International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 11 (2018) pp. 9957-9964
© Research India Publications. http://www.ripublication.com

A Survival Study on Different Aerial Image Classification Techniques

A.Gokila Vani

Ph.D. Scholar, Hindusthan College of Arts and Science,
Coimbatore - 641028, Tamil Nadu, India.

Dr. V. Saravanan

Associate Professor & HEAD,
Department of IT (PG), Hindusthan College of Arts and Science,
Coimbatore - 641028, Tamil Nadu, India.

Abstract

Image processing is a technique to achieve the operations on image for improving image quality or to extract valuable information. It is a signal processing method where input is an image and output is an image or features. Classification is defined as the process of categorizing image into number of classes depending on similarity, threshold, etc. An aerial image is projected image and not viewed in normal manner. Aerial image classification categorizes and detects the objects on the digital maps. Color, edge, shape and texture are taken out from input aerial images to categorize the objects through different existing classification techniques. Scene classification is defined as the process of offering information about the semantic type or function of the given image. Scene classification is taken as the complex task due to lack of discriminative features at the high level. Though the many scene classification methods were introduced, the performance was not improved. Our main objective of research work is to improve the classification performance by studying existing issues.

***Keywords*:** image processing, scene classification, aerial images, discriminative features, semantic type, texture

INTRODUCTION

Image processing is the process of transforming the image into digital form for carry outing desired task through extracting relevant information. Scene classification method classifies the input aerial images depending on the scenes. Scene classification is an essential problem for computer vision and received large attention in recent days. Multi-spectral aerial image comprises 3-15 bands (i.e., objects) while the hyper-spectral aerial image includes hundreds of objects. Multi-spectral and hyper-spectral aerial image classification is the process of classifying objects into number of classes depending on the extracted features of the objects. Multispectral image collects the image data within specific wavelength ranges across the electromagnetic spectrum. The wavelengths are divided by filters or by instruments that are sensitive to specific wavelength. Hyper spectral imaging gathers and processes the information from electromagnetic spectrum. The main objective of the hyper spectral imaging is to attain spectrum for every pixel in image of scene for detecting the objects, materials, or processes. In hyper spectral imaging, the recorded spectra include better wavelength resolution and cover broad range of wavelengths. Hyper spectral imaging determines contiguous spectral bands while the multispectral imaging determines the spaced spectral bands.

This paper is structured as follows: Section II explains the review on different aerial image classification techniques, Section III depicts the study and analysis of existing aerial scene image classification, Section IV describes the possible comparison of existing techniques. In Section V, the discussion and limitations of the existing aerial image classification are discussed with future direction and Section VI concludes the paper.

LITERATURE REVIEW

A new framework based on pre trained VGG-Net model was introduced to automatically learn feature descriptor for VHR images. Pre trained visual geometry group network (VGG-Net) model was introduced in [1] as deep feature extractors to extract the informative feature from VHR images. But, discriminant correlation analysis creates the predicted probabilities outside the range [0, 1]. A volumetric texture and reduced-spectral feature was designed in [2] for hyper spectral image classification. Volumetric textural features are extracted through volumetric gray-level co-occurrence matrices (VGLCM). But, the performance was not improved using VGLCM.

A high-order context information extraction method depending on fast sparse representation classification method was introduced in [3] to remove the context information of patch. A sample selection plan was introduced with the sparse representation classification method to design the complete training subset. Multi order features of selected training subset were employed to identify the vehicles depending on super pixel segmentation results of aerial images. But, the sparse representation classification method was not employed for the large database.

An automatic feature extraction from aerial imagery was designed in [4] for ground-based methods. Identification of roof and non-roof objects in informal settlement termed Diepsloot was described. Though the feature extraction was carried out, the classification was not performed with the extracted features. A single remote sensing data source was employed in [5] for 3-D point cloud classification and object detection. A modified Z-buffer algorithm was introduced to handle the false visibility problem and estimated the performances of commercial available dense image matching technology. But, back projection in histogram model recognized only feature in image. Back projection was not appropriate for the feature extraction.

