Road extraction in RGB images acquired by Low Altitude Remote Sensing from an Unmanned Aerial Vehicle: A Neural Network Based Approach

Ramesh K N[1],Yogitha A N[2]
[1]Department of Electronics and Instrumentation, Bangalore Institute of Technology, Bangalore
[2]Department of Information Technology, National Institute of Technology Karnataka,Surathkal

Ravi V M[3],S N Omkar[3],M B Meenavathi[1]
[3]Department of Aerospace Engineering, Indian Institute of Science,Bangalore

Keywords: UAV,Extreme learning machine,Spatial methods, Morphological close.

ABSTRACT

The high growth rate of urban population has led to an increase in the demand for better urban planning and monitoring which mainly includes road network development. Manual monitoring of road development is time consuming and inefficient. In this paper, we propose a method for automatic extraction of roads in vision spectrum (RGB) images acquired by remote sensing from a UAV also known as Low Altitude Remote Sensing (LARS) or Near Earth Remoste Sensing . Extreme Learning Machine (ELM), a neural networks based classier is used for spectral classication. Spectral classication is further improved by applying spatial techniques. The spatial techniques include a combination of Shape Index(SI), Density Index(DI) and mathematical morphological close operations. Seven images of diverse road stretches are analyzed to verify the robustness of the proposed method. The classification results are analysed using confusion matrix. The performance parameters derived from confusion matrix are analyzed for a range of hidden neurons of the ELM model and an optimum number of hidden neurons are chosen. Successful road extraction demonstrates the potential of using UAV imagery for monitoring road development.

I. INTRODUCTION

Extraction of information from images acquired from higher altitudes was always a challenging task. Having Precise and updated road network information plays a vital role for geographic information system (GIS) databases, transportation, urban planning, automated road navigation, and emergency response applications (Rajeswari et al. 2011). In the current practices the road network information is acquired using physical surveys which are very tedious, costly and highly inefficient in terms of time and money. Automated approaches use satellites imagery for road network extraction, these approaches are use data at a macro level such as a county or a city scale application to obtain road network information. With the increase in road network infrastructure there is a requirement to monitor the road infrastructure progress which requires detailed information at a micro level such as road dimensions, composition of roads and pavement under construction. Satellite based systems do not provide these micro level details, however Uninhabited/Unmanned Aerial Vehicle (UAV) provide an interesting and promising option.

Remote sensing form UAV also known as Low Altitude Remote Sensing (LARS) allows the acquisition of high spatial resolution images in the vision spectrum (RGB) bands using light weight and cost effective optical sensors which are typical of UAV payloads category. Another interesting microlevel application of LARS in road extraction is to determine the extent of roads and pathways in agriculture fields. The number of fragments have a direct impact on the cost of agricultural production An addition of one fragment is estimated to reduce the output by 2 to 10 percentage points (Deininger et al. 2017). In the past, several groups of researchers have carried out road extraction and extraction of linear structures on images that are acquired using LARS. Extraction is carried out using several methods such as use of semi-automated techniques that address the problem of delineation (Wiedemann et al. 1998) , geometrically constrained template matching (Doucette et al. 2001) and texture analysis and graph partitioning (Senthilnath, Rajeshwari, and Omkar 2009). However not all of these methods are fully automatic and need human supervision and user provided clues for identifying feature locations. In addition, many of the papers on the road extraction make use of secondary cartographic information which is not the case in the proposed approach (Teo, Logenthiran, and Woo 2015). From Literature it can be observed that the current Road Extraction classification methods predominantly use detection of linear structures this however leads to identification of other linear structure such as fences and powerlines. Machine Learning and Neural Network based approaches have noteworthy success in image classification and object detection. Feedforward neural networks have been widely used to perform image classification by their ability to approximate and map complex nonlinear distributions directly from the input samples. in the context of image classification the nonlinear spectral distribution is addressed but also to provide models for enormous class of natural and articial phenomena that are problematic to handle using traditional parametric techniques; provides more detailed information on classification without approximations (Zhou et al. 2015).

