International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Vehicle Detection In Remote Sensing Images

Mohammed A.-M. Salem, Sultan Almotairi

Abstract: ***Traffic monitoring and management is one of the most crucial tasks of governing bodies in modern big cities. With each passing day the traffic problem grows in complexity due to the continuous increase of participating vehicles and the hard expansion of the road network and parking places. In this article we introduce a new method for vehicle detection and localization in parking lots using high resolution UAV images. In order to end up with practical and yet effective approach, which could be implemented on low computing hardware resources and integrated with the camera in the UAV, we considered simple steps in the proposed algorithm for optimization. It follows the machine learning pipeline such as preprocessing, sensing, feature extraction, training and classification. In preprocessing the images are thresholded iteratively in multiple color spaces to extract the candidate regions of interest (ROI). The algorithm relies on point and shape features using fast techniques in the feature extraction. The features are then clustered by the K-means algorithm and represented by the resulted clusters' centers. Region based linear classification is finally applied using SVM to classify if the object is a vehicle or else. The proposed approach proved high detection and classification accuracy more than 86% and still running under the low complexity constraint..***

Keywords: ***Unmanned Arial Vehicles, Traffic Monitoring, Image Segmentation; Machine Learning; Point features; Shape features; Speeded Up Robust Features extractor (SURF); Support Vector Machines (SVM).***

I. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) were developed for military purposes at first, but then due to their great capabilities, they found their way into many civilian applications. Today, UAVs are widely available for any people and can be used in many applications such as surveillance, object discovery and even pizza delivery. Smart cities management is one of the promising applications which is very demanding for deploying UAVs for remote high-resolution imagery [1, 2, 3].

UAVs images could be used for cultured land, crops and land usage, which based on some indices it is possible to extract meaningful information for farm management and decision-making [4,5]. Yoo et. al. introduced one use of a UAV to create 3D mapping of beach images form monitoring where eroded and deposited sand volumes were estimated using combination of quarterly ground measurements and elevation data [6].

Vehicle localization processes in roads are utilized for vehicle tracking, counting, recording, activity investigation, [7,8,9], etc. They are multiple ways to do so, but three main approaches are used to detect and segment vehicles out from the road background [10,11,12] which are: feature-based methods, frame differencing & motion-based methods, and background modeling and subtraction.

In machine learning, SVM (Support Vector Machines) is one of the most popular classification techniques with related learning calculations that break down information provided for learning and further investigation [13]. Given a set of prepared data, each instance of an object is set to belong to one of two classifications, an SVM prepares calculations to fabricate a model that is used to classify new data into one class or the other, and this makes it a non-probabilistic paired straight classifier.

Many vehicle detection techniques were proposed recently. Cheng, et al. [14], proposed a framework for car detection in aerial traffic monitoring using dynamic Bayesian networks. Luo-Wei Tsai, et al. [15] proposed another novel method for detection in static images using color transformation models. The method for detection is done in multiple stages. The first stage colors are transformed to easily identify the pixels that have vehicles from the background pixels. Afterwards edge features are extracted from the segmented regions and used by a cascaded multichannel classifier. The authors refer the achieved high accuracies to the combination of the global color features and the local edge features.

Hansen et al. [16], proposed a model for motion detection which is followed by background subtraction and then binary classification of vehicles. Classification is done using an SVM applied on features generated from HOG (Histogram of Gradients) [17] where is represents the object contours and size in order to distinguish between vehicles and false alarms as vehicles have different shapes and sizes. The last step is road context extraction, multi object tracking is carried out in the categorization step by exploring the trajectories of objects. In order to determine roads networks from the obtained trajectories, parallel object tracking is done by analyzing the short tracks and then further associating them to complete longer tracks.

Moranduzzo et al. [18] proposed another method to detect cars obtained by a UAV sensor which generates very high resolution images (2cm). The first step is screening of asphalted zones where car would be detected. The aim of this step is to reduce the false alarms that may raise from the urban regions in the images. Then, local feature extraction processing is performed based on either SIFT (Scalar Invariant Feature Transform) [19] or SURF extractor.

