IJRSET DECEMBER 2020 Volume 7, Issue 12 Pages: 1-7

ISSN 2394-739X

International Journal for Research in Science Engineering and Technology

SURVIVAL STUDY ON OBJECT DETECTION BASED CLASSIFICATION WITH AERIAL IMAGES

[1] S. Poonkodi, [2] Dr. S. M. Jagatheesan
[1] Research Scholar, [2] Associate Professor
[1, 2] PG & Research Department
[1, 2] Department of Computer Science
[1, 2] Gobi Arts & Science College (Autonomous)
[1, 2] Gobichettipalayam-638453.

ABSTRACT: Aerial imagery is the process of capturing the photos from flying entity. The aerial imagery has several objects each belonging to different classes. Object detection is more demanding one as it includes bounding box around every object in image and allocates into the class label. Object recognition denotes the group of associated tasks for detecting the objects in photographs. Image classification is defined as the process of assigning the class label to every image. Feature plays an essential part in aerial image classification for performing the object detection. Feature extraction includes the extraction of relevant shape information for performing accurate classification by formal process. Classification in aerial images is one of the key issue over last decade which categorize the image based on object exists. But, multiclass classification of objects in aerial images is still difficult one as classification accuracy was not improved. In our research work, many object detection based image classifications are studied and identified the existing problems.

Keywords: [aerial imagery, object detection, object recognition, digital photographs, image classification, shape information.]

1. INTRODUCTION

Classification is a regular arrangement in groups and classes depending on the features. Image classification is attained through differentiating the image into prescribed category based on the objects. Image classification minimized the gap between the computer vision and human vision through training the computer with data. Image classification is an essential process in the object detection and image analysis. The pixel is allocated to definite class when it satisfies rule of class. The classes are considered as known or unknown class. When user separates the classes depending on the training data, classes is considered as known. Otherwise, the class is said to be unknown. Different studies are carried out to conclude the object detection.

This paper is organized as follows: Section 2 explains existing object detection based classification techniques, Section 3 shows the study and analysis of existing object detection based classification techniques, Section 4 identifies the possible comparison between them. Section 5 presents the discussion and limitations of object detection based classification techniques. Conclusion of the paper is described in section 6.

2. LITERATURE SURVEY

A flexible unsupervised feature extraction (FUFE) was introduced in [1] for image classification. But, designed model was not combined with convolutional neural networks for handling the gesture recognition, behavior recognition and other applications. A three-step approach was introduced in [2] to enhance BIQA performance. The feature extraction was carried out to identify the distortion type. The features employed using regression model for forecasting the quality score. However, time consumption for feature selection was not reduced by three step approaches.

A HSI denoising method was introduced in [3] with redundancy and correlation (RAC) in spatial/spectral domains. But, denoising performance gets minimized when noise was strong. Transfer learning with sparse representation was introduced in [4] for classifying the high spatial resolution images. However, the classification accuracy was not improved through transfer learning with sparse representation.

An end-to-end automatic image annotation depending on deep CNN model (E2E-DCNN) was introduced in [5] for feature learning through multiple loss functions. But, classification time was not reduced by end-to-end automatic image annotation model. A new sparse coding model was introduced in [6] to attain the optimal result. However, the complicated background as key factor reduced the classification accuracy.

A partition method was introduced in [7] to analyze spatial information limitations of feature. However, image was not accurately represented and performance was not improved. A multi-scale feature extraction method depending on stacked sparse autoencoder (SSAE) termed multi-scale SSAE (MS-SSAE) was introduced in [8] to increase the classification results. But, the deep architecture like CNN was not employed and failed to examine the use for PolSAR image classification. The designed method failed to learn the deep multi-scale features for PolSAR data.