The feed forward neural networks parameters are tuned iteratively using gradient-based learning algorithm as the training function, in the past this often resulted as a bottleneck for using them in applications because of the slow learning speeds. This paper proposes Extreme Learning Machine (ELM), a single-hidden layer feedforward neural (SLFNs) network. Unlike the conventional feedforward network learning algorithms like Back-Propagation (BP) algorithm, the learning speed the ELM can potentially be thousands of times faster, while obtaining better overall performance. (Huang, Q. Zhu, and C.K. Siew 2006) ELMs ability to handle large volumes of data and quick learning speed without the hassle of back propagation, over fitting and multiple iterations make it a better classifier than most of the traditional classifiers (Huang, K. Zhu, and C.K Siew 2004). In this paper we propose the ELM classifier for classification. Further, the classification is improved by using spatial techniques. The spatial techniques are Shape Index(SI), Density Index(DI) and mathematical morphological close operations. The road extraction performance is analysed using confusion matrix. This paper is organized as follows Section II explains the methodology undertaken. Performance evaluation of the proposed algorithm are discussed in Section III followed by Section IV conclusion.

II. METHODOLOGY

The proposed road extraction methodology comprises of three steps $S1$ to $S3$ as shown in Figure 1.Two class classification is performed i.e., road and non-road regions in step 1. The classification is a supervised per pixel classification carried out on the image using Extreme Learning Machine(ELM). Extraction is further improvised by eliminating

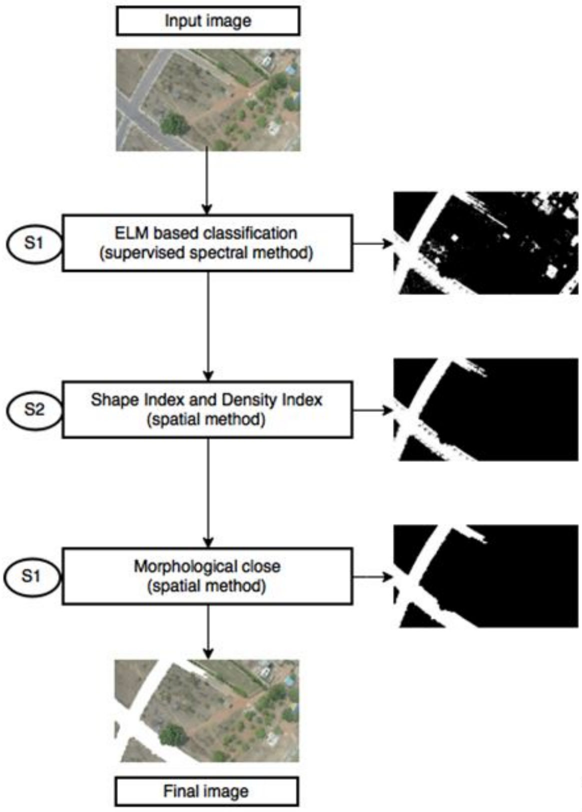

Figure 1: Road Extraction Steps

the non-road regions using spatial methods in step 2 and step 3.The spatial methods used are geometric operations of Shape and Density index (Ramesh et al. 2015) in step 2 and morphological close in step 3 .

A. Data Acquisition

The roads image data-set is acquired through UAV remote sensing. A Fixed wing modular UAV fitted with a Go Pro camera is flown over the land cover at an altitude of 100m and the video is recorded. Image frames containing

