# Classification and Area Computation Modeling of Remote Sensing Images Using Histogram and Convolutional Neural Network

Swarna Kamal Pradhan, Dipon Das and Ujjwal Mondal



August 27, 2021

# Classification and Area Computation Modeling of Remote Sensing Images using Histogram and Convolutional Neural Network

Swarna Kamal Pradhan[1[0000-0003-3940-0119]], Dipon Das[2[0000-0001-6800-8939]] and Ujjwal Mondal[3[0000-0001-7462-1959]]

[1]Department of Applied Physics, University of Calcutta, Kolkata, India
[2]Department of Applied Physics, University of Calcutta, Kolkata, India
[3] Associate Professor, Department of Applied Physics, University of Calcutta, Kolkata, India

**Abstract.** Remote Sensing is an important field in science and technology and consists of the images of Earth and satellites taken by the means of artificial satellites or aircraft. Satellite images or high-resolution aerial images, is flexible to work with and easy to monitor. Since, the total area of the earth is so large, high resolution remote sensing images produce vast amount of data, even image processing is time consuming. This work represents a combination of unsupervised and supervised process to classify high spatial resolution satellite images so that minimal human intervention is needed. For this purpose, histogram peak-based classification approach is used to classify remote sensing image into subcategories like urban land, vegetation land, water body etc. To detect different objects, present in the image, convolutional neural network-based approach is used. The neural network model is trained using custom dataset. Then object localization operation is performed to get the co-ordinates of the object present in the image. Then histogram-based segmentation operation is performed to compute the area of different objects present in the image. After that 3d model is constructed using the co-ordinates obtained. Georeferencing technique is used to calculate the area of different classes observed.

**Keywords:** Remote Sensing, Image Classification, Convolutional Neural Network, 3D-Modeling.

## 1 Introduction

The use of Image signal is huge and until today, there is still need to produce regional land use/ land cover maps for the variety of purposes of government, public, private, and national security applications. The main purpose is spatial information retrieval from remote-sensing images. Field based measurements give results at small areas with high temporal data, usually requires a lot of measuring and hard work. Additionally, this process is time consuming, may be less efficient, and varies from person to person. Satellite images or high-resolution aerial images, on the other hand, is flexible to work with and easy to monitor. Colour and texture features in colour image provide better result [1]. This information can be well observed by plotting colour histogram of the image. Colour histogram is related closely to scene properties and colour histogram has features that are identifiable and relates in a mathematical way to scene properties precisely [2]. When histogram is plotted for an image, it represents different colour classes present on that image and cluster formed on that histogram. These clusters together form different type of shapes, height and width, which represent distinct feature. Though histogram plotting is a powerful process for image information retrieval, it can be time consuming for a large image dataset. There are several methods proposed for calculating histogram in real time [3]. Histogram peak-based approach is proposed by V.V. Strelkov in 2008[4]. A peak or local maximum is defined as any sample whose two direct neighbours have smaller amplitude. In this approach a similarity between ordered histogram is used. Where, similarity is related to the closeness of positions and shapes of peaks in the compared histograms. Histogram peak-based approach also used for histogram stretching for contrast enhancement purpose using plateau equalization method with a scene-adaptive plateau threshold to enhance the raw image [5]. The histogram peaks that are above average value, represents number of clusters. A cluster is a small group of objects having similar property. Here a cluster represents a group of histogram peaks [6]. There are basically two types of clustering of data: (1) Unsupervised classification and (2) Supervised classification. Unsupervised classification does not require human involvement and it is a faster process on the other hand supervised classification needs large human intervention [7]. There are several unsupervised clustering techniques are available. K-means is a very popular unsupervised algorithm and widely used. It is used for partitioning N-dimensional points into k sets on the basis of a sample [8]. Supervised image classification is the task of assigning an input image, one label from a fixed set of categories. Since 2012 machine learning methods are used in various fields for classification purposes such as medical imaging [9], hand written letter recognition [10], face recognition, security systems, automatic zone detection of UAVs [11] etc. Machine learning in the field Geographical image classification is relatively new and has provided better result in the area of vegetation detection [12]. Using machine learning method, we obtain features from the image and train a neural network model using those features so that neural network model can predict the class of an unknown image.