Different techniques were introduced in [6] for classification of aerial images with the descriptors computed through visible spectral bands and information attained from the infrared band. Though dimensionality reduction was performed, the classification performance was not improved. A new classification efficient technique was designed in [7] for Hyper-Spectral Images. The classification of Interior and Exterior Pixels were performed through calculating the Posterior Probability with the pixel intensities. The main objective is to detect the Interior and Exterior Pixels depending on the Optimal Threshold and Probability.

The single-layer and deep convolution networks were developed in [8] for remote sensing data analysis. To perform unsupervised sparse features learning, A greedy layer wise unsupervised pre training linked with effectual algorithm. The algorithm was designed the sparse representations and enhanced lifetime of sparse extracted features. However, large quantity of time was used to train the deep convolution networks.

AERIAL IMAGE CLASSIFICATION

Scene classification is an essential issue for computer vision and received significant attention in recent years. Scene classification differs from conventional object detection/ classification to extent that scene comprises several entities in unpredictable layout. Scene classification addresses issues like differentiating the indoor from outdoor scenes. By using classification technique, large number of scene categories is employed. A multispectral image has different bands of data. Every band of image is displayed one band at time called as grey scale image or in mixture of three bands at time called color composite image.

Deep Feature Fusion for VHR Remote Sensing Scene Classification

The pre trained visual geometry group network (VGG-Net) model was presented to extract informative features from original VHR images. Fully connected layers are constructed through VGG-Net where each layer is taken as the separated feature descriptors. The layers are joined between them to construct final representation of VHR image scenes. Discriminant correlation analysis (DCA) is used as the feature fusion strategy to refine original features extracting from VGG-Net that allow fusion approach with lesser cost than traditional feature fusion strategies. The discriminant correlation analysis (DCA) reduces dimension of features and utilizes adequate method for feature fusion. The pre trained deep CNN models are designed for VHR images scene classification. VGG-Net is developed as feature extractor through choosing useful layers to attain the good demonstration of images scene.

Full connected layers of VGG-Net model are combined where output of each layer is supposed as the feature descriptor and construct the final feature representation of input image. The fused deep feature learning performs better than other feature representation methods like SIFT, speeded up robust features (SURF) and histogram of oriented gradients (HOG), and current methods depending on pre trained CNNs. DCA is introduced to represent the fused features in a low dimension with better classification results. Deep VGG-Net is employed as feature extractor to explain the VHR image scene with representative features. DCA is developed as feature fusion method. The designed approach includes three essential steps. Feature extraction with deep VGG-Net by DCA approach combined the extracted features for VHR scene classification.

VGG-Net outperforms generation of CNN models with public ImageNet dataset and improved the classification results. The scene classification method is depending on low-level features such as SIFT, SURF, HOG, or deep learned features. It is combination of features by VGG-Net model. The output of chosen layers is taken as feature descriptor of input to explain the images scene by informative and essential features. The deep VGG-Net employs the feature extraction from many layers to explain the VHR image scene with informative features. VGG-net framework includes five convolution layers where everyone tracked by pooling layer and three fully connected layers.

Vehicle Detection in High-Resolution Aerial Images based on Fast Sparse Representation Classification and Multi order Feature

A fast sparse representation classification method is introduced to improve the low detection efficiency when the trained dictionary comprises large number of items. The training samples are partitioned into many classes to instruct the small dictionaries. Depending on the trained small dictionaries, detection efficiency is improved. Contextual information is essential for increasing the object detection performance. The vehicles in aerial images comprise rich contextual information.