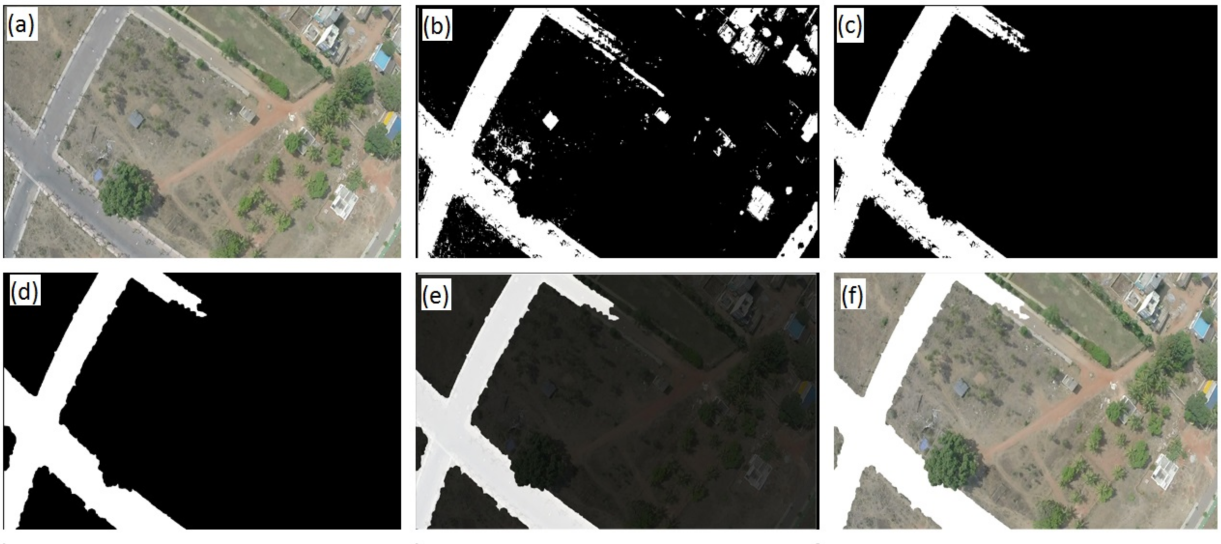

Figure 2: (a) Original Image (b) ELM classified Image (c) Shape Index and Density Index (d) Morphological Close (e),(f) Final image of the extracted road

roads are extracted from the video. These extracted images are used for processing. It can be observed from Fig. 2(a). that the image is acquired at an oblique aerial orientation as opposed to a nadir orientation, this is to ensure that there is a larger coverage of roads in the image.

B. Extreme learning machines in road extraction

An Extreme Learning Machine (ELM) is a single-hidden layer feed forward neural (SLFNs) network with a very fast learning algorithm invented by (Huang, Q. Zhu, and C.K. Siew 2006)

In general, ELM models have only one hidden layer with parameters of input weights w and biases b of hidden neurons $(1, 2, ...L)$. Hidden neuron parameters w, b are assigned randomly and the output weights β are determined analytically using the Moore-Penrose generalized inverse H^{ψ} [9]. All hidden neuron parameters are independent of the target function y and the training data G .

In a training data set of G distinct samples (x_i, y_i) where $[x_{i1}, x_{i2}, ..., x_{in}]^T \in R^n$ where $R^n = R^3$ which is RGB input of every pixel of the image to the model and $[y_{i1}, y_{i2}, ..., y_{im}]^T \in R^m$ where $R^m = R^2$ and the target function $y_{n,m} \in \{0, 1\}$ (i.e., road and non-road) , such that $(x_i, y_i) \in R^n \times R^m$ where $(i = 1, 2, ..., G)$ where G is the number of samples. In our model we use $L = 10$ hidden neurons and sigmoid activation function $g(x) = \frac{1}{1+e^{-\lambda x}}$. The classification of ELM with Lhidden neurons and output function $g(x)$ of ELM is mathematically represented as (Huang, Q. Zhu, and C.K. Siew 2006):

$$f_L(x_j) = \sum_{i=1}^{L} \beta_i g(w_i x_i + b_i), j = 1, ..., G \tag{1}$$

Where,

$$g(w_i x_i + b_i) = \frac{1}{1 + e^{-(w_i x_i + b_i)}} \tag{2}$$

Here,$w_i = [w_{i1}, w_{i2}, ..., w_{in}]^T$ weight vector between input nodes and $i^t h$ hidden nodes; $\beta_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{in}]^T$ weight vector between the output nodes and $i^t h$ hidden nodes. b_i (bias) is the threshold of the $i^t h$hidden neuron. $w_i x_j$ denotes the inner product of w_i and x_j . A standard SLFN with L hidden neurons can approximate$G$ training data with zero mean error. This implying that there exists , w_i, β_i and b_i such that:

$$y_i = \sum_{i=1}^{L} \beta_i g(w_i x_i + b_i) = H\beta \tag{3}$$

where,

$$H = \begin{bmatrix} h(x_i) \\ \vdots \\ h(x_G) \end{bmatrix} = \begin{bmatrix} g(w_1 x_1 + b_1) & \cdots & g(w_L x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1 x_G + b_1) & \cdots & g(w_L x_G + b_L) \end{bmatrix} \tag{4}$$

and β is given by

$$\beta = \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,m} \\ \vdots & \ddots & \vdots \\ \beta_{L,1} & \cdots & \beta_{L,m} \end{bmatrix} \tag{5}$$

The i^{th} column of H is the i^{th} hidden node output with respect to inputs $x_1, x_2, ..., x_G$. The n^{th} row of H is the hidden layer feature mapping with respect to the n^{th} input x_n . The output weight vector β can be calculated by:

$$\beta = H^{\psi} Y \tag{6}$$

Where H the Moore-Penrose pseudo inverse Courrieu 2008 of the hidden layer output matrix and Y is given by:

$$Y = \begin{bmatrix} y_{1,1} & \cdots & y_{1,m} \\ \vdots & \ddots & \vdots \\ y_{G,1} & \cdots & y_{G,m} \end{bmatrix} \tag{7}$$