The contributions are as follows:

- We introduced an object-based remote sensing image classification method using machine learning techniques.
- Used CNN based deep learning method for object detection and K-mean based unsupervised classification technique to classify the urban land.
- Have successfully managed to classify image and calculate their area using Georeferencing technique.
- Have successfully built a 3-dimensional model of the target image using the object coordinates.

## 2 Proposed Method

In this proposed method, the steps required for image classification and 3d modeling are described. This approach exploits a combination of unsupervised and supervised image classification technique to broadly classify the urban land.

### 2.1 Image Classification Using Histogram and K-Mean Clustering

Different types class present in a terrain image impact highly the shape of the histogram of the image. Depending on the classes present on the image, the histogram will produce unimodal, bimodal or multimodal distributions. These peaks of histogram distribution represent one, two and multiple classes (depending on colour) present in the image. Depending on the location of the peak and width, images can be classified into colour based sub categories. To plot the image histogram, at first the RGB model image is converted to HSV model image. HSV is a colour model where RGB pixel values are represented in cylindrical co-ordinate system. Then the position of the peaks in the histogram has to be calculated. A peak or local maximum is defined as any sample whose two direct neighbours have smaller amplitude. Then histogram peaks are detected. This process takes a 1-dimentional array and finds all local maxima by simple comparison of neighbouring values. The detected peaks are plotted with respected to their position. Then the peaks are classified using one dimensional K-mean clustering. K-mean clustering is a method of vector quantization.

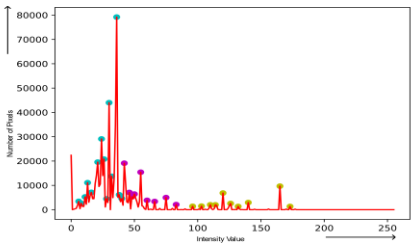


**Fig. 1.** Peak detected of a Histogram

### 2.2 Waterbody detection using Edge based operator

Some part of water body cannot be detected properly by only looking at the histogram as the colour of water has wide range of variety. So frequency based analysis is used. Water body region on the image has very low frequency noise compared to urban land. Our goal is to filter that low frequency signal for identifying water body. This is done using Canny-Edge detection. The Canny-Edge detector is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images. It has basically four stages. They are noise reduction with a Gaussian or median filter, finding intensity gradient, non-maximum suppression and hysteresis threshold.

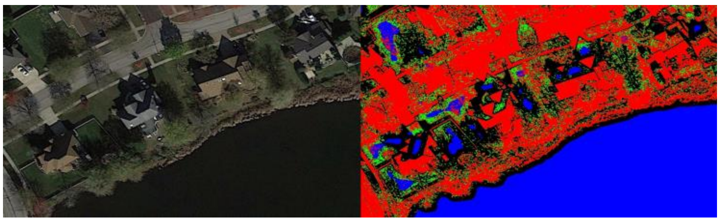

**Fig. 2.** On the left side image before false colour decomposition and on the right side image after false colour decomposition

### 2.3 Area Computation

To calculate the area of the highlighted part of the images we had to geo reference the image. The purpose is to transform image from geometric domain to geographic domain using base map or base image. To geo reference an image we followed certain steps.

**Choice of Ground Control points.** Ground Control Points (GCPs) are point in terms of latitude and longitude on surface of earth of known location used to geo reference map or remote sensing images of a region on earth.

**Assigning Latitude and Longitude value.** Also, this is necessary to determine the pixel values of the above-mentioned places on the raw image that we are going to geo reference, so that we can determine latitude and longitude variation per pixel.

### 2.4 Object Detection Using Convolutional Neural Network

**Structure of convolutional neural network.** Convolutional neural networks traditionally made of several types of layers, which are Convolutional layer, Pooling layer, Flatten Layer, Fully-connected (FC) layer.

