MAN-MADE OBJECT EXTRACTION FROM REMOTE SENSING IMAGES USING GABOR ENERGY FEATURES AND NEURAL NETWORKS

Md. Abdul Alim Sheikh
Dept. of Electronics & Communication Engineering
Aliah University
Kolkata, India
alim.sheikh16@gmail.com

Alok Kole
Dept. of Electrical Engineering
RCC Institute of Information Technology
Kolkata, India
alokkole93@gmail.com

ABSTRACT

This chapter presents a novel approach for man-made object extraction in remote sensing images. This paper focuses on the design and implementation of a system that allows a user to extract multiple objects such as buildings or roads from an input image without much user intervention. The framework includes five main stages: 1) Pre-processing Stage. 2) Extraction of Local energy features using edge information and Gabor filter followed by down sampling to reduce the redundant information. 3) Further reduction of the size of feature vectors using Wavelet decomposition. 4) Classification and recognition of man-made structures using Probabilistic Neural Network (PNN) 5) NDVI based post-classification refinement. Experiments are conducted on a dataset of 200 RS images. The proposed framework yields Overall Accuracy (OA) of 93%. Experimental results validate the effective performance of the suggested technique for extracting man-made objects from RS images. Compared with other methods; the proposed framework exhibits significantly improved accuracy results and computationally much more efficient. Most notably, it has a much smaller input size, which makes it more feasible in practical applications.

Keywords—Remote Sensing Image, Man-made Object Extraction, Gabor Wavelets, Probabilistic Neural network

I. INTRODUCTION

Automatic extraction of man-made objects such as buildings and roads from RS images has been an important and popular research topic of interest in photogrammetry and computer vision for many years [1]. Extracting man-made objects such as buildings, roads from RS images is extremely useful in various Geographical Information System (GIS) applications like map generation and update, urban planning, disaster management, traffic management and land use analysis, change detection and military reconnaissance, socioeconomic parameter analysis, etc. [2-7],[8][9]. Although it is possible to manually extract objects from these images, this operation may not be robust and fast and will exceed the capacity of manual processing to carry out the extraction duties timely and economically [10][11]. Thus, there is a clear need for automatic/semi-automatic methods to detect and recognize man-made objects from RS imagery for turning data to information and extracting knowledge from information.

However, extraction of man-made structures like buildings, roads accurately and efficiently from RS imagery is still a challenging task with several difficulties. A number of strategies for extracting man-made objects from RS imagery have been presented in recent years. Some comprehensive reviews on man-made objects extraction from satellite can be found in [9], [12] [13], [14], [15-17]. Buildings and roads are one of the most important groups of man-made objects and automatic/semi-automatic extraction of building and roads can minimize the human labour in the application of map generation and updates.

The fundamental issue with these approaches is that the building is confused with other objects having similar spectral reflectance. The diverse characteristics and nature of objects appearing in urban environment such as colour, sizes and shapes, material, and interference of building shadows and trees make precise and reliable building extraction more complex and challenging [18], [12]. Many objects in high-resolution RS images, such as highways and parking lots, appear to be quite similar to buildings [19][20]. In practice, this has been the most significant stumbling block for RS applications. Therefore, automatic extraction of man-made structures accurately and efficiently from RS imagery remains a challenge that attracts huge research interests [21].

Based on recent advances, the use of Deep Learning methods is taking off to extract objects from RS data [22]. The Convolutional Neural Network (CNN) [23], [24] and the Fully Convolutional Neural (FCN) [25] network are widely used for these purposes. Though many techniques have been developed for road detection,

they still suffer from disadvantage. Due to the variety of buildings, road forms and the complexity of surrounding environment in reality, most of the existing methods extract objects from specific RS images and specific areas.

This chapter focuses on the design and implementation of a system that allows a user to extract multiple objects such as a building or a road from an input image without much user intervention. A novel self-assessed approach for detection and recognition of manmade object and natural objects (e.g., vegetation, natural water body, and seashore) from RS images is presented. Most of the man-made object, such as buildings and roads, have regular shapes with largely straight lines and consistent texture, whereas natural objects, such as vegetation and lakes, have irregular shapes with disorderly boundaries and textures [13]. In images containing manmade structures, the probability of possessing more edges segments, vertical and horizontal straight-line segments is very high, and the existence of vertical and horizontal straight-line is not accidental compared to natural structures scenes. For this reason, we try to exploit some non-accidental characteristics of images of both classes. To aim this, an automatic manmade object detection and recognition method is proposed, based on Gabor wavelet and Neural Network (NN). The features are extracted using the Gabor wavelets followed by down sampling by a factor to reduce the redundant information. In addition, dimension reduction method is used to further reduce the size of the feature vectors. Finally, the features are applied to a classifier for recognition. Spectral information is also used as a source for man-made object extraction by eliminating trees. For further refining of the building extraction, NDVI is utilized to eliminate the tree-generated lines. The main novelties of the proposed approaches consist of

(i) capable to extract multiple objects e.g., roads and buildings.

(ii) Techniques based on the efficient differentiation of sharp edges present or absent serve as a discriminative characteristic in separating man-made items from natural objects.

(iii) Exploiting domain knowledge and local interaction to achieve correct classification percentages with high accuracy

(iv) problem of large size feature is avoided. Efficient technique has been used that reduces the dimensionality of the feature vector by eliminating redundant information in the input feature vector. It speeds up the system. The proposed algorithm is validated by numerical results of real remote sensing images.