Manuscript published on 30 September 2019.
***Correspondence Author(s)**
Mohammed A.-M. Salem, Faculty of Media Engineering and Technology, German University in Cairo, Egypt. Faculty of Computer and Information Sciences, Ain Shams University, Egypt
Sultan Almotairi*, Department of Natural and Applied Sciences, Faculty of Community College, Majmaah University, Majmaah, 11952, Saudi Arabia, Email: almotairi@mu.edu.sa

The Algorithm then tries to classify each of these key points by using SVM (Support Vector Machine) classifier which discriminates if each of the points belong to cars or anything else. The final step of the procedure is to group all the key point that belong to the same car to get a one-to-one relationship between key-points and cars. After the procedure is done, the car count in the scene is simply determined by the final number of found key points.

Many recent works are based on deep learning for the application of counting moving vehicles or for traffic scene understanding [20, 21]. Zheng et al. [22], used five deep unsupervised learning models to learn driving modes. However, the data collected from different sensors mounted in the vehicles. Biswas et al. [23], used speed estimation for moving object detection based on UAV system. Faster R-CNN is applied to detect the objects, and a discriminative correlation filter with CSRT is used for tracking. The main challenge in deep learning methods is the existence of labeled data. These methods are very hungry to data and the results are sensitive to the environment used to generate the label data.

II. IMAGE ACQUISITION AND PREPROCESSING

A. Overview

The proposed system uses the most common modules for pattern recognition such as preprocessing, sensing, training, feature extraction and classification. The proposed method is presented in Figure 1, it illustrated the training phase as well as the testing phase. The input is a set of remote sensed images captured from a UAV in RGB color space which is then converted to different color models in order to facilitate the extracted the candidate regions of vehicles. During the training phase, images vehicles as well as non-vehicles are used. The SURF algorithm extracts the interest points in order to build the feature vector for each image. Then the feature vectors and the classifications are fed to the SVM for training. After training is complete, testing then could be done. The first step is to examine each pixel in order to extract the ROI (Region of Interest), then feature extraction is carried out using the same method as the training phase, then finally the extracted features get classified using the SVM.

B. Dataset

The dataset was created by the University of Trento, lab of Intelligent Information Processing, Italy [18]. The camera used for image acquisition is Canon EOS 550D which contains 18 MP CMOS APS-C sensor. The images are 2 cm spatial resolution and are captured in the RGB color space. The images have typical resolution of 5184×3456 pixels with color depth of 8 bits per pixel.

C. Region of Interest Extraction

Determining ROI (Region of Interest) is done to decrease the amount of false alarms and speed up the detection process. This method is approached by applying processing on multiple color spaces [24]. Where images are processed by changing the color space components. Mainly HSV and YCbCr were used for investigation. HSV is a hue color model which separates the pure color from the saturation and brightness. YCbCr is one of the most popular color spaces in computing. Its naming comes from combing three letters which have different properties where Y (represents luminance which is measurement of the amount of energy that is perceived from a light source). Cb (represents the difference between the blue component of the images and a certain reference value). Cr (represents the difference between the red component of the images and a certain reference value) [25].