A Rotated-CorneR Local Binary Pattern (RCRLBP) was introduced in [9] with Contours filter through orthogonal depiction. But, long delay temporal series was not observed to automate the land cover reconstruction fromHSItorical aerial images through multi-spectral images. A bidirectional adaptive feature fusion strategy was introduced in [10] to manage scene classification. The deep learning feature and SIFT feature were combined to attain the discriminative image presentation. However, the error rate during the classification process was not reduced by bidirectional adaptive feature fusion strategy. An end-to-end learning framework was introduced in [11] with the spectral and spatial information. But, the robustness was not improved

3. OBJECT DETECTION BASED IMAGE CLASSIFICATION TECHNIQUES

Object Detection analyzes the image and identifies the objects in it. Object recognition includes the collection of linked tasks for finding the objects. Classification process is depending on the description, texture or similarity of items or things. Image classification considers two processes, namely supervised classification and unsupervised classification. Pixels are unit represented in an image. Image classification groups the pixels in different classes. The image classification includes image acquisition, image pre-processing, image segmentation, feature extraction and classification. Image classification is an essential process in different fields like remote sensing, biomedical images and automation. A classification system comprised the camera placed on interested zone where the images are gathered and processed.

3.1. Flexible unsupervised feature extraction for image classification

The key objective is to recover the low-dimensional representation for collecting the geometric structure hidden in high-dimensional data and make class distribution more apparent one to increase the machine learning results. A FUFE was introduced for image classification process. Principal Component Analysis (PCA) and Locality Preserving Projection (LPP) were representative unsupervised dimensionality reduction methods. The designed method was appropriate for dealing with certain types of nonlinear manifolds to classify the local and global geometric structures. A regression residual term was joined into objective function for low-dimensional data with projected training data by projection matrix. The designed method handle data sampled from nonlinear manifold near linear subspace. The unsupervised technique employed complicated iterative optimization solutions. PCA and LPP were two representative unsupervised models used to classify local and global geometric data structures.

3.2. Distortion-specific feature selection algorithm for universal blind image quality assessment

Blind image quality assessment (BIQA) employed the objective measure for forecasting quality score of distorted images without preceding information. A three-step approach was introduced to enhance the results of BIQA techniques. The feature extraction was carried out through BIQA techniques to determine the distortion type. BIQA techniques extracted the features in spatial and transform domains where model deviation in distorted image features compared with natural images. BIQA techniques extracted the features to perform the feature extraction and classification by regression. The feature forecast the quality score. BIQA techniques improved the correlation of predicted quality score with mean observer score (MOS) and minimized the processing time. Natural Scene Statistics (NSS) assessed the image quality depending on deviation between distorted as well as natural images. The key aim was to employ the generalized approach for distortion specific feature selection.

3.3. Hyperspectral Image Denoising via Sparse Representation and Low-Rank Constraint

HSI denoising method was introduced with the global and local RAC (redundancy and correlation) in spatial/spectral domains. Image patches of spectral bands approximated through dictionary learned from noisy features. The sparse coding was developed to model the global RAC in spatial domain and local RAC in spectral domain. Local RAC in spectral domain resulted in spectral distortion. Low-rank problems were addressed with global RAC in spectral domain to compensate the needs of local spectral RAC. The low rank of HSI was incorporated as supplementary regularization term. The global RAC in spectral dimension was developed for HSI to reduce spectral distortion. The regularization addressed ill-posed denoising issues. In addition, the error gets minimized through enforcing the low rank on denoised data for sparse coding and dictionary learning.

4. PERFORMANCE ANALYSIS ON OBJECT DETECTION FOR IMAGE CLASSIFICATION

The different object detection techniques are compared with varying number of aerial images. The object detection performance is determined by using various parameters. An experimental evaluation is implemented using MATLAB software with Caltech 101 dataset http://www.vision.caltech.edu/Image_Datasets/Caltech101/. Caltech 101 dataset comprises 9144 images from 102 classes (i.e., 101 object classes and background class). Caltech 101 dataset include pizza, umbrella, watch, dolphin, and so on. The number of images of per class changes from 31 to 800. The vector quantization codes are pooled to form feature in every spatial subregion of spatial pyramid. Every class is randomly selected 20 images as training samples and rest is considered as the test samples. In our work, aerial image is considered as an input. The classification performance gets enhanced by utilizing the various parameters, namely

- Object Recognition Rate
- Object Detection Time
- Error Rate

4.1. Object Recognition Rate

Object recognition rate (OBR) is defined as the ratio of number of aerial images that are correctly recognized and classified to the total number of aerial images. It is given by,

$$OBR = \frac{\text{Number of aerial images that are correctly recognized and classified}}{\text{Total number of aerial images}} \quad (1)$$

From (1), the object recognition rate is determined. When the objection recognition rate is higher, the method is said to be more efficient.