(v) The proposed framework's high computational efficiency and ease of implementation demonstrate its promise for any object extraction from RS images.

The rest of the chapter is as follows: in section 2, the suggested technique for man-made object extraction is described in detail. Then experimental results are provided in section 4.5 with a brief overview of dataset and experimental setup. Finally, section 4.6 summarizes the chapter with future scope.

II. PROPOSED APPROACH

The data flow diagram of the proposed method is shown in Fig.1, which depicts different stages clearly. The stages are (a) Pre-processing (b) feature Extraction (c) classification (d) Refinement of Building structures. Each stage is described in detail in the following sub-sections.

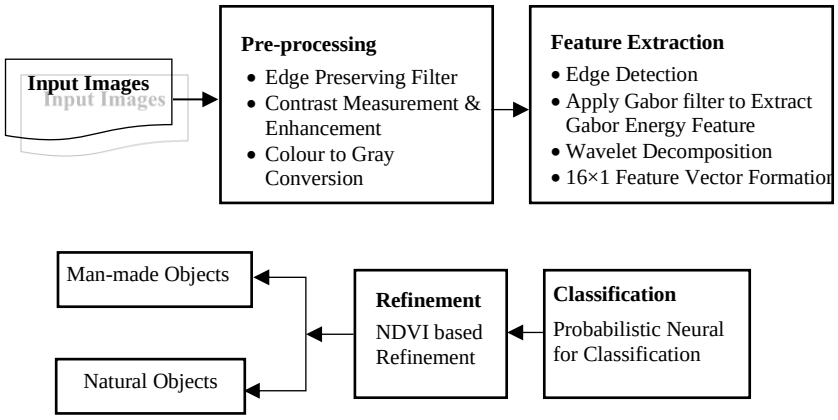


Fig.1. The data flow diagram of pattern classifier based on Gabor energy and Probabilistic Networks

A. Pre-processing

A pre-processing pipeline is adopted to cope with input images of varying quality, resolution, and channels to remove of noises and undesirable objects. In view of both noise diminution and edge preservation, bilateral filtering and Histogram Equalization (HE) are performed to preprocess the input image [26]. Bilateral filter [7] performs noise reduction and nonlinear smoothing on images by keeping the edge information by means of a

nonlinear combination of nearby image values. Fig. 2(b) shows the pre-processing result after bilateral filtering on the image shown in Fig.2(a).

Histogram Equalization (HE) [27] technique is used for image enhancement which adjusts the intensity histogram to approximate a uniform distribution. Fig.2(c) shows a HE processed image. After enhancement, the colour images are converted to gray scale images for feature extraction.

(a) (b) (c)

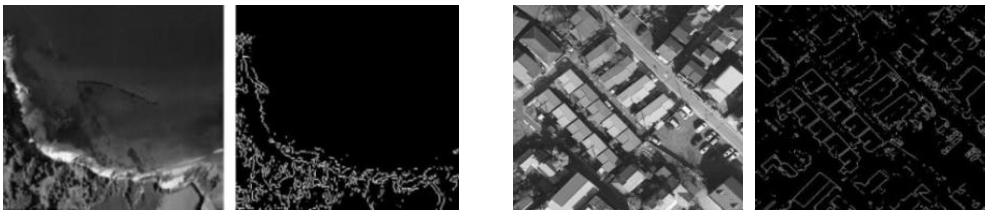
Fig.2. (a) Input Image of emerging suburban area (b) Result of bilateral Filtering (c) Histogram Equalization Processed Image

B. Feature Extraction

Local energy features are extracted using edge information and Gabor wavelets followed by down sampling to reduce redundant information and wavelet decomposition technique.

1) Edge Detection

To exploit the non-accidental occurrence of edges and straight-line segments, the Sobel operator is used. When compared to the other operators, this operator provides significantly larger output values for similar edges [33]. The presence of edge information of man-made and natural images is shown in Fig. 3.



(a) (b)

Fig. 3. (a) Image of natural scene and Sobel Edge of the image; (b) image with manmade structures and its Sobel edge output

2) Feature Extraction using Gabor Wavelets

Gabor filters may be thought of as orientation and scale tunable edge and line detectors, making them an excellent tool for detecting geometrically limited linear features from RS imagery, such as buildings and roads [28]. The invariance of Gabor filters to rotation, scale, and translation is its most significant advantage. They are also resistant to photometric disturbances such as illumination changes and image noise [28].

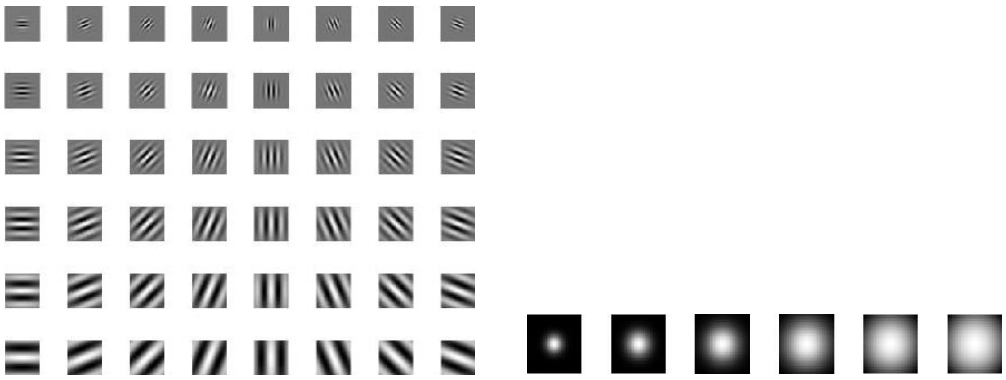
Gray-level images are directly used to extract Gabor filter-based features. Gabor functions can be obtained with a Gaussian window multiplied by a complex sinusoidal wave [29]. In spatial domain, a 2-D Gabor function is defined as:

$$g_{f,\theta,\varphi,\sigma,\gamma}(x,y) = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right) exp\{2\pi f x' + \varphi\} \quad (1)$$