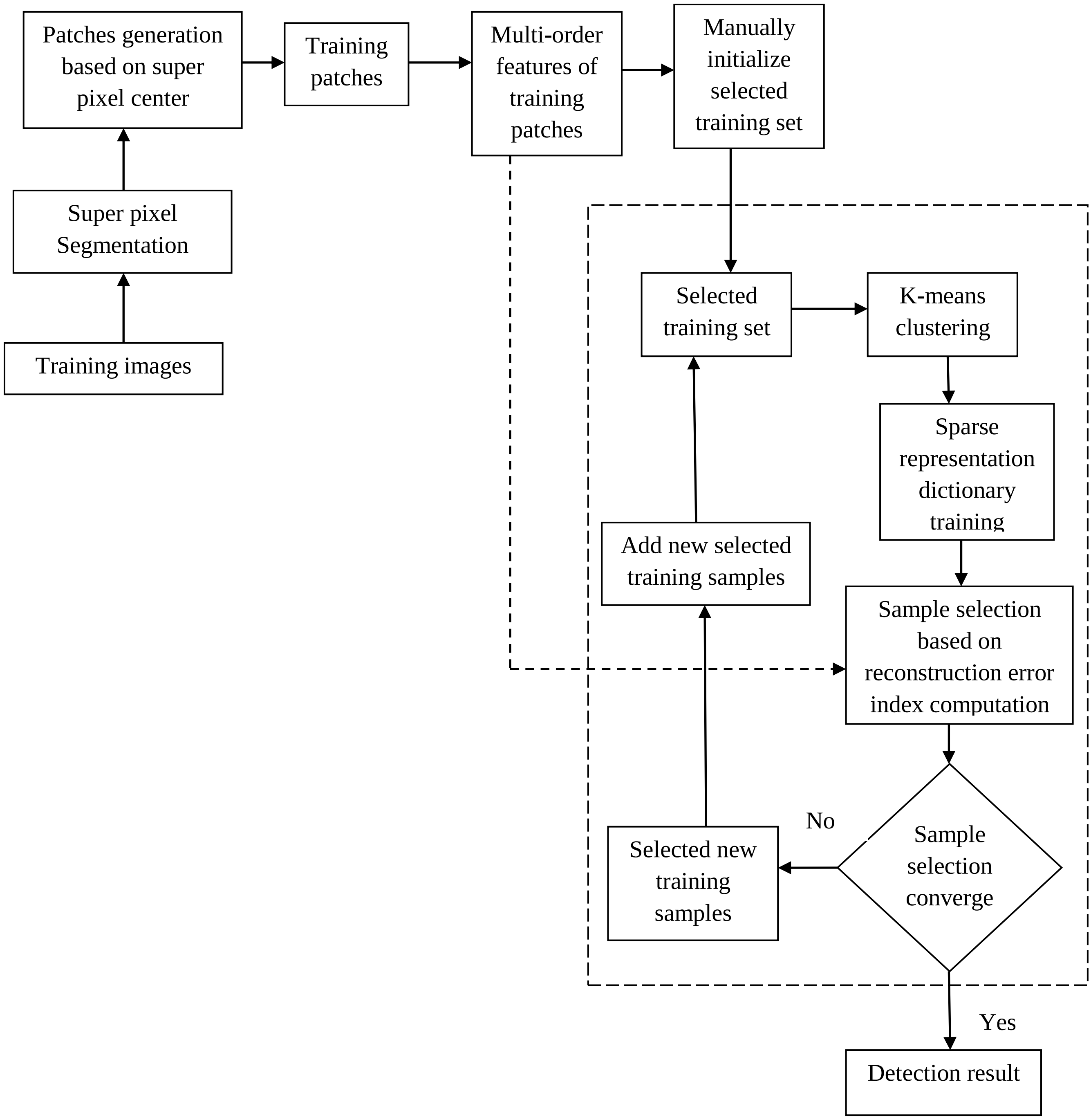

Figure 1. Vehicle Detection Framework

Color information is an essential part in vehicle detection. Color information comprises three channels per pixel in high dimensional feature. A multi-order feature extraction method is introduced to employ the contextual information, texture information and color information for vehicle detection in aerial images. The designed feature descriptor comprises two parts, namely low-order feature and high-order feature. Color and texture information are employed to construct the low-order feature. An original RGB colors are converted into the color name space to minimize the dimension of RGB color information for explaining the colors in semantics. HOG feature grids are employed to gather the texture feature. The color name feature is inserted into HOG feature to maintain the spatial information. The high-order feature is extracted depending on the collection of pre-trained sparse dictionaries that comprises the representative classes in aerial images. With pre-trained dictionaries, reconstruction errors of patches are around examining target in every dictionary. The changes of reconstruction errors between target patch and around patches are employed as high order feature.