Number of aerial Images	Object Recognition Rate		
	FUFE Model	Three-step approach	HSI denoising method
20	0.48	0.32	0.28
40	0.54	0.35	0.31
60	0.54	0.38	0.33
80	0.55	0.42	0.35
100	0.58	0.43	0.36
120	0.59	0.44	0.38
140	0.59	0.45	0.39
160	0.58	0.44	0.39
180	0.57	0.44	0.38
200	0.57	0.44	0.38

Table 1 Tabulation for Object Recognition Rate

Table 1 explains the object recognition rate with respect to number of aerial images ranging from 20 to 200. Object recognition rate comparison takes place on flexible unsupervised feature extraction (FUFE) model, three-step approach and Hyper spectral image (HIS) denoising method. The graphical illustration of object recognition rate is described in figure 1.

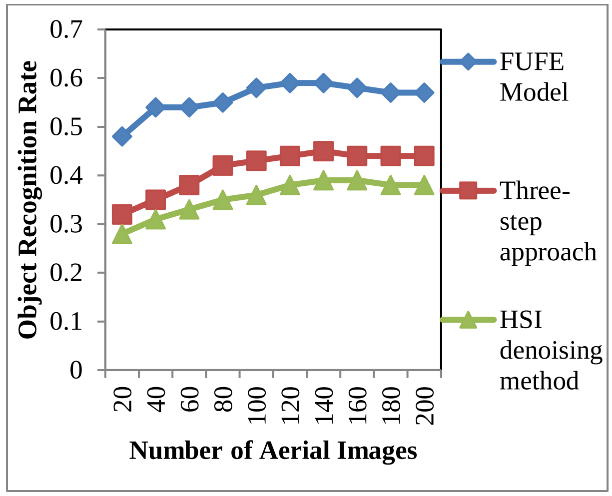

Figure 1 Measurement of Object Recognition Rate

Figure 1 describes the object recognition rate performance for different number of aerial images. From figure, it is clear that the object recognition rate using FUFE model is higher when compared to three-step approach and Hyper spectral image (HIS) denoising method. This is due to application of PCA and LPP to classify local and global geometric data structures. A regression residual term combined with reformulated objective function to implement the low-dimensional data representation. Research in FUFE model has 37% higher object recognition rate than three-step approach and 58% higher object recognition rate than Hyper spectral image (HSI) denoising method.

4.2. Object Detection Time

Object detection time (ODT) is defined as the amount of time taken to detect the object from the aerial images. It is defined as the difference of starting time and ending time of object detection from input aerial images. It is measured in terms of milliseconds. It is given by,

$$\text{ODT} = \text{Ending time} - \text{Starting time of object detection} \quad (2)$$

From (2), object detection time is determined.

Number of aerial Images	Object Detection Time (ms)		
	FUFE Model	Three-step approach	HSI denoising method
20	44	25	38
40	47	27	40
60	50	30	43
80	52	32	45
100	54	33	48
120	57	35	51
140	60	41	56
160	63	43	59
180	66	46	62
200	68	48	64

Table 2 Tabulation for Object Detection Time

Table 2 describes the object detection time with respect to number of aerial images ranging from 20 to 200. Object detection time comparison takes place on flexible unsupervised feature extraction (FUFE) model, three-step approach and Hyper spectral image (HIS) denoising method. The graphical representation of object detection time is illustrated in figure 2.

