March-April 2020
ISSN: 0193-4120 Page No. 1488 - 1494

A Combined Use of SIFT-CNN Method for Ship Detection in Optical Satellite Imagery

[1]Gurram Midhila, [2]C.M. Sheela Rani

[1]M.TECH Student, [2]Professor, Koneru Lakshmaiah Education Foundation (KLEF), Vaddeswaram, Guntur District, Andhra Pradesh-522502, India

[1]midhilaramanaiah123@gmail.com, [2]sheelarani_cse@kluniversity.in

Article Info
Volume 83
Page Number: 1488 - 1494
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 15 March 2020*

Abstract

Optical satellite imagery have been extensively used in marine environmentology, naval surveillance and sea border activities. Ship identification is a vital aspect of optical satellite imagery in harbor dynamic reconnaissance and oceanic administration. Prominent ship detection techniques are bounded by their time complexity and execution accuracy. This paper proposes a strategy of scale invariant feature transform (SIFT) algorithm and a convolution neural network (CNN) which is a class of deep learning. In this paper, a hierarchical method to amalgamate the feature extraction, efficient object masking and an accurate object identification model is proposed where feature extraction is done by applying SIFT algorithm and object detection by CNN. First, in the feature extraction phase SIFT algorithm is applied where scale space extrema detection, key point localization, orientation assignment, key point descriptor are considered specifically to improve the robustness regarding detection, patented local feature detector and image masking. Finally for the object detection, these sub-models are integrated and output is given to convolution neural network. Ship detection is done by bounding boxes. By implementing CNN-SIFT technique, 97% of accuracy is achieved.

Keywords: *SIFT-CNN, Feature Extraction, Masking of ships, Ship Detection, Optical Satellite Imagery*

1. Introduction

Using of satellite data for feature extraction and ship identification have become very important requirement in the present day image processing techniques. Many advanced technologies are relaying on image processing methods for their functionality. Basically security related techologies where we need to identify a ship and classify various ships into groups we have a great requirement of accuracy. For feature extraction techniques like SIFT can be used.Scale Invariant Feature Transform, commonly known as SIFT was presented by David G. It is a merger of keypoint extractor and descriptor. It is invariance to image scale and image rotation. The working of SIFT algorithm takes place in 4 steps. The SIFT algorithm will workout in different levels initially the corners are detected in the images and this is names as Scale-space Extrema Detection here state space filtering is used for the the detection of the corners and accordingly the window size is decided for detection of corners.Then we will go for the Keypoint Localization, here we will finalize the potential keypoint locations and then will refine them for more accurate results by using taylor series and here the need of the contrast threshold comes up, any pixel intensity whose level is 0.03 then it will be rejected and only the pixel levels greater than 0.03 are considered. The next step will be of Orientation Assignment which is necessary to eliminate any errors in image orientation, here we will orientation for every key point identified in previous step by considering the neighborhood pixels. Then we will go for Keypoint Descriptor where we will take a region around the keypoint and will partition it into sub squares and make a histogram for each sub square. A vector is created from these histograms called as keypoint descriptor and finally these keypoints are matches with remaining neighbors and is called as Keypoint Matching. here we will consider the ratio of distance between the similar

March-April 2020
ISSN: 0193-4120 Page No. 1488 - 1494

neighborhood and will consider only those neighbors whose distance is greater than a minimum threshold value of our algorithm.

In CNN First stage is feature learning and the second one is classification. CNN is a part of deep learning algorithm in which we will give a image as input to the algorithm then some weights are assigned to different objects in the image so as to differentiate the objects from one another. These weights are assigned based upon the biases. Here for all this to happen we need to do some preprocessing to the image called as ConvNet which is necessary for the feature extraction but here in our model we have used SIFT algorithm for the feature extraction so we don't need to have ConvNet for our proposed model. Once the features are extracted the first stage in CNN is convolution and this is done by using a kernel. The size of the kernel is decided depending upon the requirement and for efficient output. Once the image is convolved using the kernel then the Pooling stage comes up in which the spatial size is reduced in the convolved input image which is necessary to reduce the computational requirements. Then we will implement valid padding on the image this is done to increase the dimensionality of the convolved feature. Then this layer is provided for the fully connected network for the classification of the image.