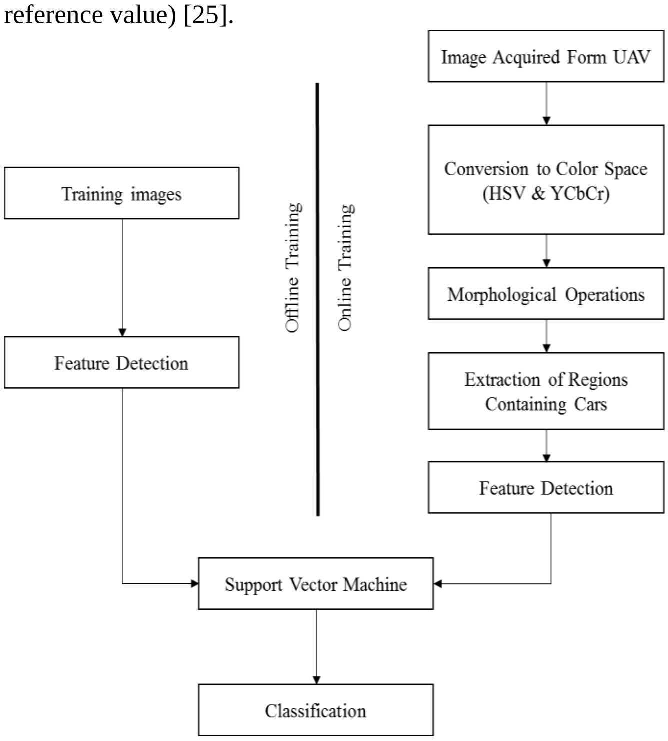


Fig. 1. Block diagram of the proposed method. The system is divided into training (offline phase) and testing (online phase).

Since that many elements can be found in the acquired. Many researchers used the hue and luminance to classify them in five categories [15]. The chrominance channels C_b, C_r and the pixel's values aren't used because they do not provide any useful information from the evaluations. The listing below contains the relationships between the color space components:

1. The smallest "hue" values used to represent buildings, which ranges from 0 to 0.1
2. Roads have high saturation value of about 0.3, hue value between 0.05 and 0.2, and luminance between 110 and 160 lumens.
3. Sandy regions have minimum saturation value about 0.3, and similar luminance value with roads.
4. Trees have a luminance value below 100

Therefore, these essential adjustments are applied to the image using the condition in equation (1), which is applied to each image pixel in order to separate between the ROI and the background.

$$(100 < Y) \text{ AND } (S < 0.3) \text{ AND } (H \geq 0.05) \qquad (1)$$

The resulted image of the above step is a binary image, which undergoes some enhancement steps to create a mask image for ROI extraction.

Retrieval Number: K18070981119/19©BEIESP
DOI: 10.35940/ijitee.K1807.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

First a filling operation is used to fill the holes in the input binary image, then all small objects or connected components that are smaller than 20 pixels are removed from the binary image. The size threshold was chosen after several trials. The next morphological operations used are the closing operation followed by dilation.

III. FEATURE EXTRACTION AND CLASSIFICATION

A. *Feature Extraction*

The resulted image from the ROI algorithm undergoes an automated cropping of size (30x30) pixels. The cropped images were then subject to interest point extraction using SURF. The SURF feature descriptor generates huge amount of redundant features that could not be used directly as an input for the classification step. Therefore, a feature representation is usually following this step. For that we used the Bag-of-visual-Words (BoW). A bag of words is a scattered vector of occurrence counts of words. In computer vision however, a bag of words is a visual vector of occurrence counts in a vocabulary of local image features. Utilizing the BoW model for picture is done by developing an extensive vocabulary of numerous visual words and speak to every picture as a histogram of the recurrence words that are in the picture.

Three steps are followed to represent an image using BoW: feature extraction, feature representation, and generating a codebook. The SURF is applied on the extracted ROI's to extract the representing features. SURF is a local feature extractor that extracts the interest points based on edge images. It used widely for image and object matching as it is scale and transformation invariant and is robust against illumination changes [26]. Despite its robustness, SURF is appropriate for embedded systems as investigated by Liénard et. al. [27] for real-time execution on the low-cost processing hardware embedded on the UAV.

The features are then arranged in a vector and is called feature descriptors. The K-means algorithm is used afterwards to cluster the vectors in k-non-overlapping clusters [28]. Similar features are grouped together in one cluster and are separated from other different features that are associated to the other clusters. As a result, each cluster center represents a representative and differentiable feature or a visual word. These words are then fed into the classification stage afterwards.

B. *ROI Classification using SVM*

The input for the classifier is a set of training sample points, where each sample point (ROI) is marked to belong to one of two categories. The training set contains positive images containing the required object which is car in our case and negative images of different objects. The training set is passed to the bag of features to create features which is then fed to the SVM to train on it. The SVM takes the features and divided into two parts, the first part is used to train, and the other part is for validation. After the training of the SVM, a confusion matrix is supplied to be able to choose the best SVM model to be imported to be used in new data prediction. The training set images is exposed to the same previous steps. However, there is no morphological operations is done on them.

