# Deep Learning for Satellite Image Classification

Mayar A. Shafaey[1(✉)], Mohammed A.-M. Salem[1,2], H. M. Ebied[1], M. N. Al-Berry[1], and M. F. Tolba[1]

[1] Faculty of Computers and Information Sciences,
Ain Shams University, Cairo, Egypt
mayar.al.mohamed@fcis.asu.edu.eg,
{salem,maryam_nabil}@cis.asu.edu.eg,
hala.m@outlook.com, fahmytolba@gmail.com
[2] Faculty of Media Engineering and Technology,
German University, Cairo, Egypt

**Abstract.** Nowadays, large amounts of high resolution remote-sensing images are acquired daily. However, the satellite image classification is requested for many applications such as modern city planning, agriculture and environmental monitoring. Many researchers introduce and discuss this domain but still, the sufficient and optimum degree has not been reached yet. Hence, this article focuses on evaluating the available and public remote-sensing datasets and common different techniques used for satellite image classification. The existing remote-sensing classification methods are categorized into four main categories according to the features they use: manually feature-based methods, unsupervised feature learning methods, supervised feature learning methods, and object-based methods. In recent years, there has been an extensive popularity of supervised deep learning methods in various remote-sensing applications, such as geospatial object detection and land use scene classification. Thus, the experiments, in this article, carried out on one of the popular deep learning models, Convolution Neural Networks (CNNs), precisely *AlexNet* architecture on a standard sounded dataset, *UC-Merceed Land Use*. Finally, a comparison with other different techniques is introduced.

**Keywords:** Remote-sensing · Satellite image · Deep learning
Convolution Neural Networks (CNNs) · UC-Merceed Land Use
Parallel computing

## 1 Introduction

A Satellite Image is an image of the whole or part of the earth taken using artificial satellites. It can either be visible light images, water vapor images or infrared images [1]. The different types of satellites produce (high spatial, spectral, and temporal) resolution images that cover the whole Earth in less than a day. The large-scale nature of these data sets introduces new challenges in image analysis.

The analysis and classification of remote-sensing images is very important in many practical applications, such as natural hazards and geospatial object detection, precision

A. E. Hassanien et al. (Eds.): AISI 2018, AISC 845, pp. 383–391, 2019.
https://doi.org/10.1007/978-3-319-99010-1_35

agriculture, urban planning, vegetation mapping, and military monitoring [2]. Despite decades of research, the degree of automation for remote-sensing images analysis still remains low [3].

The main objective of this paper is to present a literature review on the recent deep-learning based techniques for satellite image classification and the available training and testing datasets. Moreover, testing results will present on one popular dataset using the *AlexNet* architecture of the Convolution Neural Networks (CNNs).

In the next section, a list of available datasets and their specifications are presented. A review on recent classification approaches applied on one or some of these datasets is presented in Sect. 3. The experimental work followed by results and evaluations are presented in Sect. 4. Finally, conclusions are highlighted in Sect. 5.

## 2 Review on Publicly Remote Sensing Images Datasets

In the past years, several high resolution remote-sensing image datasets have been introduced by different groups to enable machine-learning based research for scene classification and to evaluate different methods in this field. The authors will review some publicly available sets in this section, as given in Table 1. The table below shows the number of scene classes, images per class, total images, size of images, and spatial resolution.

**Table 1.** Comparison between the different remote-sensing datasets proposed

| Data set | Scene classes | Images/class | Total images | Spatial resolution | Image sizes |
|---|---|---|---|---|---|
| AID [4] | 30 | 200–400 | 10000 | High | 600 × 600 |
| Patter Net [5] | 38 | 800 | 30400 | Up to 0.8 | 256 × 256 |
| RSI-CB256 [6] | 35 | Various | 34000 | 0.3–3 | 256 × 256 |
| SAT_4 & SAT_6 [7] | Patches (500000 + 405000) | | | Low | 28 × 28 |
| UC-Merced Land Use [8] | 21 | 100 | 2100 | 0.3 | 256 × 256 |
| WHU-RS19 [9] | 19 | ~50 | 1005 | Up to 0.5 | 600 × 600 |
| SIRI-WHU [10] | 12 | 200 | 2400 | 2 | 200 × 200 |
| RSSCN7 [11] | 7 | 400 | 2800 | – | 400 × 400 |
| RSC11 [12] | 11 | ~100 | 1232 | 0.2 | 512 × 512 |
| Brazilian Coffee [13] | 2 | 1438 | 2876 | Low | 64 × 64 |
| NWPU-RESISC45 [14] | 45 | 700 | 31500 | ~30–0.2 | 256 × 256 |

The most images in these datasets are imported from Google Earth Engine and cover the areas of: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and so on. Except the dataset in [13] "Brazilian Coffee Scene dataset", cropped from SPOT satellite images and contains only two scene classes, which is appropriate for multi-class scene classification methods. In contradiction, the large number of classes and images in NWPU-RESISC45 [14] dataset, will impact positively the classification results.