Where $x' = xcos\theta + ysin\theta$ and $y' = -xsin\theta + ycos\theta$

Where f is the frequency of the sinusoidal factor, θ (value lies between 0 and π) specifies the orientation, φ is the phase offset, σ denotes standard deviation and γ is the spatial aspect ratio which clarifies the ellipticity of the cooperation of the Gabor function.

Different filters can be generated with varying values for orientation and scale. As a result, a Gabor filters bank is created, which is made up of a set of Gaussian filters with various radial frequencies and orientations that cover the frequency domain. The whole frequency spectrum, both amplitude and phase are captured by the Gabor filter family. In this step, forty-eight Gabor filters are employed in six scales and eight orientations as shown in Fig. 4.



(a) (b)

Fig.4. Gabor wavelets. (a) The real part of the Gabor kernel at scales = 6 and orientations = 8 (b) the magnitude of Gabor kernels at six different scales

3) Gabor Feature Representation

The Gabor wavelet representation of the image is obtained by convolving the image *I* of dimension M × N with every Gabor filter of the Gabor filter family as defined in Eqn. (2) at every pixel (*x; y*). The power spectrum of the filtered image at each pixel position is used as a discriminative feature to characterize that pixel.

$$r^2{}_{f,\theta,\varphi,\sigma,\gamma} = i\,(x,y) * g_{f,\theta,\varphi,\sigma,\gamma} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} i(m,n)\; g(x-m, y-n) \tag{2}$$

Where * is the convolution operation. The filtered responses of the symmetric and anti-symmetric filters are combined to form Gabor energy quantity. Then Gabor energy matrix E for each orientation is formed by finding vector sum of the corresponding filtered outputs from each filter bank

$$E_{f,\theta,\varphi,\sigma,\gamma}\,(x,y) = \sqrt{r^2{}_{f,\theta,\sigma,\gamma,0}(x,y) + r^2{}_{f,\theta,\sigma,\gamma,\frac{-\pi}{2}}(x,y)} \tag{3}$$

Where $r^2{}_{f,\theta,\sigma,\gamma,0}(x,y)$ and $r^2{}_{f,\theta,\sigma,\gamma,\frac{-\pi}{2}}(x,y)$ are outputs of symmetric and anti-symmetric filters.

The input images used in our experiments are 256×256 pixels in size. The Gabor wavelets depicted in Fig. 4 are used to extract the features. Fig. 5 shows the Gabor wavelet representation of a sample image of our database. These representations exhibit locality, scale, and orientation properties corresponding to those Gabor wavelets in Fig. 4.

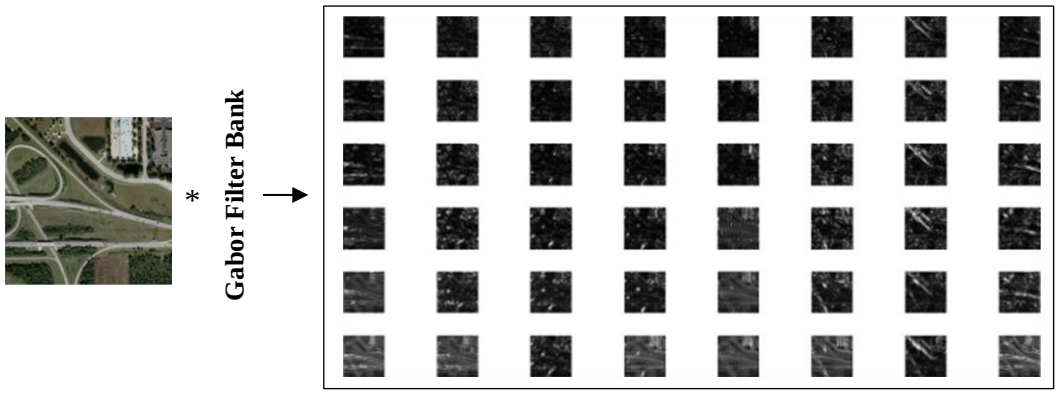


Fig.5. Results of feature extraction. To create a Gabor filter image, the input image is convolved with the Gabor filter banks. The Gabor filtered images' amplitudes at 6 scales and 8 orientations are shown

The feature images obtained from Gabor filters are further down sampled by a factor 8 to reduce the redundant information as the adjacent pixels in an image are usually highly correlated. The feature vector will have a size of (256×256×6×8)/(8×8) = 49,152 in total. After that, the vectors are normalized to have a zero mean and unit variance. In addition to down sampling, dimensionality reduction method is used to further minimize the size of the feature vectors.