Training samples are essential part in object detection through the machine learning techniques. Because of complex back-ground, there are many redundant negatives in vehicle detection from the aerial images. A training subset is chosen in random manner for training. A sample selection method is introduced to choose the representative training samples depending on fast sparse representation classification method.

Vehicle detection framework is introduced depending on super pixel segmentation performances and estimated patch orientations as described in figure 1.

The training images are divided into super pixels depending on centers creating the training patches. The multi-order features of created patches are taken out. The dozens of positives and negatives initialize the small sized training set. K-means method partitions the clustered positives and negatives. The sparse representation dictionaries were trained depending on the positive and negative classes. After initialization of sparse representation dictionaries, sample selection iteration is carried out to create complete training subset where set of complete sparse representation dictionaries are trained for vehicle detection with better accuracy.

Hyper spectral image classification based on volumetric texture and dimensionality reduction

A new volumetric texture and reduced-spectral feature approach is introduced for hyper spectral image classification. The volumetric textural features were extracted by VGLCM. The spectral features were taken out by minimum estimated abundance covariance (MEAC) and linear prediction (LP)-based band selection and semi-supervised k-means (SKM) clustering method with deleting worst cluster (SKMd) band-clustering methods. The four feature combination scheme was introduced for hyper spectral image classification through spectral and textural features as described in figure 2.