Here we presented a technique to extract and detect features from the optical satellite data. For extraction and detection here we have used SIFT and convolutional neural network techniques. Once the features are extracted they are classified based on CNN technique. Here to verify the proposed technique an Optical Satellite image of a sea shore is considered which is having ships in the sea. This image is provided as input to the proposed system and after processing we can observe that 97% of the features that is the ships in the images were detected by using the proposed method. The language used here for the development of the proposed technique is Python with a front end navigator of Anaconda Navigator – Spyder.

2. Literature Survey

Image Picture based characterization of ships was finished by numerous specialists around the globe. A significant part of the ongoing exploration work was centered around identifying different ships.Identifying ships from satellite data was very scarce. [4] Zambanini et al., highlights about the use of computer vision is an interdisciplinary logical field that manages how computers can be made to increase significant level comprehension from computerized pictures or recordings. They recognized five domains in which computer vision can help the image processing. computer visioncan be applied onimage processing in different zones they are identification of objects, segmentation of different objects and classification of objects. Recognizing Ships by [8] Sebastian Zambanini et al., Their research mostly focused on detection and recognition of Ships from images.Classification of ships by Y. Xiang et al., [6] by utilizing the local feature matching techniques like SIFT and SIFT flow. Next in the another attempt to identify the ships based on the motifs Bag of Visual Words (BoVW) method was utilized. Maia Zaharieva et al., [7] executed tests on various ship pictures to perceive the ships excluding the other objects. Classification algorithm and the feature extraction methods didn't perform well for classification of different ships. F. Zhang and H. Lang et al., in [5] used (BoVW) for local feature detection. They achieved an accuracy of 72.7% for one training sample per class. Maia Zaharieva et al., [9] given an overviewof experimental results that are obtained for an automatic prediction of small shipsbased on the Bayesian decision theory. Keyur D. Joshi et al. Machine vision system was used by [3] to identify and count the fish eggs.Segmentation, and classification made it possible for them to recognize and count the eggs.Guangzhi Shi et al., [2] structured a target recognition framework using SVM and ANN.Artificial Neural Networks (ANN) for classication of the ships.M. Kartalet al., proposed an alternate methodology by utilizing deep learning in [10]so as to perceive the ships. Applying the deep neural network to perceive the ships from satellite images. The accuracy that achieved here is 71.4, 87.6 and 87.4 % for fine, exceptionally fine and incredibly fine reviewed delivers separately. [11] Martin Kampel et al., planned a framework using computer vision algorithms that perceives the ships. To extract candidate regions visual saliency detection is used, a homogeneous filter is introduced to totally remove speculated targets, to characterize ship targets a novel S-HOG descriptor is presented. Also the framework is robust for images with clouds and other disturbances. J.Wang et al., in [12] used Ant Colony Optimization (ACO) to adjust threashold dynamically and also in dealing with image edge detection. Scale invarient target detection approach based on the pose consistency voting used by Y. Lin et al., [13] for manmade targets.This technique is robust to the shape-comparative distractor impedance. Jongpil Kim et al., [14] used novel cascade method that is ship head classification in satellite images using SVM, here the ship direction is also specified. In order to remove false alarms of sea areas context information is brought up. D. Zhang et al., proposed a new method for detecting inshore ships. Here Hough Transform method is used and also to increase the efficiency itration is performed.Y. Xiang et al., [16] employed the SIFT methodolgy to identifyships from the dataset that having differentships. [17] Deepthi Sudhaet al., used image processing techniques, edge detection methods and false detection methods for currency. [18] Sandeep Dwarkanath Pandeet al.,CNN was compared with SVM. Here CNN generalization ability is tested and results show CNN performs better than SVM.

3. Features Extraction & Matching

A. Features:

Features assume a significant job in distinguishing the image.As opposed to accepting the entire image as a

contribution for the framework, now these features willsupport the framework in order to comprehend the image in a superior manner.It consist particular samples of the picture that are utilized to recognize one picture from different pictures. This process of spotting the features is known as feature detection. Characterizing the features in such a way that the computer can understand is called as\ feature description. By considering the features in image. Better outcomes can be achieved in less time by taking into account of neighborhood features like SIFT and SURF instead of considering the global features like shape based and texture-based features.