However, the UC-Merced Land-Use [8] in Fig. 1 is the most popular and has been widely used for the task of remote-sensing image scene classification and retrieval so far. So, the authors will choose it to carry out the classification experiment.

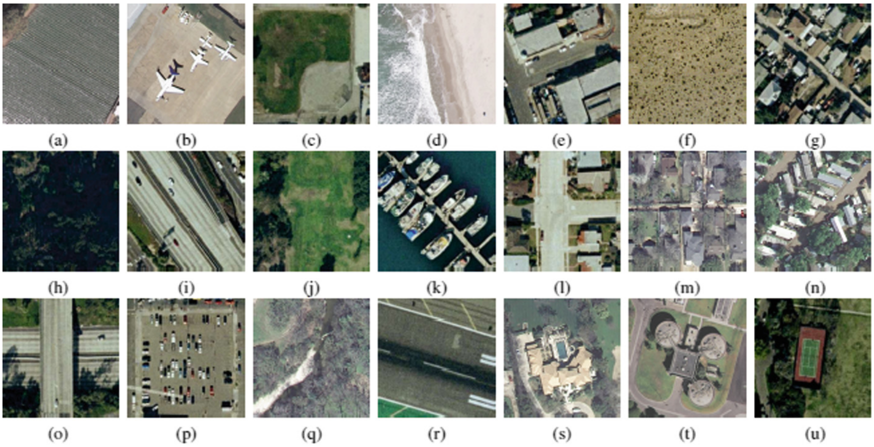

**Fig. 1.** 21 Classes representative [(a)–(u)] of the UC-Merced Land-Use dataset [34].

## 3 Remote Sensing Images Classification Methods

There are long and proud researches during the last and current decades that were carried out on the satellite images for the task of scene classification. From the vast publications of this topic, generally, the existing scene classification methods could summarized into four main categories according to the features they used: **manually feature based methods, unsupervised classification methods, supervised learning methods, and object-based methods**.

### 3.1 Manually Feature Based Methods

A fundamental step in image classification is based on handcrafted features. These methods measure the skills of researchers to design and extract important features, such as color, orientation, texture, shape, spatial and spectral information, or their combination. Some of the most common and essential features that are used for scene classification are: ***Color histograms*** - ***Texture descriptors*** – ***GIST****: describe orientations of a scene* – ***SIFT****: describe sub-regions of a scene* – ***HOG****: describe gradient of objects* [15–17, 40].

## 3.2 Unsupervised Classification Methods

The limitations of manually feature based methods could be solved by self-learning features from images. This strategy is called unsupervised learning method. In recent years, unsupervised feature learning from unlabeled input data has become an attractive alternative to handcrafted features [18].

The idea behind that strategy is first grouping the image pixels into clusters based on their properties. By learning features from images instead of relying on manually designed features, we can obtain more discriminative feature that is better suited for the classification problem [19]. Such clustering algorithms are: principal component analysis (PCA) [20], *k*-means clustering [21], sparse coding [22], and so on.

In real applications, the aforementioned unsupervised feature learning methods have achieved good performance for land use classification, especially compared to handcrafted based methods. For example, authors in [23–25] applied unsupervised methods and made a significant progress for remote-sensing scene classification.

## 3.3 Supervised Learning Methods

Starting year 2006, the volcano of researches relied on supervised learning methods which need to use labeled data to extract more powerful features, especially, a deep learning method which made by *Hinton and Salakhutdinov* [26]. There exists different numbers of deep learning models, such as deep belief nets (DBN) [27], deep Boltzmann machines (DBM) [28], stacked auto-encoder (SAE) [29], Convolutional Neural Networks (CNNs) [30], and so on. In this article, the authors mainly review the widely used deep learning method CNNs.

The basic concept of CNN is to train huge multi-layer networks for giving impressive classification results of large scale input images. The CNN itself has different models like: *AlexNet, GoogleNet, ResNet, VGGNet, CaffeNet* … etc. [31]. Limited by the space, a short and highlight description of *AlexNet* architecture was given. The net consists of 25 layers: 5 convolution layers, max-pooling layers, dropout layers, and 3 fully connected layers, as shown in Fig. 2. It is trained on *ImageNet* data, which contained over 15 million annotated images from a total of over 22,000 categories and Used ReLU for the nonlinearity functions [32].