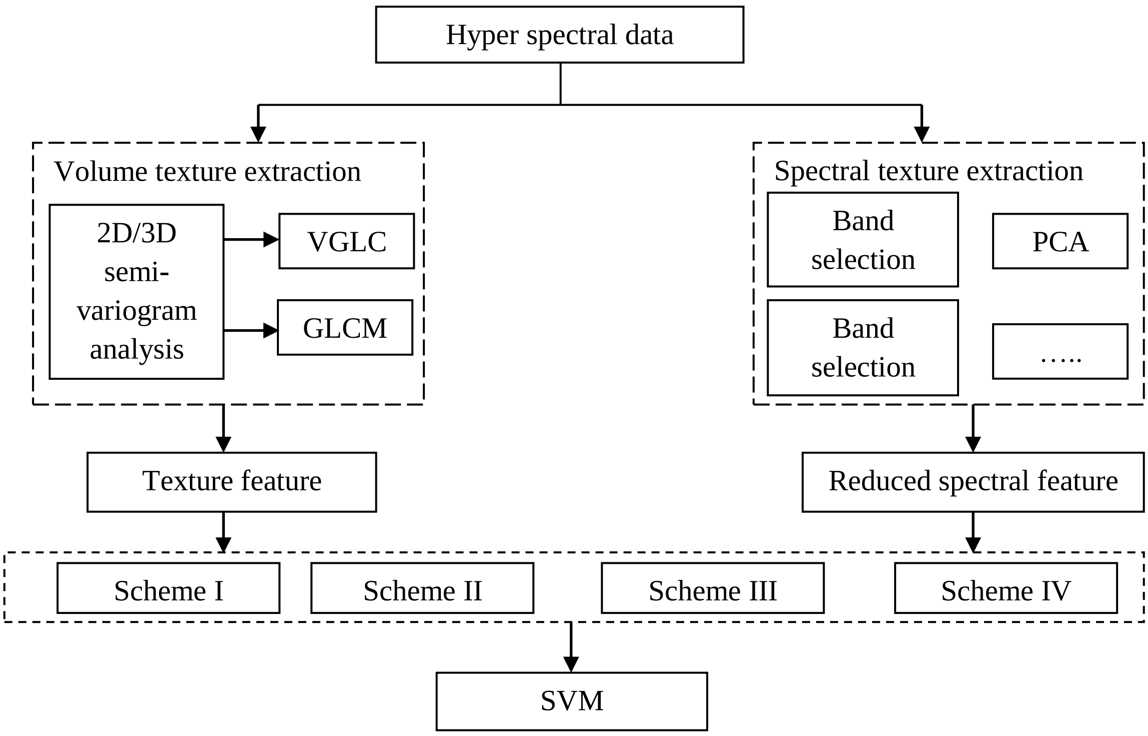

Figure 2. Classification Framework for Hyper spectral Image Analysis

The volumetric textural features were taken out by VGLCM algorithm and dimension-reduced spectral features were attained through MEAC-based band selection and SKMd-based band-clustering algorithm. In spectral-textural fusion schemes, the chosen band, clusters, PCA components and texture features are combined for classification. Every feature from VGLCM is taken as new band where the features are fused by vector-stacking method. The fused feature vectors are taken as inputs for support vector machines. Four spectral–textural fusion techniques are introduced to authenticate that texture features improves the performance of classification accuracy. The texture features are combined with all bands of the original data. In addition, the texture features are fused with principle components of original data after PCA compression. The texture features are combined with the selected bands from original data. The band-clustering centroids are combined with the texture features.

PERFORMANCE ANALYSIS OF AERIAL IMAGE CLASSIFICATION TECHNIQUES

In order to compare the classification performance of aerial images, no. of spatio-temporal features is taken to perform the experiment. Various parameters are employed for improving the classification performance.

Classification Time (CT)

Classification time is defined as the amount of time taken to extract the feature from the aerial images. It is difference of ending time and starting time of feature extraction. It is measured in terms of milliseconds (ms). Feature extraction time is formulated as,

$$CT = Ending\ time - Starting\ time\ of\ aerial\ image\ classification \quad (1)$$

From (1), the classification time is measures. When the classification time is lesser, the method is said to be more efficient.

Table 1. Tabulation for Classification Time

Number of aerial images (Number)	Classification Time (ms)		
	VGG-Net model	Fast Sparse Representation Classification Method	Volumetric Texture and Reduced-Spectral Feature Approach
10	15	21	25
20	17	25	29
30	22	28	34
40	28	33	38
50	21	27	32
60	16	23	26
70	24	30	34
80	28	35	40
90	32	39	46
100	35	44	50

Table 1 describes the classification time with respect to number of aerial images ranging from 10 to 100. Classification time comparison takes place on existing visual geometry group network (VGG-Net) model, Fast Sparse Representation Classification Method and Volumetric Texture and Reduced-Spectral Feature Approach. From table value, it is clear that the classification time using VGG-Net model is lesser when compared to Fast Sparse Representation Classification Method and Volumetric Texture and Reduced-Spectral Feature Approach. The graphical representation of classification time is shown in figure 3.

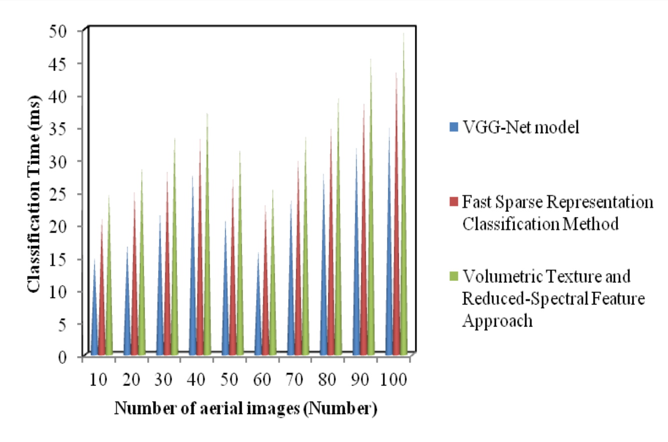

Figure 3. Measure of Classification Time

From figure 3, classification time for different number of aerial images is illustrated. It is clear that classification time of VGG-Net model is lesser than Volumetric Texture and Reduced-Spectral Feature Approach and Fast Sparse Representation Classification Method. This is because VGG-Net model extract informative features from original VHR images. Fully connected layers are employed through VGG-Net where each layer is considered as the separate feature descriptors. DCA is employed as the feature fusion plan to refine the original features extracting from the VGG-Net that allow fusion approach with lesser cost than feature fusion strategies. The classification time consumption of VGG-Net model is 23% lesser than Fast Sparse Representation Classification Method and 34% lesser than Volumetric Texture and Reduced-Spectral Feature Approach.