B. SIFT:

Scale Invariant Feature Transform known as SIFT was presented by David G. Lowe [17], which is a mergerof keypoint extractor and descriptor. It is invariance to image scale and rotation. The SIFT algorithm process carries out in 4 different steps. The very first step is, it estimates a scale space by using the Difference of Gaussian where the blurring can be done. Then At that point next a key point localization is performed, where the key core interestsare limited and cleared by wiping out the lowest differentiation focuses just as that of edge focuses. Then coming to the third step a key point orientation assignment which is dependent on local image gradient is done. Finally a descriptor generator to enumerate each key point dependent on picture size and direction. The SIFT algorithm will workout in different levels initially the corners are detected in the images and this is names as Scale-space Extrema Detection here state space filtering is used for the detection of the corners and accordingly the window size is decided for detection of corners. The coefficients are calculated by taking the below equations.

$$LL_{i,j}=\left|\frac{(OR_{2i,2j}+OR_{2i+1,2j})}{2}\right| \quad (1)$$

$$HL_{i,j} = Or_{2i+1,2j} - Or_{2i,2j} \quad (2)$$

$$LH_{i,j} = Or_{2i,2j+1} - Or_{2i,2j} \quad (3)$$

$$HH_{i,j} = Or_{2i+1,2j+1} - Or_{2i,2j} \quad (4)$$

The inverse transform is given as,

$$Or_{2i,2j} = LL_{i,j} - \left|\frac{HL_{i,j}}{2}\right| \quad (5)$$

$$Or_{2i,2j+1} = LL_{i,j} + \left|\frac{HL_{i,j}}{2}\right| \quad (6)$$

$$Or_{2i+1,2j} = Or_{2i,2j+1} + LH_{i,j} - L_{i,j} \quad (7)$$

$$Or_{2i+1,2j+1} = Or_{2i+1,2j} + HH_{i,j} - LH_{i,j} \quad (8)$$

The inverse coefficient were represented as follows,

$$O_{2i,2j} = LL_{i,j} - \left|\frac{HL_{i,j+1}}{2}\right| \quad (9)$$

$$O_{2i,2j} = LL_{i,j} + \left|\frac{(HL_{i,j+1})}{2}\right| \quad (10)$$

$$O_{2i+1,2j} = O_{2i,2j+1} + LH_{i,j} - L_{i,j} \quad (11)$$

$$O_{2i+1,2j+1} = O_{2i+1,2j} + HH_{i,j} - LH_{i,j} \quad (12)$$

Then we will go for the Keypoint Localization, here we will finalize the potential keypoint locations and then will refine them for more accurate results by using taylor series and here the need of the contrast threshold comes up, any pixel intensity whose level is 0.03 then it will be rejected and only the pixel levels greater than 0.03 are considered. The next step will be of Orientation Assignment which is necessary to eliminate any errors in image orientation, here we will orientation for every key point identified in previous step by considering the neighborhood pixels. Then we will go for Keypoint Descriptor where we will take a region around the keypoint and will separate it into sub squares and make a histogram for each sub square. A vector is created from these histograms called as keypoint descriptor and finally these keypoints are matches with remaining neighbors and is called as Keypoint Matching. Here we will consider the ratio of distance between thesimilar neighborhood and will consider only those neighbors whose distance is greater than a minimum threshold value of our algorithm.

C. CNN:

CNN is vital aspect of image processing uses perceptron to analyze data of given image or object. The procedure can be extensively isolated into two phases. First stage is feature learning and the subsequent one is classification.CNN is a part of deep learning algorithm in which we will give a image as input to the algorithm then some weights are assigned to different objects in the image so as to differentiate the objects from one another. These weights are assigned based upon the biases. Here for all this to happen we need to do some preprocessing to the image called as ConvNet which is necessary for the feature extraction but here in our model we have used SIFT algorithm for the feature extraction so we don't need to have ConvNet for our proposed model. Once the features are extracted the first stage in CNN is convolution and this is done by using a kernel. The size of the kernel is decided depending upon the requirement and for efficient output. Once the image is convolved using the kernel then the Pooling stage comes up, for reducing the spatial size of the convolved input image which is necessary to reduce the computational requirements. Then we will implement valid padding on the image this is done to increase the dimensionality of the convolved feature. Then this layer is provided for the fully connected network for the classification of the image.