As compared to traditional feed forward or Back Propagation ELM networks for training the ELM model consists of only three steps Huang, K. Zhu, and C.K Siew 2004:

1) Assign arbitrary input weight w_i and bias $b_i, i = 1, 2, ..L$
2) Generate the random hidden layer weight matrix H .
3) Calculate the output weights $\beta = H^{\psi} Y$

The final classification result for an unseen case is given by (Malik and Mishra 2015) :

$$f(x_{unknown}) = argmax f_j(x_{unknown}), j \in 1, ..., m \tag{8}$$

where $f(x_{unknown})$ is the predicted label of an unseen case. The labels are obtained after getting the output weight matrix β after training the ELM model. ELM output for Figure 2(a) can be seen in Figure 2(b).

C. Spatial methods

1) Shape index (SI) and Density Index (DI): The output from extreme learning machine assigns two class labels the to the image (road and non-road) as seen in Figure 3, however some of the non-road regions are falsely classified as road regions by ELM due to spectral similarity. To eliminate these regions Shape and Density Index Rajeswari et al. 2011 are used. In the paper, there are two scenarios to be considered (i) Images with a single road (ii) Images with multiple roads. Density Index (DI) is given by

$$DI = \frac{\sqrt{A}}{1 + \upsilon} \tag{9}$$

A represents the area of the object υ is given by

$$\upsilon = \sqrt{Var(X) + Var(Y)} \tag{10}$$

$Var(X)$ represents the variance of x-coordinates of all pixels and $Var(Y)$ represents the variance of y-coordinates of the region ,approximates the radius of the region. Shape index(SI) is a geometrical parameter given by:

$$SI = \frac{P}{4 * \sqrt{A}} \tag{11}$$

where P represents the perimeter of the object (i.e. the number of pixels on the boundary of the object) . A

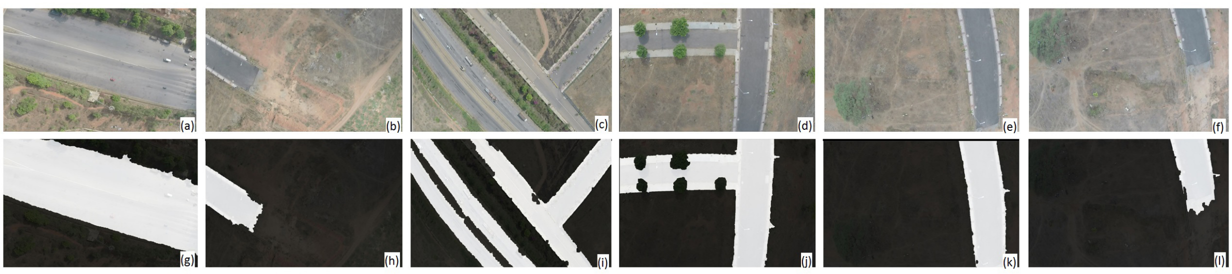

Figure 3: (a) Original Image-1 (b) Original Image-2 (c) Original Image-3 (d) Original Image-4 (e) Original Image-5 (f) Original Image-6 (g) Final ELM Classified Image-1 (h) Final ELM Classified Image-2 (i) Final ELM Classified Image-3 (j) Final ELM Classified Image-4 (k) Final ELM Classified Image-5 (l) Final ELM Classified Image-6

high DI indicated large area objects and a high SI indicates line segments. Roads normally occupy lesser area indicating that they have a lower DI. Since they are relatively long line segments compared to the non-road regions they have high SI. A white pixel indicates a road region and a non-white pixel indicates a non-road region. In our observation, we have found that the road regions have a SI of range between 1.3-3.9 and DI of range between 1-2.2 . Result of application of SI and DI on Figure. 2(b) is shown in Figure (c) and a SI value chosen is 3.84 and DI value chosen is 1.38 for Figure 2(b).

Further, to improve the results we apply morphological operations the processed image and is shown in Figure 2(d).

2) Morphological Close: Morphological closing of A by B is denoted as $A.B$, is a process of dilation followed by erosion. Dilation is an operation that ”grows” objects in a binary image and Erosion ”shrinks” objects in a binary image. The extent and manner of growing or shrinking of object is controlled by structuring element B. Morphological closing of a geometric structure is mathematically defined as

$$A.B = (A \oplus B) \ominus B \tag{12}$$

Here, A represents the image obtained after applying SI and DI operations on the binary image (i.e. ELM output). Structuring element B chosen is a flat disk shaped structure with radius 10 and 6 periodic line structuring elements. It can be observed that after the application of close operation small holes are removed. Further, the extracted roads are super imposed on the original image final ELM classified image is shown in 2(d), 2(e).