Feature Extraction Accuracy (FEA)

Feature extraction accuracy is defined as the ratio of number of features that are correctly classified to the total number of features. It is measured in terms of percentage (%). Feature extraction accuracy is formulated as,

$$FEA = \frac{Number\ of\ features\ that\ are\ correctly\ classified}{Total\ number\ of\ features} \quad (2)$$

From (2), the feature extraction accuracy is calculated. When the feature extraction accuracy is higher, the method is said to be more efficient.

Table 2. Tabulation for Feature Extraction Accuracy

Number of features (Number)	Feature Extraction Accuracy (%)		
	VGG-Net model	Fast Sparse Representation Classification Method	Volumetric Texture and Reduced-Spectral Feature Approach
10	75	84	78
20	79	87	84
30	77	85	78
40	80	88	82
50	82	91	85
60	85	93	88
70	83	90	86
80	81	88	83
90	84	92	87
100	88	95	90

Table 2 explains the feature extraction accuracy with respect to number of features ranging from 10 to 100. Feature extraction accuracy comparison takes place on existing visual geometry group network (VGG-Net) model, Fast Sparse Representation Classification Method and Volumetric Texture and Reduced-Spectral Feature Approach. From the table, it is observed that the feature extraction accuracy using Fast Sparse Representation Classification Method is higher when compared to VGG-Net model and Volumetric Texture and Reduced-Spectral Feature Approach. The graphical representation of feature extraction accuracy is shown in figure 4.

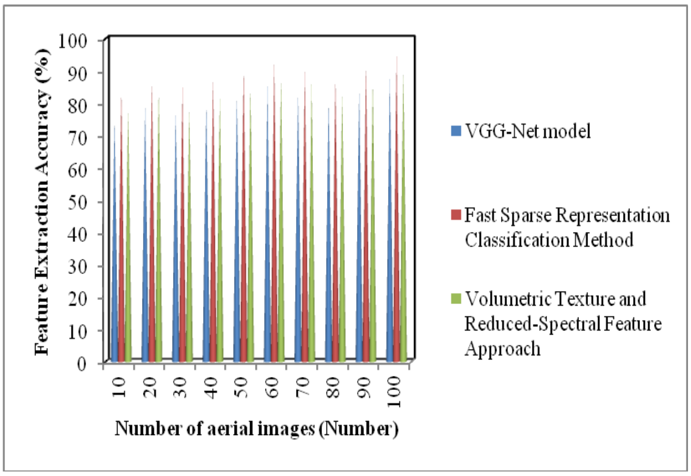

Figure 4. Measure of Feature Extraction Accuracy

From figure 4, feature extraction accuracy for different number of aerial images is explained. It is clear that feature extraction accuracy of Fast Sparse Representation Classification Method is higher than VGG-Net model and Volumetric Texture and Reduced-Spectral Feature Approach. This is because the training images are divided into super pixels depending on the centers of training patches. The multi-order features of created patches are taken out. The dozens of positives and negatives initialize small training set. A sample selection method selects the representative training samples based on the fast sparse representation classification method. The feature extraction accuracy of Fast Sparse Representation Classification Method is 10% lesser than VGG-Net model and 6% lesser than Volumetric Texture and Reduced-Spectral Feature Approach.

False Positive Rate

False Positive Rate is defined as the ratio of number of aerial images is incorrectly classified to the total number of aerial images. It is measured in terms of percentage (%). False Positive Rate is formulated as,

$$FPA = \frac{Number\ of\ aerial\ images\ that\ are\ incorrectly\ classified}{Total\ number\ of\ aerial\ images} \quad (3)$$

From (3), the false positive rate is calculated. When the false positive rate is lesser, the method is said to be more efficient.

Table 3. Tabulation for False Positive Rate

Number of features (Number)	False Positive Rate (%)		
	VGG-Net model	Fast Sparse Representation Classification Method	Volumetric Texture and Reduced-Spectral Feature Approach
10	21	16	12
20	24	19	15
30	29	22	18
40	32	26	20
50	28	24	17
60	24	21	14
70	29	25	18
80	33	29	21
90	37	32	24
100	42	36	27

Table 3 describes the false positive rate with respect to number of aerial images ranging from 10 to 100. False positive rate comparison takes place on existing visual geometry group network (VGG-Net) model, Fast Sparse Representation Classification Method and Volumetric Texture and Reduced-Spectral Feature Approach. From the table, it is clear that the false positive rate using Volumetric Texture and Reduced-Spectral Feature Approach is lesser when compared to VGG-Net model and Fast Sparse Representation Classification Method. The graphical representation of false positive rate is shown in figure 5.