*Convolutional layer.* The convolutional layer is the first layer of a convolutional network and it is used for feature extraction. Earlier layers focus on simple features, such as colours and edges. As the image data progresses through the layers of the CNN, it starts to recognize larger elements or shapes of the object until it finally identifies the intended object. Convolution are obtained by sliding a filter w, over the input matrix, x, and computing, at each position, the dot product between the filter and the patch of x starting at the current position. Here square filters are used to detect object, meaning that they have a size (200,200) pixels and (100,100) pixels. Though various combinations of height and width can be used. Mathematically, convolution of a matrix, x, with a filter, w, for all valid positions (i, j) can be given by:

$$(w * x)_{i.j} = \sum_{l=0}^{k-1} \sum_{m=0}^{k-1} w_{l,m}\, x_{i-l,j-m} \quad (1)$$

*Pooling layer*. Pooling layers, also known as down-sampling, conducts dimensionality reduction. It is used to reduce the number of parameters in the input. Similar to the convolutional layer, the pooling operation sweeps a filter across the entire input, but the difference is that this filter does not have any weights. There are two main types of pooling max pooling and average pooling.

*Flatten Layer.* In this layer two-dimensional data is converted to one-dimensional array so that it can be passed through the fully connected layer.

*Fully-Connected Layer*. However, in the fully-connected layer, each node in the output layer connects directly to a node in the previous layer. This layer performs the task of classification based on the features extracted through the previous layers and their different filters. Fully- connected layer performs the following operation on the input and return the output.

$$\text{Output} = \text{activation (dot (input, weights) + bias)} \quad (2)$$

**Evaluating the loss.** In this part we define a loss function. The goal of the loss function is to evaluate how well the network is performing. This function creates relation between predictions and network parameters, such as its weights and biases. Here we used binary cross-entropy (BCE) loss, which converts the predicted probabilities into logarithmic scale before comparing them to expected values. The BCE function can be expressed by the equation given below:

$$B(y, y^{true}) = \sum_i [-y_i^{true} \log(y_i) + (1 - y_i^{true}) \log(1 - y_i)] \quad (3)$$

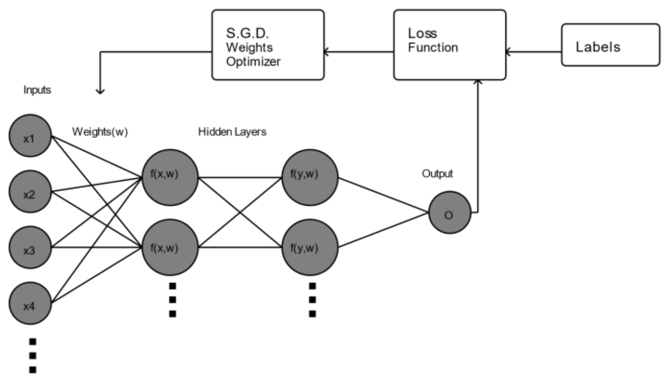


**Fig. 3.** Structural model of Convolutional Neural Network

After calculating the loss, we have to update the network parameters such that loss is minimum. So, stochastic gradient descent and back propagation operation is performed to optimize each parameter. At the time of optimizing, getting stuck in suboptimal local minima occurs, so stochastic gradient descent is used as it is able to leap out of local minima.

**Training the convolutional neural network model.** To train the model a python program is written and images are imported into an array. Images are imported into an array with their corresponding labels. For an example, house images are labelled as 1 and not house is labelled as 0, this is done to create house detection dataset. Similar operation is also done to create tree detection model. After importing the image data, they are converted to grayscale image for 2-Dimensional convolution. Then the image pixel values are normalized from 0-255 to 0-1.0 value for the ease of calculation. Next, training the dataset is shuffled such that stochastic gradient descent operation can be performed. After that the whole dataset is divided into training and testing dataset so that we can test the model after training. Then the created dataset is imported into this model for training. After training, the model is saved for further use and also validation accuracy is also observed for both house and tree model.

**Object detection.** The generated model basically works as a filter. The filter is then applied to an area of the image, and a dot product is calculated between the input pixels and the filter. This dot product is then fed into an output array. Afterwards, the filter shifts by a pixel, repeating the process until the filter has swept across the entire image. We used sliding window method to detect the objects present in the image. While performing image classification operation using sliding window, we given an input image and we present it to our neural network, and we obtain a probabilistic value associated with the class label prediction. These probability values represent how much close an image is, compared to the class of image we want to find. Then we chose a threshold value for the probability value so that upon which the object is present and below that value the object is not present. There is multiple object detection of the same object. The reason is that, we choose a threshold value for the probability output of the convolutional neural network output. So, we applied non-maximum suppression operation to remove the repetitive detection of the neural network. In this operation we took all the probability values for a single object and performed non maximum suppression operation so that only the probability value that represents the object most accurately remains. Then the coordinates of the final result are stored in a list so that these coordinates can be used for 3-Dimensional modeling purposes.