IV. RESULTS

The results of each step in the block diagram will be shown in order to summarize the work done in this paper and to simplify the process. Finally, up on the results shown, it could be concluded that the proposed system has proved a high performance and robustness.

A. *Test Dataset*

The dataset are prepared from high resolution colored images captured by controlled UAV for parking lot in the University of Trento, Italy as mentioned before. lab of Intelligent Information Processing, Italy [18]. The camera used for image acquisition is Canon EOS 550D which contains 18 MP CMOS APS-C sensor. The images are 2 cm spatial resolution and are captured in the RGB color space. The images have typical resolution of 5184×3456 pixels with color depth of 8 bits per pixel.

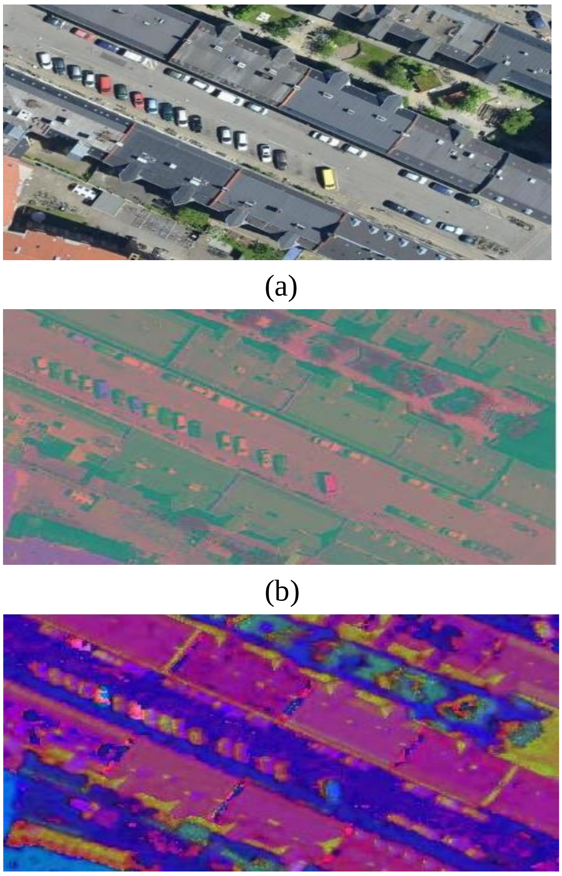
B. *ROI Extraction*

As shown in Figure 2 a), the input image for the system is in three color channel (RGB), the image first is converted to two color spaces, YC_bC_r as shown in Figure 2 b), and HSV, as shown in Figure 2 c).

These images undergo a pixel by pixel loop to apply equation (1) to create a mask image as shown in Figure 3. After applying iterative morphological operations on the binary image as opening and closing, we obtain the final resulted mask image as shown in Figure 4.

C. *SVM Training and Classification*

For training and testing purposes, 596 cropped images are created by the previous step. All the images are labeled manually. 400 images are used for training and validation of the SVM and 196 cropped images are used for testing the trained model. Three trials are performed to test the overall accuracy of the training versus the size of the training dataset. In each trial 70% of the images are used for training and 30% are used for validation. The number of images in the dataset for the first test was 200 images, then 300 images for the second test and 400 images for the third test. It has been found that increasing the number of the training images leads to increase the accuracy of the SVM up to 86.2%. Summary of the results and the accuracy of the training after validation against the dataset size are shown in Table 1. Figure 5 shows that the curve of the accuracy increased slightly when the number of samples in the training dataset increased from 300 to 400.



(a)

(b)

(c)

Fig. 2. Different representation of a UAV's image in the color models RGB, YC_bC_r and HSV, respectively.