III. PERFORMANCE EVALUATION

Confusion matrix is used to evaluate the performance of classification. The parameters computed are True positive (TP),True Negative (TN), False Positive (FP) and

TABLE I: Confusion Matrix

Predicted (expected)	Ground truth (actual)		
		Number of non-road pixels	Number of road pixels
	Number of non-road pixels	1721067 (TN)	11307(FN)
	Number of road pixels	35772 (FN)	305454 (TP)

TABLE II: Evaluation features

Features	**Results**
Precision	96.43%
Recall	89.53%
Accuracy	97.73%
MCC	0.92

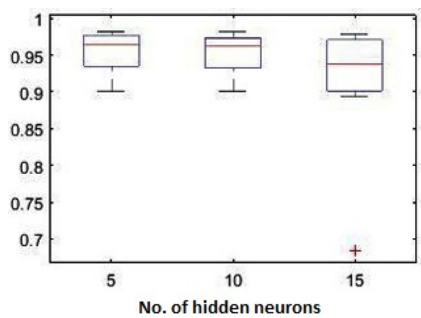

Figure 4: Precision

False Negative(FN). The following evaluation features are obtained. The confusion matrix for Figure 2 is tabulated in Table I. And evaluation features obtained from the confusion matrix for Figure 2 is tabulated in Table II.

Precision (PPV) describes the fraction of retrieved instances that are relevant .Recall (TPR) also know as sensitivity, is the fraction of relevant instances. Accuracy (ACC), is the degree of closeness of predicted values to the original values. Matthews correlation coefficient (MCC) , is a measure of quality of binary classification which takes into account true and false positives and negatives and is regarded as a balanced measure. The value

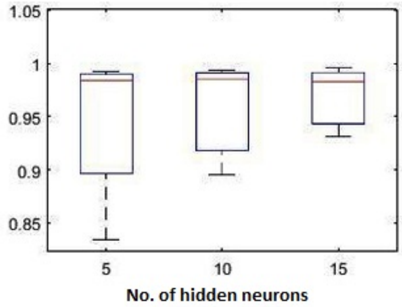

Figure 5: Recall

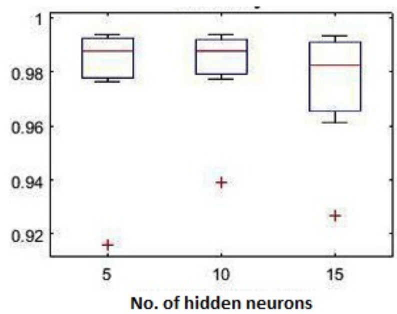

Figure 6: Accuracy

returned by MCC lies between -1 and +1 . A coefficient of 1 represents a perfect prediction. The values returned by precision, recall and accuracy lie between 0 to 100%.Values of evaluation features being close to 100% signifies that a good classification has been carried out. This is also true in the observed results.

The images and their extracted output images are shown in Figure 3 and Figure 8 shows precision and recall performance of the seven images for a varying number of hidden neurons. With 5 hidden neurons, precision and recall is greater than 85% for 5 images however recall is less than 90% for two images i.e, Figure 2(a) and Figure 3(c). In the case of 15 hidden neurons both precision and recall is more than 90% for four images, however precision is 68% for Figure 3(b). and 88% for Figure 3(c). In the case of 10 neurons, the precision recall performance is consistently close to or greater than 90% for all the seven images. Hence 10 hidden neurons was chosen for the ELM model.

The performance parameters for various hidden neurons are represented using box-plots McGill, Tukey, and Larsen 1978 as shown in Figure 4,5,6 and 7. The optimized performance can be observed for 10 hidden neurons.