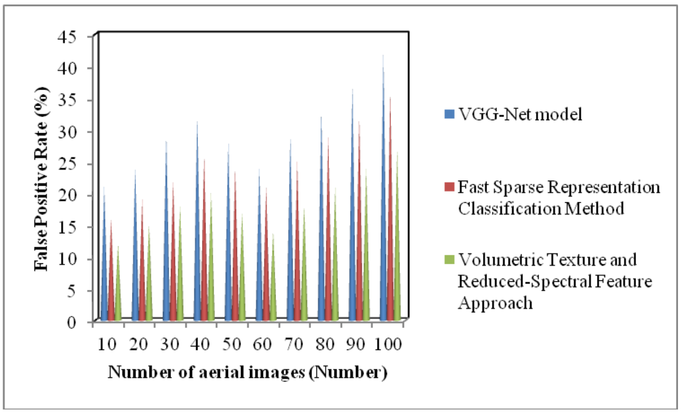

Figure 5. Measure of False Positive Rate

From figure 5, false positive rate for different number of aerial images is described. It is clear that false positive rate gets increased or decreased due to excess number of aerial images. False positive rate comparison of Volumetric Texture and Reduced-Spectral Feature Approach is lesser than VGG-Net model and Fast Sparse Representation Classification Method. This is because spectral features were taken out by MEAC, linear prediction (LP)-based band selection and SKM clustering method with deleting worst cluster (SKMd) band-clustering methods. The texture features are joined with bands of original data. The texture features are combined with principle components of original data after PCA compression. The texture features are combined with the selected bands from original data. The false positive rate of Volumetric Texture and Reduced-Spectral Feature Approach is 38% lesser than VGG-Net model and 26% lesser than Fast Sparse Representation Classification Method.

DISCUSSION AND LIMITATION ON AERIAL IMAGE CLASSIFICATION TECHNIQUES

VGG-Net model is an effective fusion approach with lesser cost. Feature fusion technique improves the accuracy by means of raw deep features. DCA is taken as the feature fusion strategy to refine the original features extracting from VGG-Net. The designed model utilizes DCA fusion that creates good quality informative features to portray the images scene with minimal dimension. But the discriminant correlation analysis creates predicted probabilities in the range [0, 1].

Using dense matching, the photogrammetric point clouds are created. Gradient image depending on photogrammetric point cloud are essential features for detecting the elevated urban objects that includes occlusion performances. The designed classification technique is based on exact, dense and entire point clouds. Volumetric texture and reduced-spectral features approach failed to improve the performance through enhanced dense matching technique.

A fast sparse representation classification method improves the performance of detection efficiency in high-resolution aerial images depending on information of texture, color, and high-order context. With help of trained small dictionaries, detection efficiency and detection accuracy are improved simultaneously. But, the sparse representation classification method was not used for large database.

RELATED WORKS

A new car detection and localization framework was designed in [9] with better precision rate at specified recall. The designed framework employed sliding-window approach with four stages like evaluation, extraction and encoding, classification, and post processing. But, the classification accuracy was not enhanced by means of sliding-window approach. A generic iterative enhancement process was employed in [10] using the partial differential equation termed recurrent neural network (RNN). However, the designed process employed the descent downhill problem with lesser classification accuracy using RNN.

A comprehensive review of up-to-date algorithms and new large-scale benchmark data set of aerial images (AID) was carried out in [11] for performing the aerial scene classification task. But, the aerial scene classification task was not performed with higher accuracy by using AID. A machine learning based framework for random forest classification was introduced in [12] for mapping the lava flows of tropical volcanoes from single image through evaluating the pixel-based versus object-based mapping approaches. But, random forests had been monitored to over fit with noisy classification/regression tasks.