4) Feature Reduction by Wavelet Decomposition

The size of extracted features in remote sensing image analysis is often large, which increases computing complexity and reduces system performance. To address these issues, Wavelet decomposition is used to reduce

feature dimension and their redundancies to a level that is easy for applying to classifier. Here, at first the available feature data is transformed to a wavelet form with a substantial proportion of its total energy packed into a small number of transform coefficients. So, taking these few coefficients, while neglecting the rest, we can retain the image feature. This is how the transformed data is reduced to a lower dimensional space.

The wavelet transform decomposes a signal into *wavelets* and *scaling*, which are formed by scaling and translating of a base function known as the mother wavelet [27]:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \quad (4)$$

Where "a" denotes the scaling and "b" is the translation. Each of the elemental functions or wavelets is applied to the original function to obtain the wavelet decomposition:

$$W_{a,b}(x) = \int_{-\infty}^{+\infty} x(t)\Psi^*_{a,b}(t)dt \quad (5)$$

Ψ^* denotes the complex conjugate of the function Ψ, and this is defined on the open (b, a) half-plane $(b \in R, a > 0)$

Due to the sparseness of the wavelet transform, the important coefficients of the transformed data have a larger magnitude than the unimportant coefficients. Coif3 wavelet is used due to its easy implementation, fast speed, shorter filter and easy to describe small texture structure, good resolution and smooth traits. Sixth level decomposition is used for reducing the feature size to 16×1.

C. Classification

A PNN network [13] is adopted to classify the input feature vectors into a specific class. PNN network is adopted for its many advantages e.g. a) training is easy and instantaneous b) training speed is many times quicker than back propagation c) an inherently parallel structure d) As the size of the representative training set expands, it is assured to converge to an optimal classifier (No local minima issues) e) Without extensive retraining, training data can be added or withdrawn f) Additionally, it is robust to noise examples.

The network structure in our proposed algorithm is illustrated in Fig. 6. The PNN used here has four layers: the input layer, Radial Basis Layer, the Competitive Layer, and the output layer.

1) Input Layer: This layer's sole purpose is to distribute input to all neurons in the pattern layer. The black vertical bar in Fig. 6 represents the input vector, designated as P. It has an R×1 dimension. R = 16 in this study.

2) Radial Basis Layer: It calculates the vector distance between an input feature vector p and the weight vector resulting from each row in weight matrix W.

3) Competitive Layer: Competitive layer has no bias. The binary output of competitive function is denoted by *a2*. The vector '*a1*' is multiplied with layer weight matrix $LW_{2,1}$ in this layer, yielding an output vector *a2*.

The network has categorized the test vector into one of K classes with the highest likelihood of being accurate. Once the classifier is mapped or trained, it is ready for future classification.

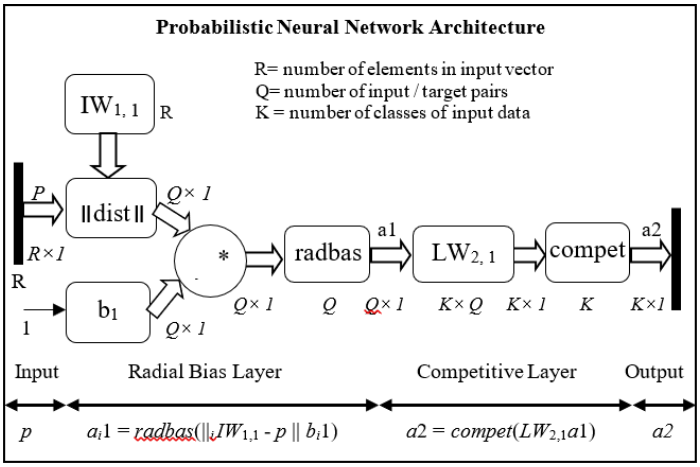


Fig. 6. Architecture of the Probabilistic Neural Network, R=16, Q=108, K=2

Between the input and output layer, the adopted neural network contains two hidden layers of 108 and 16 nodes, respectively. It has a 16-node input layer that receives the 16 features of each image region (8 Gabor energy matrices) and a single-node output layer that produce the classification decision.

D. Refinement of Results

The initial man-made object extraction results produced in the previous step may contain some non-building objects, e.g. trees. The NDVI is one key parameter, which is used here to differentiate between vegetated and non-vegetated objects [30]. The NDVI values show how much green vegetation is present in each pixel. The

classification is based on the simple assumption that objects with an NDVI of more than a specific value must be trees; and NDVI of man-made class is low. Those objects with NDVI values higher than the threshold values (i.e. trees) were removed from the classification result produced in the previous step. The NDVI can be calculated as follows:

$$NDVI = \frac{NIR+RED}{NIR+RED} \quad (6)$$

Where NIR: Near-infrared reflectance value
RED: Visible red reflectance value

Fig.7 indicates the refinement process employed for extraction of buildings and roads from images. The general rule for this refinement process is that man-made class will be found when NDVI is low, and vegetation class e.g. trees will be found NDVI is very high.