IV. CONCLUSION

This paper proposes a novel approach of using Machine leaning techniques, for road extraction. The images that are analyzed were acquired from an UAV flown at an altitude of 100 metres. At this height spatial information in the RGB images provides micro level (sub decimeter) resolution. Sub decimeter satellite images are currently unavailable to researchers in public domain. Machine learning methods used in this work is unlike the traditional methods that detect roads as linear structures. The use of Machine learning for road classification which is independent of the traditional methods of detecting roads as linear structures is presented. The robustness of the

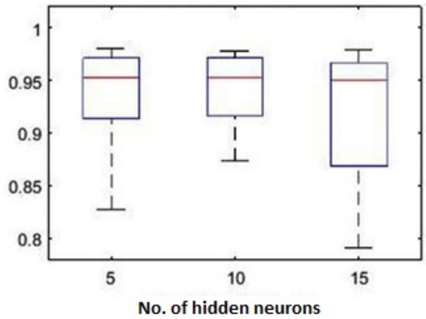

Figure 7: Matthews Correlation Coefficient

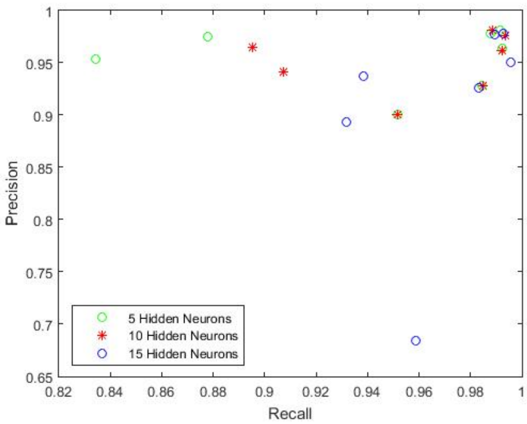

Figure 8: Precision-Recall graph

proposed method is demonstrated by extracting roads from diverse road stretches in multiple images. Successful extraction of roads in images acquired using LARS demonstrate the potential of using UAV imagery for microlevel applications such as road construction monitoring and and determining the actual areas available for cultivation in fragmented farms.

V. FUTURE WORKS

This work can be EXTENDED for multi-class classification to classify multiple road types. Further, ELM can be applied to classify using texture features to improve classification.

ACKNOWLEDGMENT

The authors would like to thank members of UAV Lab, Aerospace Department, Indian Institute of Science for providing the UAV images for analysis.

REFERENCES

Courrieu, P (2008). "Fast computation of Moore-Penrose inverse matrices". In: *arXiv preprint arXiv:0804.4809*.

Deininger, K et al. (2017). " Does Land Fragmentation Increase the Cost of Cultivation ? Evidence from India." In: *The Journal of Development Studies* 53(1), pp. 82–98.

Doucette, P et al. (2001). "Self-organised clustering for road extraction in classified imagery". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 55(5), pp. 347–358.

Huang, G.B, K.Y Zhu, and C.K Siew (2004). "Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks." In: *International Joint Conference on Neural Networks, (Budapest, Hungary)*, pp. 1–5.

Huang, G.B, Q.Y. Zhu, and C.K. Siew (2006). "Extreme learning machine: theory and applications". In: *Neurocomputing* 70(1), pp. 489–501.

Malik, H and S Mishra (2015). "Extreme learning machine based fault diagnosis of power transformer using IEC TC10 and its related data." In: *Annual IEEE India Conference (INDICON), IEEE*, pp. 1–5.

McGill, R, J.W Tukey, and W.A. Larsen (1978). " Variations of box plots". In: *The American Statistician* 32(1), pp. 12–16.

Rajeswari, M et al. (2011). "Automatic road extraction using high resolution satellite images based on Level Set and Mean Shift methods". In: *In Electronics Computer Technology (ICECT), 2011 3rd International Conference on, IEEE*. Vol. 2, pp. 424–428.

Ramesh, K.N et al. (2015). "Automatic detection of powerlines in UAV remote sensed images". In: *International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), IEEE*, pp. 17–21.

Senthilnath, J, M Rajeshwari, and S.N. Omkar (2009). "Automatic road extraction using high resolution satellite image based on texture progressive analysis and normalized cut method". In: *Journal of the Indian Society of Remote Sensing* 37(3), pp. 351–361.

Teo, T.T, T Logenthiran, and W.L Woo (2015). "Forecasting of photovoltaic power using extreme learning machine". In: *In Smart Grid Technologies-Asia (ISGT ASIA), IEEE*, pp. 1–6.

Wiedemann, C. et al. (1998). "Automatic extraction and evaluation of road networks from MOMS-2P imagery". In: *International Archives of Photogrammetry and Remote Sensing* 32, pp. 285–391.

Zhou H ang Kong, H et al. (2015). "Efficient road detection and tracking for unmanned aerial vehicle". In: *IEEE Transactions on Intelligent Transportation Systems* 6(1), pp. 297–309.