp. 1895–1925.
[28] BARALDI, A. and PARMIGGIANI, F., 1995, An investigation of the textural characteristics associated with gray level co-occurrence matrix statistical parameters. IEEE Transactions on Geoscience and Remote Sensing, 33, pp. 293–304.
[29] PODEST, E. and SAATCHI, S., 2002, Application of multiscale texture in classifying JERS-1 radar data over tropical vegetation. International Journal of Remote Sensing, 23, pp.1487–1506.

[30] BUTUSOV, O.B., 2003, Textural classification of forest types from Land sat seven imagery. Mapping Sciences and Remote Sensing, 40, pp. 91–104.
[31] AUGUSTEIJN, M.F., CLEMENS, L.E. and SHAW, K.A., 1995, Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier. IEEE Transactions on Geoscience and Remote Sensing, 33, pp. 616–625.
[32] SOARES, J.V., RENNO, C.D., FORMAGGIO, A.R., YANASSE, C.C.F. and FRERY, A.C., 1997, An investigation of the selection of texture features for crop discrimination using SAR imagery. Remote Sensing of Environment, 59, pp. 234–247.
[33] J. Sivic and A. Zisserman, "A text retrieval approach to object matching in videos," in Proc. 9th IEEE Int'l Conf. Computer Vision, 2003.
[34] F. Perronnin, "Universal and adapted vocabularies for Generic Visual Categorization," IEEE Trans. Pattern Anal. Machine Intell. 30 (2008) 1243-1256.
[35] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2005.
[36]. Junwei Han, Dingwen Zhang, Gong Cheng, Lei Guo, and Jinchang Ren, (2015), “Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning”, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 53, NO. 6, pp. 3325-3337.
[37]. Yi Yuan and Xiangyun Hu, (2015), “Bag-of-Words and Object-Based Classification for Cloud Extraction from Satellite Imagery”, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 8, NO. 8, pp. 4197-4205.