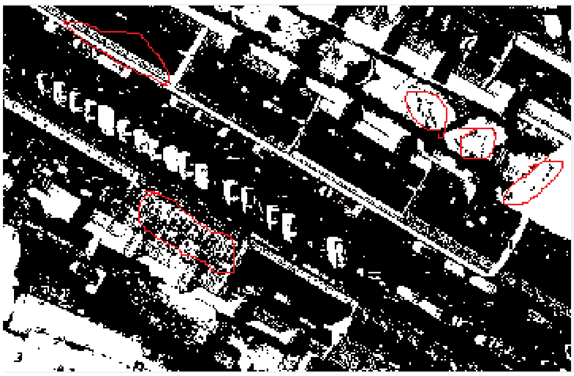


Fig. 3. Mask image

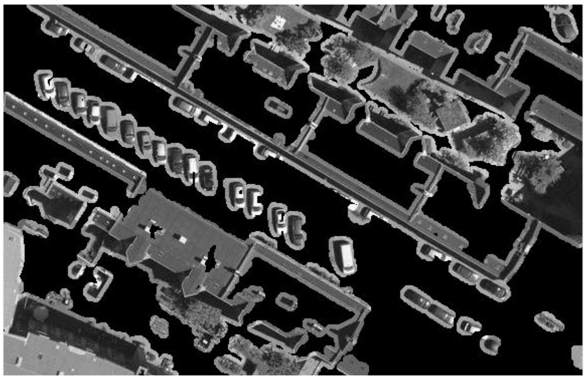


Fig. 4. ROI image

D. *Testing and Classification*

After training the SVM a test set is created for manual testing. The image set created form the automatic cropper is passed to the bag of word to create visual features and the prediction is done on each image to know whether it is a car or not. Table 2. shows the confusion matrix. The overall accuracy achieved is very close to that obtained in the training, namely 86.1%. The true-positive rate (TPR) or the sensitivity is 83.6%, while the precision or positive predictive value (PPV) is 80.3%.

Table 1. Comparison between the numbers of training images

Number of training images	Accuracy
200	78.1 %
300	84.7 %
400	86.2 %

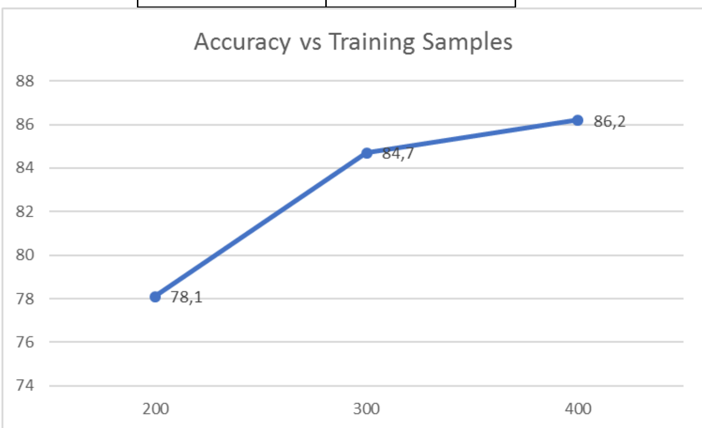


Figure 5: Overall classification accuracy of the ROI against the number of training samples.

Table 2. Confusion matrix of the manual test dataset.

		Predicted	
		Vehicle	Background
Actual	Vehicle	61	12
	Background	15	106

Retrieval Number: K18070981119/19©BEIESP
DOI: 10.35940/ijitee.K1807.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

V. CONCLUSION

Urban communities everywhere throughout the world manage a large number of vehicles going through their roads every day. The expanding number of vehicles is viewed as the primary reason for traffic jam. This makes it hard likewise to discover a spot to park the car. The greatest problem in these years is the security and the activity control of the roadways, attempting to make individuals life less demanding. Numerous answers for taking care of this issue is proposed. One way is utilizing remote sensing field, particularly Unmanned Aerial Vehicle (UAV) which these days picks up an awesome fame in this field.

In this paper a we have proposed a simple yet robust technique for vehicle localization in the traffic monitoring applications. The algorithm is able to process single high-resolution image in real time, and hence is suitable for further implementation on low computing hardware embedded on UAV.

ACKNOWLEDGMENT

The authors their appreciation to the Deanship of Scientific Research at Majmaah University for funding this work under project number NO (RGP-2019-29).