### 2.5 3-Dimentional Modeling

Once the location coordinates of the object are found, the 3-Dimensional modeling is done using Ursina library using those co-ordinates. For that purpose, we had built a model for each class in Blender software. The 3-dimensional shape file is created for those two classes and stored into ".blend" file format. Also, colour texture files are created using blender software and stored into the ".png" file format. Then those 3-Dimensional model files and texture files are imported into Ursina model using python. The ground colour is set using k-means clustering of the original image as described. The final output is given in the Figure 4.

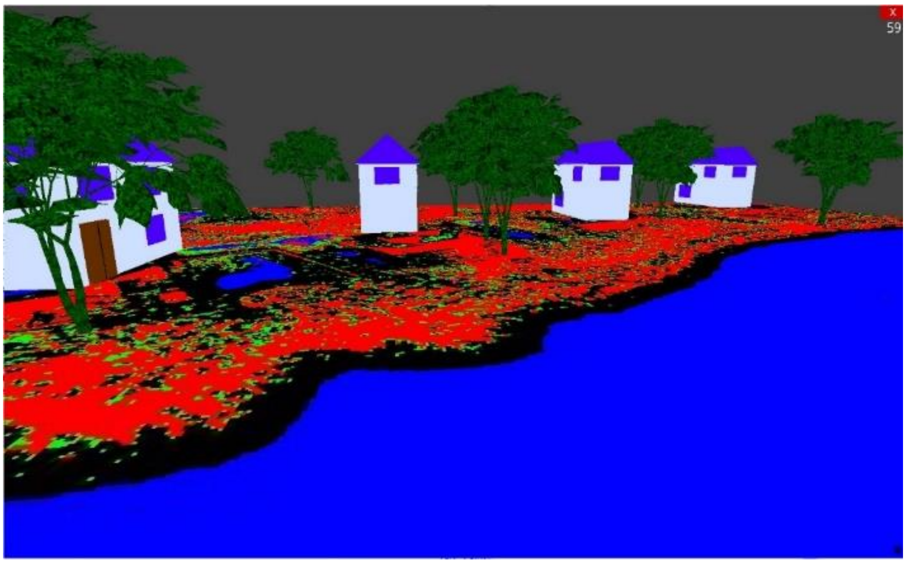

**Fig. 4.** Side view of generated 3-dimensional model

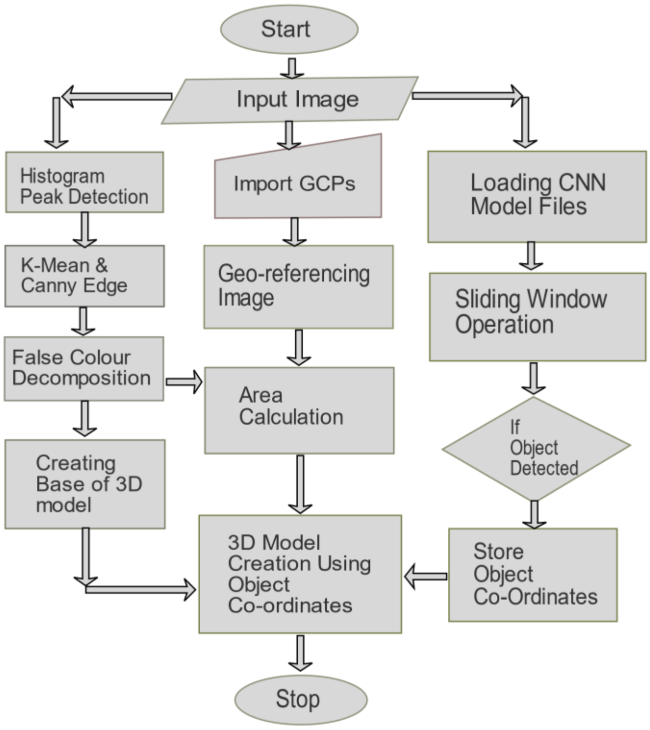


**Fig. 5.** Flowchart of the complete process

# 3 Conclusion and Scope for Future Work

## 3.1 Conclusion

We have successfully managed to classify image and calculate their area by histogram. Though we have faced certain difficulties and errors while performing this work, some of them are mainly atmospheric errors caused due to pollution, smog, dust particles present in the atmosphere and cloud cover. The histogram provides a complete solution for colour-based classification problem. It can be extended easily to higher (two or three) dimensional data spaces. Although, we have managed to classify images using convolutional neural network, which helped us to find objects present in the image, such as house, tree, etc. Using those object co-ordinates, we have successfully built a 3-dimensional model of the target image. Though rendering synthetic 3-dimensional models has enabled a variety of computer vision applications, it is, however, not the perfect remedy for data scarcity. Gathering hundreds of images for each new element to recognize is costly, time consuming and sometimes completely impractical, for instance, when the target objects are not produced or are only available at some remote location.

### 3.2 Scope for future work

We have only used grayscale images to train and test the neural network. So, in case of image classification colour based analysis can be used for further study and for better accuracy. This work includes limited data (one thousand images for each class). So, in future the image classification technique can be improved by using augmentation techniques with deep learning methods to diversify the training data