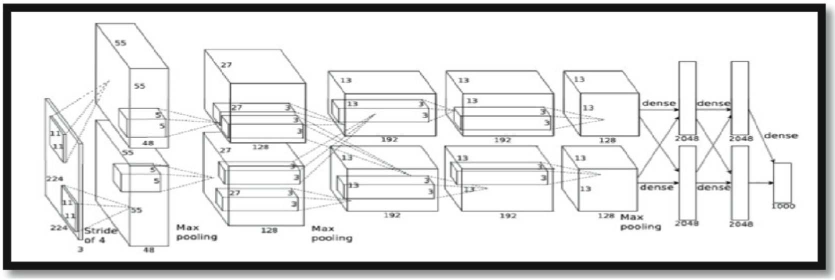

**Fig. 2.** *ImageNet* classification with *AexNet* CNN [32]

Table 2 represents some of authors who used the CNN models in their experiments for large scale image scene classification and gave the proud accuracy values which demonstrate the power of CNN learning model.

**Table 2.** Survey of recent publications applied CNNs in their experiments on large scale remote-sensing (RS) images, *UC-Mercced* dataset

| References | Year | Application | Method | Accuracy |
|---|---|---|---|---|
| [33] | 2015 | Multi-spectral land use classification | Deep CNN | 93.48% |
| [34] | 2015 | Land Use RS classification | GoogleNet, and CaffeNet | 97%, and 95.48% |
| [35] | 2016 | RS scene classification | Large patch CNN | Effective results |
| [36] | 2016 | Large scale image classification | CNN | 92.4% |
| [37] | 2018 | Remote sensing scene classification | CNN | 92.43% |

### 3.4 Object-Based Methods

Unlike pixel-based or image-based classification, object-based image classification groups pixels into representative shapes and sizes and assigns each group to a semantic object. This process relies on multi-resolution segmentation. Multi-resolution segmentation produces homogenous objects by grouping pixels. It generates objects with different scales in an image simultaneously. These objects are more meaningful because they represent features in the image [38, 41].

The question here is how to select the appropriate image classification techniques. It is based on common sense of the engineering. Let's say you want to classify water in a high spatial resolution image containing grasses. You decide to choose all pixels with low NDVI (Normalized Difference Vegetation Index) in that image. NDVI is used to analyze remote sensing measurements and assess whether the target being observed contains live green vegetation or not. But this could also misclassify other pixels in the image that aren't water i.e. pixels of the sky. For this reason, pixel-based classification as unsupervised and supervised classification gives a salt and pepper look.

As illustrated in this article, spatial resolution is an important factor when selecting image classification techniques. Hence, when you have low spatial resolution, both traditional pixel-based and object-based image classification techniques perform well. But when you have high spatial resolution, object-based image classification is superior to traditional pixel-based classification [39].

## 4 Experiments and Results

Taking advantages of the availability of *UC-Merceed* dataset [8], the *AlextNet* CNN approach was applied to represent the large scale image classification process. In this section, the experiment's steps will be described, i.e., software, hardware specification, results, comments, and comparisons.

### 4.1 Experimental Procedure

The experiment ran on two different computers. Machine 1 has a processor: Intel® Core™ i7-2670QM CPU @ 2.20 GHz–8 GB RAM. Machine 2 equipped with NVIDIA GTX 1050 4G cc: 6.1 GPU: Intel® Core™ i7-7700HQ @ 2.20 GHz–16 GB RAM. The time elapsed on machine 1 was 1800 s and on machine 2 was 14 s. Thanks to Graphical Processing Unit (GPU) for giving an impressive and significant execution time. The parallel computing optimizes the performance 100 times than the serial computations.

Hence, The experiment ran on Machine 2 and Matlab® software using *alexnet()* built-in function which is trained on a subset of the *ImageNet* database – ~1.2 million images - and can classify images into 1000 object categories. This function requires Neural Network Toolbox™ Model for *AlexNet* Network. The basic three steps are firstly resizing the image dimension from *256* × *256* to *227* × *227* as a required input for the CNN. The second step is to choose the training set percentage. And thirdly, train the multiclass SVM classifier, extract test features using the CNN, and pass them to the trained classifier to get the known labels. Finally, the classification results are given by computing the summation of main diagonal of the confusion matrix divided by the diagonal elements number.