CONCLUSION

A comparison of different existing aerial image classification techniques is studied. From the survival study, it is clear that the existing techniques failed to improve the classification performance. The review explains that the existing volumetric texture and reduced-spectral features approach failed to improve the performance by using enhanced dense matching technique. In addition, the sparse representation classification method was not employed for the large dataset. The wide range of experiments on existing methods describes the performance of many scene classification techniques with its limitations. Finally, from the result, the research work can be carried out using machine learning techniques for improving the performance of scene classification from the aerial images.

REFERENCES

[1] Souleyman Chaib, Huan Liu, Yanfeng Gu, and Hongxun Yao, “Deep Feature Fusion for VHR Remote Sensing Scene Classification”, IEEE Transactions on Geoscience and Remote Sensing, Volume 55, Issue 8, August 2017, Pages 4775 – 4784

[2] Hongjun Su, Yehua SHENG3, Peijun Du, Chen Chen, Kui Liu, “Hyper spectral image classification based on volumetric texture and dimensionality reduction”, Frontiers of Earth Science, Springer, Volume 9, Issue 2, June 2015, Pages 225–236

[3] Ziyi Chen, Cheng Wang, IEEE, Huan Luo, HanyunWang, Yiping Chen, Chenglu Wen, Yongtao Yu, Liujuan Cao and Jonathan Li, “Vehicle Detection in High-Resolution Aerial Images Based on Fast Sparse Representation Classification and Multiorder Feature”, IEEE Transactions on Intelligent Transportation Systems, Volume 17, Issue 8, August 2016, Pages 2296 – 2309

[4] Nathaniel Williams, Duncan Quincey and John Stillwell, “Automatic Classification of Roof Objects from Aerial Imagery of Informal Settlements in Johannesburg”, Applied Spatial Analysis and Policy, Springer, Volume 9, Issue 2, June 2016, Pages 269–281

[5] Jiann-Yeou Rau, Jyun-Ping Jhan, and Ya-Ching Hsu, “Analysis of Oblique Aerial Images for Land Cover and Point Cloud Classification in an Urban Environment”, IEEE Transactions on Geoscience And Remote Sensing, Volume 53, Issue 3, March 2015, Pages 1304-1319

[6] Aleksej Avramovic and Vladimir Risojevic, “Block-based semantic classification of high-resolution

multispectral aerial images", Signal, Image and Video Processing, Springer, Volume 10, Issue 1, January 2016, Pages 75–84

[7] Deepak Kumar Jain, Surendra Bilouhan Dubey, Rishin Kumar Choubey, Amit Sinhal, Siddharth Kumar Arjaria, Amar Jain and Haoxiang Wang, "An Approach for Hyper spectral Image Classification by Optimizing SVM using Self Organizing Map", Journal of Computational Science, Elsevier, July 2017, Pages 1-14

[8] Adriana Romero, Carlo Gatta, and Gustau Camps-Valls, "Unsupervised Deep Feature Extraction for Remote Sensing Image Classification", IEEE Transactions on Geoscience and Remote Sensing, Volume 54, Issue 3, March 2016, Pages 1349-1362

[9] Mohamed ElMikaty and Tania Stathaki, "Detection of Cars in High-Resolution Aerial Images of Complex Urban Environments", IEEE Transactions on Geoscience and Remote Sensing Volume 55, Issue 10, October 2017, Pages 5913-5924

[10] Emmanuel Maggiori, Guillaume Charpiat, Yuliya Tarabalka, and Pierre Alliez, "Recurrent Neural Networks to Correct Satellite Image Classification Maps", IEEE Transactions on Geoscience and Remote Sensing, Volume 55, Issue 10, October 2017, Pages 5913 - 5924

[11] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang and Xiaoqiang Lu, "AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification", IEEE Transactions on Geoscience and Remote Sensing, Volume 55, Issue 7, July 2017, Pages 3965 – 3981

[12] Long Li, Carmen Solana, Frank Canters, Matthieu Kervyn, "Mapping lava flows from satellite remote sensing imagery through random forest classification", Journal of Volcanology and Geothermal Research, Elsevier, 2017, Pages 1-39