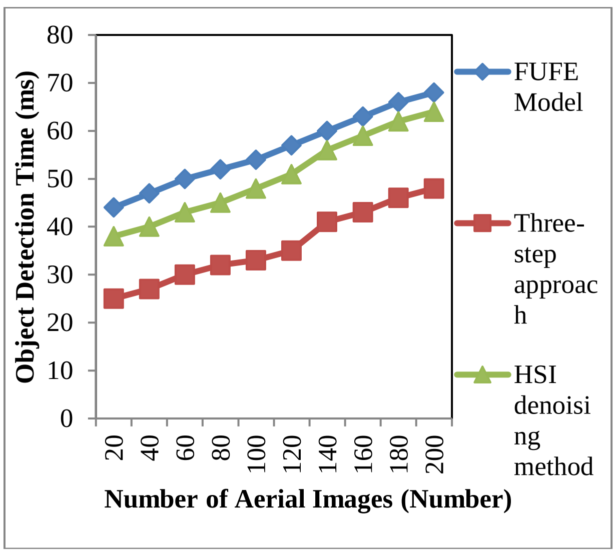

Figure 2 Measurement of Object Detection Time

Figure 2 illustrates the object detection time performance for different number of aerial images. From figure 2, it is clear that the object detection time using three-step approach is lesser when compared to FUFE model and Hyper spectral image (HSI) denoising method. This is due to the application of BIQA techniques for performing the feature selection. The distortion-specific features forecasted the quality score through regression model. BIQA techniques enhanced the correlation of predicted quality score and reduced the processing time. Research in three-step approach reduces the error rate 36% than FUFE model and 29% than Hyper spectral image (HSI) denoising method.

4.3. Error Rate

Error rate (ER) is defined as the ratio of number of aerial images that are incorrectly recognized and classified to the total number of aerial images. It is measured in terms of percentage (%). It is formulated as,

$$ER = \frac{\text{Number of aerial images that are incorrectly classified}}{\text{Total number of aerial images}} * 100 \quad (3)$$

From (3), the error rate is computed. When the error rate is lesser, the method is said to be more efficient.

Number of aerial Images	Error Rate (%)		
	FUFE Model	Three-step approach	HSI denoising method
20	21	29	16
40	23	32	18
60	25	34	21
80	28	36	23
100	30	39	26
120	32	41	29
140	35	43	32
160	36	45	35
180	39	48	36
200	42	50	38

Table 3 Tabulation for Error Rate

Table 3 portrays the error rate with respect to number of aerial images ranging from 20 to 200. Error rate comparison takes place on flexible unsupervised feature extraction (FUFE) model, three-step approach and Hyper spectral image (HIS) denoising method. The graphical representation of error rate is illustrated in figure 3.

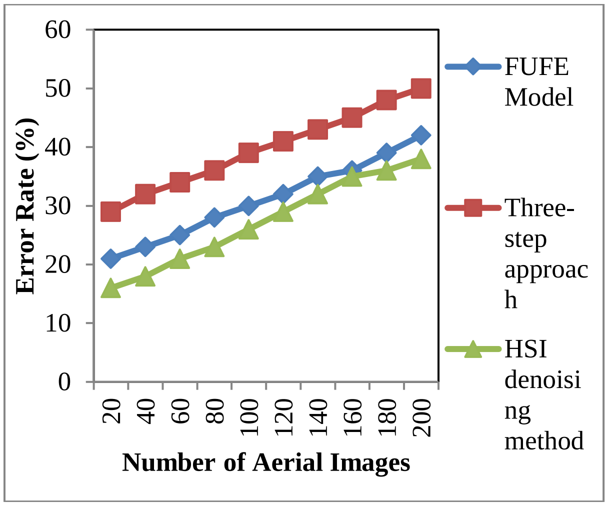

Figure 3 Measurement of Error Rate

From figure 3, error rate for different number of aerial images is described. From figure 3, it is observed that the error rate using Hyper spectral image (HSI) denoising method is lesser when compared to FUFE model and three-step approach. This is because of using local/global redundancy and correlation in spatial/spectral domains. Local RAC and global RAC in spectral dimension minimized the spectral distortion for HSI. Through determining low rank on denoised data, error rate gets reduced. Research in Hyper spectral image (HSI) denoising method reduces the error rate 13% than FUFE model and 32% than three-step approach.

5. DISCUSSION AND LIMITATIONS OF OBJECT DETECTION BASED IMAGE CLASSIFICATION TECHNIQUES

FUFE method was introduced for image classification. In addition, designed method was suitable for managing the nonlinear manifolds and classifying the local as well as global geometric structures. However, designed model was not combined with the convolutional neural networks for handling gesture recognition and behavior recognition. Three-step approach improved the blind image quality assessment performances. Feature extraction decided the distortion type. Features were chosen for all distortion types depending on the correlation constant. However, feature selection time consumption was not reduced through three step approach. HSI denoising method was introduced for RAC in spatial/spectral domains. The rank of noise-free HSI was added as regularization to develop global RAC in spectral dimension. But, denoising performance gets reduced when noise was strong.