March-April 2020
ISSN: 0193-4120 Page No. 1488 - 1494

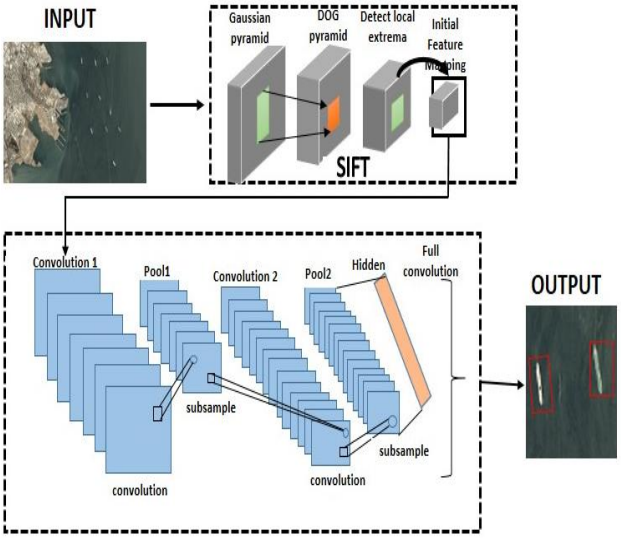


Figure 1: Architecture of SIFT-CNN

4. Present System

In Existing methodology SIFT Feature extricating strategy for cash examination has been dead. To acknowledge and portray near highlights in footage Scale-invariant feature Transformation (or SIFT) which may be a calculation in computer vision is. The SIFT highlights are separated from each one of the countenances within the information. At that time, given another face image, the highlights freed from that confront are thought of against the highlights from every face within the information. The face within the information with the largest range of coordinative focuses is taken into account the face, and is used for the order of the new face. A part is viewed as coordinated with another element once the separation to it embrace isn't specifically a selected portion of the separation to the subsequent component. This ensures we have a tendency to diminish the number of false matches. This can be on the grounds that within the event of a false match, there'll be numerous alternative shut highlights with shut separations, attributable to the high spatial property of the highlights. Then again, if there ought to arise an event of a right match, it's in all probability not reaching to discover another part that's overly shut attributable to the terribly clear nature of SIFT highlights. Another significant customary for these features is, the corresponding situations between them and inside the underlying scene should not amendment beginning with one image then onto consecutive. For instance, if just the four corners of partner degree passage were used as features, they may work paying very little mind to the entryway's position; however on the off probability that focuses within the edge were to boot utilized, the acknowledgment would come back up short if the entry is opened or shut. So likewise, features situated in clarified or adaptable articles would normally not work if any modification inside their inside unadulterated arithmetic occurs between 2 footage in the set being ready. still, much speaking SIFT identifies associate degreed utilizes an plenty larger range of highlights from the photographs, that lessens the commitment of the blunders caused by these near varieties within the traditional mistake of all part coordinative blunders.

$$D(x, y,\sigma) = L(x, y, k\sigma) - L(x, y,\sigma) \quad \textbf{(13)}$$

$$L(x, y,\sigma) = G(x,y,\sigma)* I(x,y) \quad \textbf{(14)}$$

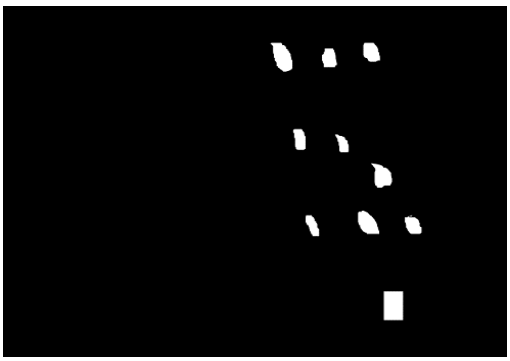
$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} \exp\{-\frac{x^2+y^2}{2\sigma^2}\} \quad \textbf{(15)}$$

$$D(Z) = D + \frac{1}{2}\frac{\partial D^{-1}}{\partial x}Z \quad \textbf{(16)}$$

5. Proposed Method

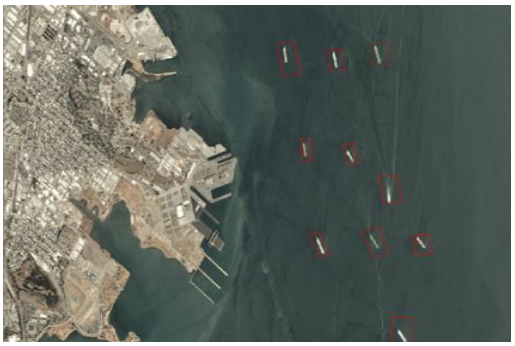
In this paper a technique to extract and detect features from the optical satellite imagery is proposed. For extraction and detection process here we have used SIFT and convolutional neural network techniques. Once the features are extracted they are classified based on CNN technique. Here to verify the proposed technique an Optical Satellite image of a sea shore is considered which is having ships in the sea.