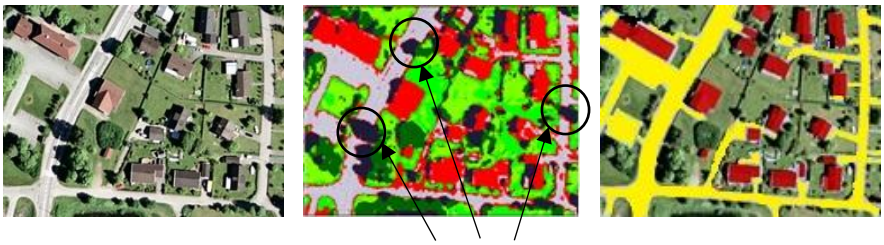


Fig.7. Input image, classified image and extracted buildings and roads after refinement

Knowledge engineer, which is included in ERDAS Imagine Software, was used to create an expert system for post-classification refinement. Building class is discovered when the NDVI is less than -0.038, road class is found when the NDVI is less than 0.02 and tree class is found when the NDVI is larger than 0.1, as in our example.

III. DATASET AND EXPERIMENTAL RESULTS

To assess the suggested method, the experiment is carried out on a balanced set of two categories of 200 RS imagery of size 256×256 each. Among them, total of 120 images, 60 from each class, are used for training phase and 40 images, 20 man-made and 20 natural scene images, are used for testing as shown in Table 1. Remaining 40 images, 20 man-made and 20 naturals, are used for validation purpose. For creating the data set, we used selected parts (256×256 pixels) of sceneries from Massachusetts buildings dataset [31], Massachusetts roads dataset [31]. For each image in the data set, a ground truth (buildings/roads) map was labelled manually.

A PC with Intel (R) Core(TM) i5-4590 CPU at 3.30 GHz and 4GB RAM is used for training and testing of the proposed method. The method is realized in MATLAB. Few samples feature vector of man-made and natural class images (wavelet compressed Gabor energy features) of training and testing sets are tabulated in Table 4 and Table 5. We set the parameters of the Gabor filter bank as follows.

- The No. of scales: 6
- The No. of orientations: 8
- The No. of rows and columns: 39
- The factors of down-sampling along the rows and along the columns: 4

It has been seen that Gabor energy maps for man-made structures tend to have dominant orientation features through Gabor filtering.

Table 1. Number of samples per class for the training, validation and test set

	Data set (200 images)		
Land Cover	Training	Testing	Validation
Man-made structures	60	20	20
Non-Manmade	60	20	20

A. Qualitative Evaluation

For qualitative evaluation, proposed method is applied for extracting two types of manmade objects, i.e., building and roads from the images in databases. To demonstrate the effectiveness of the suggested method, object extraction results are shown here.

1) Road Network Detection

For assessing the effectiveness of the proposed technique for road object extraction, the qualitative segmentation results are presented on the different images from Massachusetts road dataset in Fig.8. Roads were mostly homogeneous and not disturbed by shadows or occlusions. The region surrounds man-made structures (roads)

with near-green colors. The final result overlaid on the initial image and the ground truth data as shown in Fig.8. The algorithm is successfully detecting the whole road network with high accuracy.

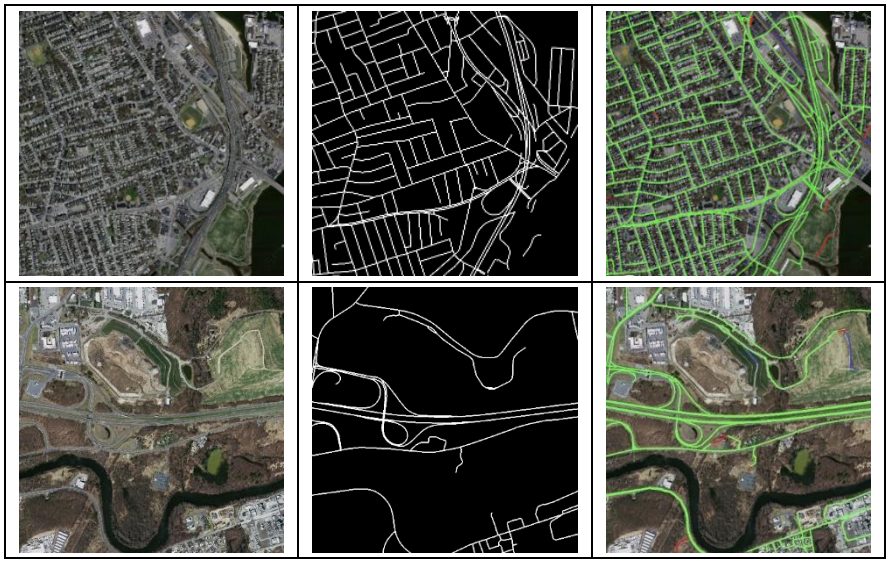


Fig. 8. Road extraction Results from Massachusetts Road Dataset achieved by proposed method. The First and second column depict the input images and their corresponding ground truth images. The third column shows the road extraction results. The green, blue, and red colours represent TPs, FPs, and FNs, respectively

2) Building Extraction

For assessing the usefulness of the proposed technique for building object extraction, the qualitative segmentation results are presented on the different images from datasets in Fig.9. The dataset contains all three bands (i.e., R, G, B) and rich information including roads, various buildings, vegetation etc. Fig.9 depicts the building objects extraction results from Massachusetts building dataset achieved by proposed method.