so that more robust responses can be obtained [13]. We have successfully completed the classification of images based on their class and 3-dimensional structural modeling using convolutional neural network which can be used in various applications such as construction surveying, virtual tour, industrial structural planning etc. Till now in 3-dimensional modeling we have successfully managed to define the outline (specific location and area) of a particular object, but we could not define the exact structure and condition of the objects. So, semantic segmentation can be used for defining the precise contours of the object for object classification.

## References


1. ‘Color Texture Based Image Retrieval System’ June 2011 International Journal Of Computer Applications, Doi: 10.5120/2958-3910
2. ‘Anatomy Of a Color Histogram’ By Cl Novak, Sa Shafer - Cvpr, 1992
3. ‘Integral Histogram: A Fast Way to Extract Histograms in Cartesian Spaces’, By Fatih Porikl Doi: 10.1109/Cvpr.2005.188
4. ‘A New Similarity Measure for Histogram Comparison and Its Application in Time Series Analysis,’ By Vasily Strelkov, Doi: 10.1016/J.Patrec.2008.05.002
5. ‘A Quantitative Measure Based Infrared Image Enhancement Algorithm Using Plateau Histogram’, Rui Lai, Yin-Tang Yang, Bing-Jian Wang, Hui-Xin Zhou
6. ‘Automated Numbers of Cluster Determination Using the Combination of Entropy and Histogram Peaks from Multiple Images,’ Swati Chowdhury, Sudipta Roy, Anirban Mitra, Prasenjit Das, Doi: 10.14257/Ijsip.2016.9.7.10
7. ‘Image Classification in Remote Sensing’, 2013 Journal of Environment and Earth Science, Vol. 3, No.10,
8. Lloyd, Stuart P. (1982). "Least Squares Quantization in Pcm"(Pdf). IEEE Transactions on Information Theory. Doi:10.1109/Tit.1982.1056489. Retrieved 2009-04-15.
9. ‘Machine Learning for Medical Imaging’, Bj Erickson, P Korfiatis, Z Akkus, Tl Kline - Radiographics, 2017
10. ‘Handwritten Digit Recognition Using Convolutional Neural Networks’, IJIRCCE, Vol. 4, Issue 2, February 2016, Doi: 10.15680/Ijircce.2016. 0402001
11. ‘An Automatic Zone Detection System for Safe Landing of UAVs’. May 2019, Expert Systems with Applications 122. Doi:10.1016/J.Eswa.2019.01.024
12. ‘Vegetation Detection Through Smoke-Filled AVIRIS Images: An Assessment Using Modis Band Passe,’ Journal of Geophysical Research, Vol. 103, No. D24, Pages 32,001-32,011
13. ‘Vegetation Detection Using Deep Learning and Conventional Methods’, MDPI, Remote Sensing 2020 Doi:10.3390/Rs12152502.