(a) Input Image



(a1) Masked Image

March-April 2020
ISSN: 0193-4120 Page No. 1488 - 1494

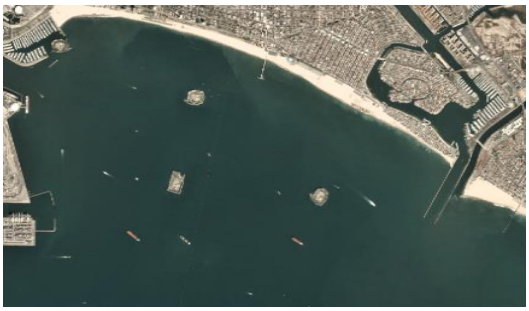


`(a2) Ship Detection

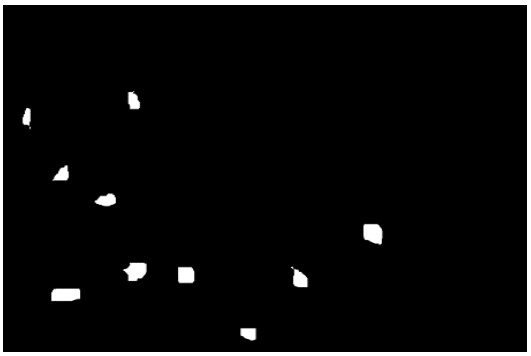
(b) Input Image

(b1) Masked Image

(b2) Ship Detection



(c) Input Image



(c1) Masked Image

(c2) Ship Detection

Figure 2: Object detection steps for three different input images

This image is provided as input to the proposed system and after processing we can observe the features that is the ships in the images were detected by using the proposed method.

The language used here for the development of the proposed method is Python with a front end navigator of Anaconda Navigator – Spyder. The steps involved in the technique are presented below. Here for feature detection, preprocessing (raw data is converted into a clear dataset), edge detection, segmentation we have used SIFT algorithm and for classification we have used CNN algorithm.

March-April 2020
ISSN: 0193-4120 Page No. 1488 - 1494

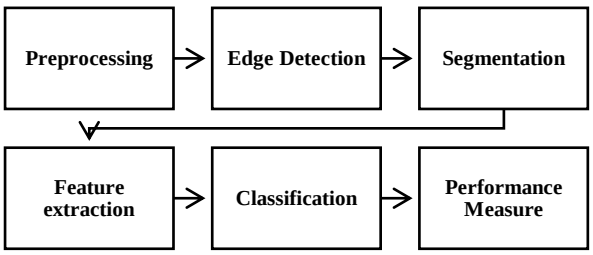


Figure 3: Proposed Strategy

Here to validate the proposed technique we have considered three different input images of a sea shore which consists of ships as shown below in the figure 2, once these input images were considered the preprocessing techniques were implemented on the input images so as to enable the image to be considered for the further stages of feature extraction and classifications. The coefficients are calculated based upon the following equations. The steps in object identification and appropriate outputs for each step are presented in the Figures 2 and 3.

For the above given inpit images in figure 2(a),(b),(c) we will apply SIFT algorithm and the maskind images were extracted in which only the ships are masked and the information of differtent features were distinguised. The masked images are of as shown in figure 2 (a1),(b1),(c1).

Once the masked images were obtained then these masked images will be overlayed on the input images and the necessary features were classified based upon the CNN algorithm. The main object of our project is to detect only the ships excludind the other objects in the image. Finally in the output the ships are localized by bounding boxes. shown in the figure 2 (a2), (b2), (c2).