5.1. Future Direction

The future direction of the object detection based image classification can be carried out using machine learning techniques with higher accuracy and lesser time consumption.

CONCLUSION

A comparison of different object detection based image classification techniques is studied. From the study, it is clear that the existing techniques reduce the denoising performance when the noise was strong. The survival review shows that the existing three step approach not reduced feature selection time. In addition, convolutional neural networks were not combined for handling the gesture recognition and behavior recognition. The wide range of experiments on existing methods determines the performance of the many object detection techniques with its drawbacks. Finally from result, the research work can be carried out using machine learning techniques for enhancing accuracy and time consumption performance during object detection.

REFERENCES

[1] Yang Liu, Feiping Nie, Quanxue Gao, Xinbo Gao, Jungong Han and Ling Shao, “Flexible unsupervised feature extraction for image classification”, Neural Networks, Elsevier, Volume 115, July 2019, Pages 65-71

[2] Imran Fareed Nizami, Muhammad Majid, Waleed Manzoor, Khawar Khurshi and Byeungwoo Jeon, "Distortion-specific feature selection algorithm for universal blind image quality assessment", EURASIP Journal on Image and Video, Springer, Volume 19, 2019, Pages 1-13
[3] Yong-Qiang Zhao, Member, IEEE, and Jingxiang Yang, "Hyperspectral Image Denoising via Sparse Representation and Low-Rank Constraint", IEEE Transactions on Geoscience and Remote Sensing, Volume 53, Issue 1, January 2015, Pages 296-308
[4] Eliezer Flores, Maciel Zortea and Jacob Scharcanski, "Dictionaries of deep features for land-use scene classification of very high spatial resolution images", Pattern Recognition, Elsevier, Volume 89, 2019 Pages 32–44
[5] Xiao Ke, Jiawei Zou, and Yuzhen Niu, "End-to-End Automatic Image Annotation Based on Deep CNN and Multi-Label Data Augmentation", IEEE Transactions on Multimedia, Volume 21, Issue 8, August 2019, Pages 2093 – 2106
[6] Han Dongmei and Liu Qigang, "Sparse Coding for Image Classification Base on Spatial Pyramid Representation", Pattern Recognition and Image Analysis, Springer, Volume 27, Issue 3, July 2017, Pages 466–472
[7] Tao He, "Image classification based on sparse coding multi-scale spatial latent semantic analysis", EURASIP Journal on Image and Video Processing, Springer, Volume 38, 2019, Pages 1-11
[8] Lu Zhang, Licheng Jiao, Wenping Ma, Yiping Duan and Dan Zhang, "PolSAR Image Classification Based on Multi-scale Stacked Sparse Autoencoder", Neurocomputing, Elsevier, Volume 351, July 2019, Pages 167-179
[9] Remi Ratajczak, Carlos Fernando Crispim-Junior, Elodie Faure, Beatrice Fervers and Laure Tougne, "Automatic Land Cover Reconstruction FromHSItorical Aerial Images: An Evaluation of Features Extraction and Classification Algorithms", IEEE Transactions on Image Processing, Volume 28, Issue 7, July 2019, Pages 3357-3371
[10] Xiaoqiang Lu, Weijun Ji, Xuelong Li and Xiangtao Zheng, "Bidirectional Adaptive Feature Fusion for Remote Sensing Scene Classification", Neurocomputing, Elsevier, Volume 328, February 2019, Pages 135-146
[11] Xu Liu, Licheng Jiao, Jiaqi Zhao, Jin Zhao, Dan Zhang, Fang Liu, Shuyuan Yang and Xu Tang, "Deep Multiple Instance Learning-Based Spatial–Spectral Classification for PAN and MS Imagery", IEEE Transactions on Geoscience and Remote Sensing, Volume 56, Issue 1, January 2018, Pages 461 - 473