A number of experiments were carried out to assess the performance of the CNN using the well-known *UC-Merceed* Land Use dataset. The *UC-Merceed* Land Use dataset contains 2100 images, 21 distinct classes and every class contains 100 different images. In the experiments, the size of training set ranged from 10 to 90% of the 100 different images per class and the remaining images where used for testing. Figure 3 shows the correct classification accuracy vs. the size of training set percentage.

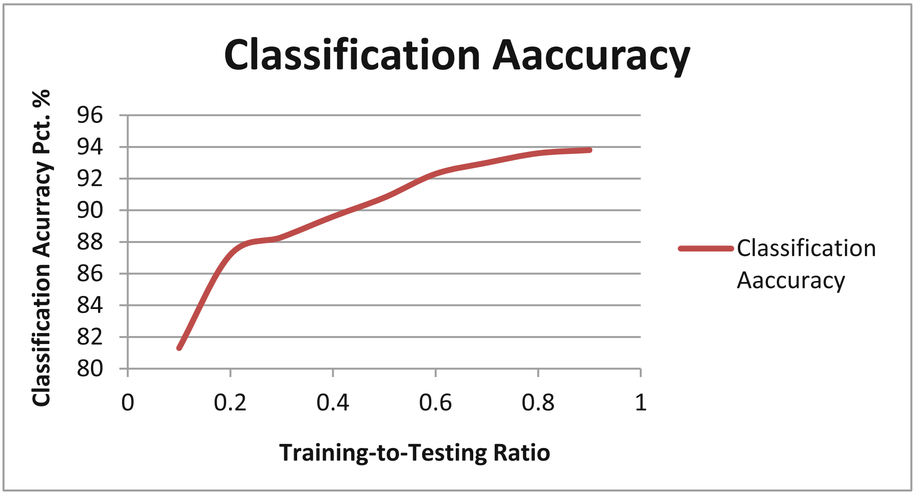

**Fig. 3.** The classification accuracy for *UC-Merceed* Land Use dataset using the *AlexNet* CNN. The *x-axis* represents the interval of training to testing set ratio [0.1–0.9]. The *y-axis* represents the classification accuracy.

The first trial started to split 10% of images into training set which gave 81.3% accuracy value. Then, repeated the experiment eight times up to 90% of images into training set which gave around 94% accuracy value. The figure below illustrates that the gradually increase of training images impacts positively the classification result.

### 4.2 Evaluation and Discussions

Compared with other CNN models, *GoogleNet* and *CaffeNet*, mentioned and discussed in [34], and applied also on UC-Merceed dataset, the authors observed that the classification accuracy gained by *GoogleNet* (~97%) was better than whose gained by *CaffeNet* (~94%) and *AlexNet* (~94%). However, The *AlexNet* is faster than *GoogleNet* model. The two models ran on the same GPU, as mentioned before, the *AlexNet* executed in only 14 s, but *GoogleNet* consumed 51 s, which is approximately 4 times slower.

On the one hand, in comparison with traditional handcrafted features that require a high mental thinking and skills, deep learning features are learned from data automatically via deep architecture neural networks. This is the key advantage of deep learning methods.

On the other hand, and compared with aforementioned unsupervised feature learning methods i.e. sparse coding, deep learning models can learn more powerful because it is composed of multiple processing layers which is more applicable for large scale and remote-sensing image scene classification. The deep feature learning methods act as a human brain in which every level uses the information from the previous level to learn deeply and accurately.

The following articles support our research. In [23], the high-resolution satellite scene classification using a sparse coding carried out on *UC-Merceed* dataset and reached about 91% accuracy. And in [24], the unsupervised feature learning via spectral clustering of multidimensional patches was carried out on the same dataset and achieved 90% right classification.

## 5 Conclusions

The automation target detection or recognition, and high resolution remotely sensed image classification are two hot topics nowadays. Hence, this paper firstly represented a comprehensive review of common and freely remote-sensing datasets to enable the community to develop the large scale image scene classification task. Then, it gave a summary of recent methods used for this task. Finally, the CNN deep learning method applied on *UC-Merceed* dataset evaluated and reported the results to compare against state-of-the-art and as a baseline for future research.

Deep learning methods can undoubtedly offer better feature representations for the related remote-sensing task, and there is a bright prospect of seeing more and more researchers dedicated to learning better features for the target detection and scene classification tasks by utilizing appropriate deep learning methods.

Thanks to parallel computing and GPUs for optimizing and enhancing the execution time 100× than the serial computations, our experiment ran in time not exceeding 14 s to classify one testing image out of 2100 images.