REFERENCES

1. Mohammed, Farhan & Idries, Ahmed & Mohamed, Nader & Al-Jaroodi, Jameela & Jawhar, Imad. (2014). UAVs for smart cities: Opportunities and challenges. 267-273. 10.1109/ICUAS.2014.6842265.
2. Kanistras, Konstantinos & Martins, Goncalo & J. Rutherford, Matthew & Valavanis, K.P.. (2013). A survey of unmanned aerial vehicles (UAVs) for traffic monitoring. International Conference on Unmanned Aircraft Systems (ICUAS), May 2013, 221-234. DOI: 10.1109/ICUAS.2013.6564694.
3. Stöcker, Claudia & Bennett, Rohan & Nex, Francesco & Gerke, Markus & Zevenbergen, J. (2017). Review of the Current State of UAV Regulations. Remote Sensing. 9. 459. 10.3390/rs9050459.
4. Simelli, I., Tsagaris , A. “The Use of Unmanned Aerial Systems (UAS) in Agriculture” Proceedings of the 7th Int. Conference on Information and Communication Technologies in Agriculture, Food and Environment, Kavala, Greece, Sep.17-20, (2015).
5. Austin, Reg. “Unmanned aircraft systems: UAVs design, development and deployment”. Vol. 54. John Wiley & Sons, (2011).
6. Yoo, C. I., and T. S. Oh. "Beach Volume Change Using UAV Photogrammetry Songjung Beach, Korea." ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2016): 1201-1205.
7. Tang, Z. Miao, Z., “Fast Background Subtraction and Shadow Elimination Using Improved Gaussian Mixture Model”, Haptic Audio and Visual Environments and Games, HAVE IEEE Int. Workshop on , vol., no., pp.38,41, 12-14 Oct. (2007).
8. Huang, Cheng-Ming, Chen, Yi-Ru, Fu, Li-Chen, “Real-time Object Detection and Tracking on a Moving Camera Platform,” ICCAS-SICE, 717 (722) (Aug 2009), pp. 18–21 (2009)
9. Everaerts, J. “The Useof Unmanned Aerial Vehicles (UAVS) for Remote Sensing and Mapping” The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B1. Beijing (2008).
10. Dhawad, S. R., Itkarkar, R. R., “Survey on Car Detection A TechniquesUsing Aerial Images: An Overview”, IOSR Journal of Electronics and CommunicationEngineering (IOSR-JECE) ISSN:2278-8735. pp57-60, www.iosrjournals.org.
11. Selvakumar S., and Kalaivani. S. , “Comparative Study on Vehicle DetectionTechniques in Aerial Surveillance”, International Journal on Cybernetics& Informatics (IJCI) Vol.2, No.4, August (2013).
12. Ramakrishnan, V., Prabhavathy, A. K., Devishree, J., “A Survey on Vehicle Detection Techniques in Aerial Surveillance”, International Journal of Computer Applications (0975 – 8887) Volume 55– No.18, October (2012).
13. Salem, M. A.-M. Multi-stage localization given topological map for autonomous robots (2012) Proceedings - ICCES 2012: 2012 International Conference on Computer Engineering and Systems, art. no. 6408483, pp. 55-60. DOI: 10.1109/ICCES.2012.6408483
14. Cheng, H.-Y., Weng, C.-C., Chen, Y.-Y., “Vehicle Detection in Aerial Surveillance Using Dynamic Bayesian Networks”, IEEE Transactions on image processing,Vol.21, No.4, (2012).
15. Tsai, L.-W., Hsieh, J.-W., Fan, K.-C., “Vehicle Detection Using Normalized Color and Edge Map”, IEEE Transactions on Image Processing, vol. 16, no. 3, March (2007).
16. Hansen, M., Anadan, P., Dana, K., Van de Wal, G., Burt, P., “Real-time Scene Stabilization and Mosaic Construction”, Proc of IEEE CVPR, 54-62, (1994).
17. Dalal, N., Triggs, B., “Histograms of Oriented Gradients for Human Detection”. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01 (CVPR '05), Vol. 1. IEEE Computer Society, Washington, DC, USA, 886-893. (2005)
18. Moranduzzo, T.; Melgani, F., "Automatic Car Counting Method for UnmannedAerial Vehicle Images," in Geoscienceand Remote Sensing, IEEE Transactions on ,vol.52, no.3, pp.1635-1647, (2014).