6. Conclusion

In the proposed method, CNN-SIFT, an object detection for ships in optical satellite imagery is implemented. It consists of two phases. Initially, features are extracted i.e., space scale extrema detection, key point localization, orientation assignment, key point descriptor and masking is performed by SIFT. In second phase, CNN is applied and ship localization by bounding-box is achieved. The presentation of the proposed strategy is evaluated by considering optical satellite datasets containing ships along with other objects resembling the shape of ship. By implementing CNN-SIFT technique, 97% of accuracy is achieved. Python with front end navigator of anaconda – Spyder is used. As a future work, time complexity can still be improved.

References

[1] C. Wang, S. Jiang, H. Zhang, F. Wu, and B. Zhang, “Ship detection for high-resolution SAR images based on feature analysis,” *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 119–123, Jan. 2014.

[2] Guangzhi Shi, Junchuan Hu, Lianglong Da and Rugang Song, "Target recognition study using SVM, ANNs and expert knowledge," *2008 IEEE International Conference on Automation and Logistics*, Qingdao, 2008, pp. 1507-1511.

[3] J. R. Powell, S. Krotosky, B. Ochoa, D. Checkley and P. Cosman, "Detection and identification of sardine eggs at sea using a machine vision system," *Oceans 2003. (IEEE Cat. No.03CH37492)*, San Diego, CA, USA, 2003, pp. 175 Vol.1.

[4] K. Eldhuset, “An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions,” *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1010–1019, Jul. 1996.

[5] N. Singhal, N. Singhal and V. Kalaichelvi, "Image classification using bag of visual words model with FAST and FREAK," *2017,(ICECCT)*, Coimbatore, 2017, pp. 1-5.

[6] Y. Xiang, F. Wang and H. You, "OS-SIFT: A Robust SIFT-Like Algorithm for High-Resolution Optical-to-SAR Image Registration in Suburban Areas," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3078-3090, June 2018.

[7] R. Shirvany, M. Chabert, and J.-Y. Tourneret, “Ship and oil-spill detection using the degree of polarization in linear and hybrid/compact dual-pol SAR,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 3, pp. 885–892, Jun. 2012.

[8] L. Huang, W. Li, C. Chen, F. Zhang, and H. Lang, “Multiple features learning for ship classification in optical imagery,” Multimedia Tools Appl., vol. 77, no. 11, pp. 13363–13389, Jun. 2018.

[9] N. Proia and V. Page, “Characterization of a Bayesian ship detection ´ method in optical satellite images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 226–230, Apr. 2010.

[10] M. Kartal and O. Duman, "Ship Detection from Optical Satellite Images with Deep Learning," *2019 9th International Conference on Recent Advances in Space Technologies (RAST)*, Istanbul, Turkey, 2019, pp. 479-484.

[11] S. Qi, J. Ma, J. Lin, Y. Li, and J. Tian, “Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1451–1455, Jul. 2015.

[12] L. Li and J. Wang, "SAR image ship detection based on Ant Colony Optimization," *2012 5th International Congress on Image and Signal Processing*, Chongqing, 2012, pp. 1100-1103.

[13] H. He, Y. Lin, F. Chen, H. M. Tai, and Z. Yin, "Seashore ship detection in remote sensing images via weighted pose voting," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, pp. 3091–3107, Jun. 2017.

[14] S. Li, Z. Zhou, B. Wang, and F. Wu, "A novel seashore ship detection via ship head classification and body boundary determination," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1920–1924, Dec. 2016.

[15] J. Xu, X. Sun, D. Zhang, and K. Fu, "Automatic detection of seashore ships in high-resolution remote sensing images using robust invariant generalized Hough transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2070–2074, Dec. 2014.

[16] Y. Xiang, F. Wang and H. You, "OS-SIFT: A Robust SIFT-Like Algorithm for High-Resolution Optical-to-SAR Image Registration in Suburban Areas," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3078-3090, June 2018.

[17] Kaza Deepthi Sudha, Pooja Kilaru, Manna Sheela Rani Chetty "Currency Note Verification and Denomination Recognition on Indian Currency System", International Journal of Recent Technology and Engineering (IJRTE), Volume-7, Issue-6S4, April 2019.

[18] Sandeep Dwarkanath Pande, Dr. Manna Sheela Rani Chetty "Analysis Of Capsule Network (Capsnet) Architectures And Applications", Jour of Adv. Research in Dynamical & Control Systems, Vol. 10, 10-Special Issue, 2018.

[19] Wong Seng Yue, "Design and Realisation of Automated Solar Tracking System", Vol. 4, 1-Issue, 2018.