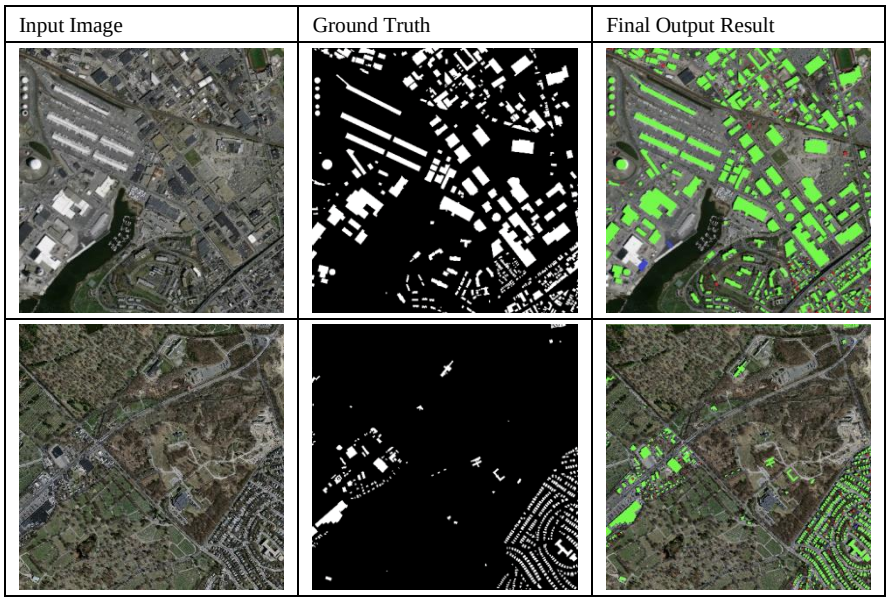


Fig.9. Building Extraction Results from Massachusetts Building Dataset achieved by proposed method. The First and second column depict the input images and their corresponding ground truth images. The third column shows the building extraction results. The green, blue, and red colours represent TPs, FPs, and FNs, respectively

B. Quantitative Evaluation

For assessing the performance of the proposed object extraction method, the following five evaluation metrics are used [32].

$$TPR = \frac{TP}{TP+FN} \tag{7}$$

$$TNR = \frac{TN}{TN+FP} \tag{8}$$

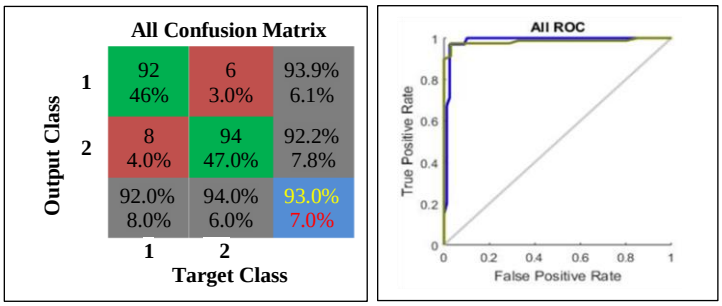
$$FNR = \frac{FN}{TP+FN} \tag{9}$$

$$FPR = \frac{FP}{FP+TN} \tag{10}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

where TP: True Positive; TN: True Negative; FP: False Positive, and FN: False Negative. The Classification Overall Accuracy (OA) is used to measure the rate of images that were correctly classified. The higher the value of an evaluation metric, the better the method's performance.

The confusion matrix and the Receiver Operator Characteristic (ROC) curve are used to assess the proposed method's performance. The confusion matrix and the ROCs is shown in Fig.10. 186 images were correctly identified and 14 images were misclassified among total 200 images by this proposed method as shown in 11(a). The proposed framework resulted in OA of 93%. ROC graph depicted in Fig. 10(b) is a plot of TP rate vs. FP rate. The ROC graph of the suggested model for all data, as shown in Fig.10(b), indicates an exceptional classification between the two classes.



(a) (b)

Fig.10. (a) Confusion matrix and (b) ROC plot of all data

Table 2 lists the OA, TPR, TNR, FNR and FPR of the proposed method of training, testing, validation and all data, respectively. As seen from the tables, the proposed system incurs the acceptable level of performance with the mean values of no less than 93 and 92.55 and 94.1 for overall accuracy, TPR and TNR, respectively. The overall performance analysis is depicted graphically in Fig.11.

Table 2. Performance analysis of Gabor feature of Training, Testing, Validation and all Data

Proposed Gabor Energy Feature and PNN					
	Accuracy	TPR	TNR	FNR	FPR
Training (%)	94.3	92.2	95.4	7.8	4.6
Testing (%)	93.4	93.5	94.9	6.5	5.1
Validation (%)	92.5	92.3	93.8	7.7	6.2
All data (%)	93.0	92.0	94.0	8.0	6.0