19. Lowe, D. “Distinctive Image Features form Scale-invariant Keypoints,” Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, (2004).
20. Attique, M., Cho, H., Jin, R., & Chung, T. (2016). Efficient Processing of Continuous Reverse k Nearest Neighbor on Moving Objects in Road Networks. ISPRS Int. J. Geo-Information, 5, 247.
21. Mayar Shafaey, Mohammed A.-M. Salem, H. M. Ebied, M. N. Al-Berry, M. F. Tolba, (2019). Deep Learning for Satellite Image Classification. Proceedings of the Int. Con. on Advanced Intelligent Systems and Informatics 2018, pp. 383--391 DOI10.1007/978-3-319-99010-1_35
22. Han Zheng, Zanyang Cui, Xingchen Zhang, (2019). Automatic Discovery of Railway Train Driving Modes Using Unsupervised Deep Learning. International Journal of Geo-Information, Vl. 8, No.7, pp 2220-9964, 2019, DOI 10.3390/ijgi8070294
23. Biswas, Debojit & Su, Hongbo & Wang, Chengyi & Stevanovic, Aleksandar. (2019). Speed Estimation of Multiple Moving Objects from a Moving UAV Platform. International Journal of Geo-Information. 2019. 259. 10.3390/ijgi8060259.
24. Bong, D. B. L., Lai, K.C., Joseph, A. “Automatic Road Network Recognition and Extraction for Urban Planning”,World Academy of Science, Engineering and Technology International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering Vol:3, No:5, (2009).
25. Kadouf, H., Mustafah, Y., "Colour-based Object Detection and Tracking for Autonomous Quadrotor UAV", IOP Conference Series: Materials Science and Engineering, Vol. 53, No. 1, (2013).
26. N. N. Saeed, M. A. -. Salem and A. Khamis, "Android-based object recognition for the visually impaired," 2013 IEEE 20th International Conference on Electronics, Circuits, and Systems (ICECS), Abu Dhabi, 2013, pp. 645-648. doi: 10.1109/ICECS.2013.
27. Liénard, Jean; Vogs, Andre; Gatziolis, Demetrios; Strigul, Nikolay. 2016. Embedded, real-time UAV control for improved, image-based 3D scene reconstruction. Measurement, Vol. 81: 6 pages.: 264-269.
28. Wang, Y., Zou, Y., Shi, H., Zhao, H., “Video Image Vehicle Detection System for Signaled Traffic Intersection”, Hybrid Intelligent Systems, 2009. HIS’09. Ninth International Conference on , vol.1, no., pp.222,227, 12-14 Aug. (2009)

AUTHORS PROFILE

Mohammed A.-M. Salem is an Associate Professor in the faculty of Computer and Information Sciences, Ain Shams University in Cairo, Egypt since 2009. He received his B.S. degree in 1997 in the field of Mathematics and Computer Science. In 2002, he received the M.Sc. degree in Scientific Computing from Ain-Shams University in Cairo. In November 2008, he received the PhD degree from Humboldt-Universität in Berlin, Germany in Multiresolution Image Segmentation. His research interests include: Robot Vision, Multimedia, Signal Processing, Multiresolution Analysis, and Wavelet Transform.

Sultan Almotairi received his B.Sc., M.Sc., and PhD. degrees in computer science from Florida Institute of Technology, Melbourne, USA, in 2010, 2012, and 2014 respectively. Currently, he is an assistant professor at Department of Natural and Applied Sciences in Community College of Majmaah University. He has acted as the dean of Community College at Majmaah University since June 2015. In 2016, he was elected as the Chairman of the Municipality Council of Majmaah. His research interests include neural networks, deep learning, pattern recognition, machine learning, image processing, and computer vision.

Retrieval Number: K18070981119/19©BEIESP
DOI: 10.35940/ijitee.K1807.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)