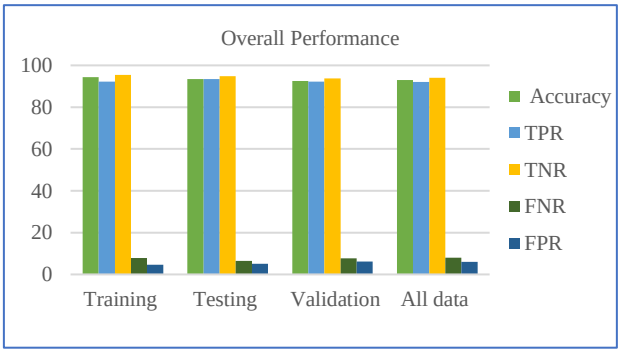
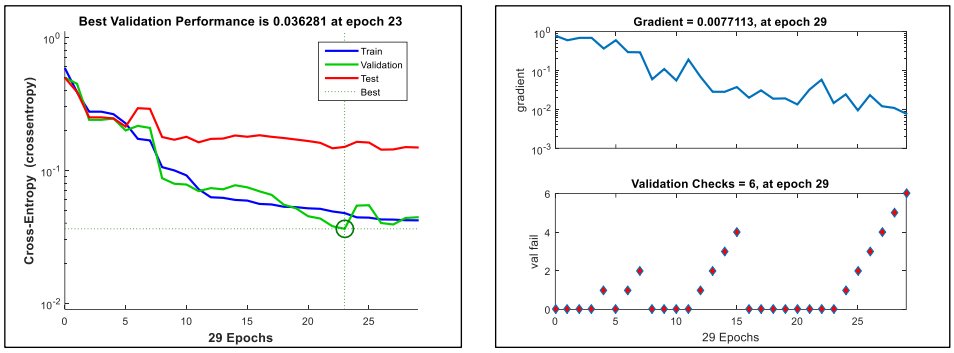


Fig.11. Overall performance analysis of training, testing, validation and all data sets

To evaluate the quality of the network, cross-entropy (CE) is used. Validation performance, based on the cross-entropy error is shown in Fig.12(a). Minimizing cross-entropy results in good classification. After iteration 29, at performance 0.0420, the training was stopped. Cross-entropy is minimized for good classification, as seen in the performance graph in Fig. 12. Prior to epoch 29, the best validation performance was 0.036 at epoch 23, which is worse than the final 0.0420. Fig.12(b) depicts the dynamics of the neural network training state in terms of cross-entropy gradient on a logarithmic scale. The gradient at the endpoint was 7.7113×10^{-3}, which is a reasonable figure to stop at for this set of data.



(a) (b)

Fig.12. a) Best Validation Performance based on Cross-Entropy. Lower values are better. Zero means no error. b) Neural Network Training State

IV. CONCLUSION

This paper proposes a technique which enables the user to extract man-made objects e.g., buildings, roads from input RS imagery without much of user interaction. We also validate the proposed algorithm by numerical results of real RS images. The generality of the proposed algorithm is demonstrated by testing the algorithm with different images. The scope for the present kind of work in the photogrammetry and computer vision field is everlasting and the demand for such work is always increasing one. The experimental results show that the proposed algorithm is practical and reliable. The proposed approach is computationally significantly more efficient than state-of-the-art algorithms while attaining superior performance. Most notably, they have a far smaller number of input parameters, which makes them more practical in practice for not only object extraction, but for other applications such as building density estimation, urban environmental monitoring, socioeconomic parameter analysis, etc.

Although our results are encouraging, the proposed method can be improved further by fusing deep features with structural ones in future studies. Recently, the uses of deep learning methods are taking off to extract objects from RS data. In the future work, we will use graphics processing unit to accelerate the feature learning process. Moreover, the proposed approach may not maintain the same classification percentage consistently because of its sensitivity to contrast. It would be promising to integrate texture features with other discriminative features e.g. spectral and radiometric characteristics, geometrical and contextual information to classify man-mad object and natural objects. Future developments are also related to more effective object selection technique so that it reduces classification mistake, which is caused by the comparable spectral components of buildings and other areas like roads and grounds.

References

[1] Md. Abdul Alim Sheikh, Tanmoy Maity, Alok Kole "IRU-Net: An Efficient End-to-End Network for Automatic Building Extraction from Remote Sensing Images," IEEE Access, vol. 10, pp. 37811-37828, 2022. DOI:10.1109/ACCESS.2022.3164401

[2] Y. Liu, Z. Li, B. Wei, X. Li, and B. Fu, "Seismic vulnerability assessment at urban scale using data mining and GIScience technology: Application to urumqi (China)," *Geomatics, Natural Hazards Risk*, vol. 10, no. 1, pp. 958-985, Jan. 2019.

[3] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathiern, and P. Vateekul, "Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning," *Remote Sens.*, vol. 11, no. 1, p. 83, 2019.

[4] Abdollahi, A., Pradhan, B., Gite, S., & Alamri, A., "Building footprint extraction from high resolution aerial images using generative adversarial network (GAN) architecture," *IEEE Access*, Article 209517-209527,2020. http://dx.doi.org/10.1109/ACCESS.2020.3038225.

[5] S. E. Park, Y. Yamaguchi, and D. J. Kim, "Polarimetric SAR remote sensing of the 2011 Tohoku earthquake using ALOS/PALSAR," *Remote. Sens. Environ.*, vol. 132**,** pp. 212-220, 2013.

[6] X. H. Tong, Z. H. Hong, S. J. Liu, X. Zhang, H. Xie, Z. Y. Li, S. L. Yang, W. A. Wang, and F. Bao, Building-damage detection using pre- and post-seismic high-resolution satellite stereo imagery: A case study of the May 2008 Wenchuan earthquake, *ISPRS J. Photogramm. Remote Sens.*, vol. 68, pp.13-27, 2012.

[7] Wenzao Shi, Z. Mao & Jinqing Liu, "Building area extraction from the high spatial resolution remote sensing imagery," Earth Science informatics, (2018) https://doi.org/10.1007/s12145-018-0355-5.
[8] Licun Zhou, Guo Cao, Yupeng Li, and Yanfeng Shang, "Change detection based on conditional random field with region connection constraints in high-resolution remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 8, pp. 3478-3488, 2016.
[9] Wenjin Wu, Huadong Guo, and Xinwu Li, "Urban Area SAR Image Man-Made Target Extraction Based on the Product Model and the Time–Frequency Analysis," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 3, pp. 943-952, 2015.
[10] K. Bittner, F. Adam, S. Cui, M. K¨orner, and P. Reinartz, "Building footprint extraction from VHR remote sensing images combined with normalized DSMs using fused fully convolutional networks," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 8, pp. 2615–2629, 2018.
[11] S. Wang, X. Hou, and X. Zhao, "Automatic Building Extraction from High-Resolution Aerial Imagery via Fully Convolutional Encoder-Decoder Network with Non-Local Block," IEEE Access, vol. 8, pp.7313–7322, 2020.DOI: 10.1109/ACCESS.2020.2964043
[12] Z. Li, W. Shi, Q. Wang, and Z. Miao, "Extracting man-made objects from high spatial resolution remote sensing images via fast level set evolutions," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 2, pp. 883–899, 2014.
[13] M. A. A. Sheikh, A novel self-assessed approach for classification of manmade objects and natural scene images from aerial images, 2011 Annual IEEE India Conference, 2011, pp. 1-7, doi: 10.1109/INDCON.2011.6139328.
[14] Abdul Alim Sheikh, S. Mukhopadhyay, Noise Tolerant Classification of Aerial Images into Manmade Structures and Natural- Scene Images based on Statistical Dispersion Measures, 2012 Annual IEEE Conference (INDICON), 2012, pp. 653-658 DOI: 10.1109/INDCON.2012.6420699
[15] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65**,** pp. 2-16, 2010.
[16] I. Sebari and D. C. He, "Automatic fuzzy object-based analysis of VHSR images for urban objects extraction," *ISPRS J.Photogramm. Remote Sens.*, vol. 79, pp. 171-184, 2013.
[17] K. Karantzalos and D. Argialas, A Region-based Level Set Segmentation for Automatic Detection of Man-made Objects from Aerial and Satellite Images, *Photogramm. Eng. Remote Sens.*, 75(6) (2009) 667-677.
[18] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834-848, Apr. 2018.
[19] Yaohui Liu, Jie Zhou, Wenhua Qi, Xiaoli Li, Lutz Gross, Qi Shao, Zhengguang Zhao, Li Ni, Xiwei Fan, and Zhiqiang Li "ARC-Net: An Efficient Network for Building Extraction from High-Resolution Aerial Images," IEEE Access, vol. 8, pp.154997-155010, 2020, *DOI: 10.1109/ACCESS.2020.3015701*
[20] N. L. Gavankar and S. K. Ghosh, "Automatic building footprint extraction from high-resolution satellite image using mathematical morphology," European Journal of Remote Sensing, vol. 51, no. 1, pp. 182–193, 2018.
[21] W. Li, C. He, J. Fang, J. Zheng, H. Fu, and L. Yu, "Semantic segmentation based building footprint extraction using very high-resolution satellite images and multi-source GIS data," *Remote Sens.*, vol. 11, no. 4, p. 403, 2019.
[22] L. Zhang, L. Zhang, and B. Du 'Deep learning for remote sensing data: A technical tutorial on the state of the art', IEEE Geoscience and Remote Sensing Magazine, 4(2) (2016) 22–40.
[23] Alshehhi, R., Marpu, P. R., Woon, W. L., and Mura, M. D. (2017) 'Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks', *ISPRS Journal of Photogrammetry and Remote Sensing, 130*:139-149. https://doi.org/10.1016/j.isprsjprs.2017.05.002.
[24] S. Saito, T. Yamashita, and Y. Aoki, "Multiple object extraction from aerial imagery with Convolutional Neural Networks," *Electron. Imag.*, vol. 60, no. 1, pp.1-9, 2016.
[25] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang, "Classification for high resolution remote sensing imagery using a fully convolutional network," *Remote Sens.*, vol. 9, no. 5, pp. 498-519, 2017.
[26] Sukhendu Das, T. T. Mirnalinee, and Koshy Varghese, "Use of Salient F eatures for the Design of a Multistage Framework to Extract Roads from High-Resolution Multispectral Satellite Images," IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no.10, pp. 3906-3931, 2011.
[27] R.C Gonzalez and R.E Woods (2018) Digital Image Processing. Fourth Edition (Pearson, ISBN: 978-0133356724).
[28] Kamarainen, J.-K.; Kyrki, V.; Kalviainen, H., "Invariance properties of gabor filter-based features-overview and applications," *IEEE Trans. Image Proc.* vol. 15, pp. 1088–1099, 2006.
[29] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 1, pp. 55-73**,** 1990.
[30] Ali Ozgun Ok, Caglar Senaras, and Baris Yuksel, "Automated Detection of Arbitrarily Shaped Buildings in Complex Environments from Monocular VHR Optical Satellite Imagery," IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 3, pp. 1701-1717, 2013.
[31] V. Mnih, "Machine learning for aerial image labeling" Ph.D. dissertation, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2013.
[32] Md. Abdul Alim Sheikh, Alok Kole, Tanmoy Maity, "A Multi-level Approach for Change Detection of Buildings using Satellite Imagery," International Journal of Artificial Intelligence Tools, vol. 27, no. 8, p. 1850031, 2018, DOI: 10.1142/S0218213018500318.