## References

1. NASA: What Is a Satellite? NASA Knows! (Grades 5–8) series (2014)
2. Zhang, L., Xia, G., Wu, T., Lin, L., Tai, X.: Deep learning for remote sensing image understanding. J. Sens. **2016**, 1–2 (2016)
3. Marmanisad, D., Wegnera, J., Gallianib, S., Schindlerb, K., Datcuc, M., Stillad, U.: Semantic segmentation of aerial images with an ensemble of CNNs. ICWG **3**(4), 1–8 (2016)
4. AID Dataset. http://www.lmars.whu.edu.cn/xia/AID-project.html. Accessed 16 Feb 2018
5. PatternNet Dataset. https://sites.google.com/view/zhouwx/dataset?authuser=0. Accessed 16 Feb 2018
6. RSI Dataset. https://github.com/lehaifeng/RSI-CB. Accessed 16 Feb 2018
7. SAT_4 & SAT_6. http://csc.lsu.edu/~saikat/deepsat/. Accessed 16 Feb 2018
8. UC-Merceed Land Use Dataset. http://weegee.vision.ucmerced.edu/datasets/landuse.html. Accessed 16 Feb 2018
9. WHU-RS19 Dataset. https://www.google.com/url?q=http%3A%2F%2Fwww.xinhua-fluid.com%2Fpeople%2Fyangwen%2FWHU-RS19.html&sa=D&sntz=1&usg=AFQjCNFzrOnViW6TWOoFbN1IaIMfyLdJhQ. Accessed 16 Feb 2018
10. SIRI-WHU Dataset. http://www.lmars.whu.edu.cn/prof_web/zhongyanfei/e-code.html. Accessed 16 Feb 2018
11. RSSCN7 Dataset. https://www.dropbox.com/s/j80iv1a0mvhonsa/RSSCN7.zip?dl=0. Accessed 16 Feb 2018
12. RSC11 Dataset. https://www.yeastgenome.org/locus/ARP7. Accessed 16 Feb 2018
13. Brazilian Coffee Dataset. http://www.patreo.dcc.ufmg.br/downloads/brazilian-coffee-dataset/. Accessed 16 Feb 2018
14. NWPU-RESISC45 Dataset. https://www.google.com/url?q=http%3A%2F%2Fwww.escience.cn%2Fpeople%2FJunweiHan%2FNWPU-RESISC45.html&sa=D&sntz=1&usg=AFQjCNGs2uMeX7KT2QvEMzcD5uF4-aQChw. Accessed 16 Feb 2018
15. Cheng, G., Han, J., Lu, X.: Remote sensing image scene classification: benchmark and state of the art. Proc. IEEE **105**(10), 1–17 (2017)
16. Thomas, M., Farid, M., Yakoub, B., Naif, A.: A fast object detector based on high-order gradients and Gaussian process regression for UAV images. Int. J. Remote Sens. **36**(10), 2713–2733 (2015)
17. Aptoula, E.: Remote sensing image retrieval with global morphological texture descriptors. IEEE Trans. Geosci. Remote Sens. **52**(5), 3023–3034 (2014)
18. Mekhalfi, M., Melgani, F., Bazi, Y., Alajlan, N.: Land-use classification with compressive sensing multifeature fusion. IEEE Geosci. Remote Sens. **12**(10), 2155–2159 (2015)
19. Cheriyadat, A.: Unsupervised feature learning for aerial scene classification. IEEE Trans. Geosci. Remote Sens. **52**(1), 439–451 (2014)
20. Jolliffe, I.: Principal component analysis. Springer, New York (2002)
21. Zhao, B., Zhong, Y., Zhang, L.: A spectral–structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery. Remote Sens. **116**, 73–85 (2016)
22. Olshausen, B., Field, D.: Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision. Res. **37**(23), 3311–3325 (1997)

23. Sheng, G., Yang, W., Xu, T., Sun, H.: High-resolution satellite scene classification using a sparse coding based multiple feature combination. Int. J. Remote Sens. **33**(8), 2395–2412 (2012)
24. Hu, F., Xia, G., Wang, Z., Huang, X., Zhang, L., Sun, H.: Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification. IEEE J. Select. Top. Appl. Earth Obs. Remote Sens. **8**(5), 2015–2030 (2015)
25. Daoyu, L., Kun, F., Yang, W., Guangluan, X., and Xian, S.: MARTA GANs: unsupervised representation learning for remote sensing image classification. National Natural Science Foundation of China (2017)
26. Hinton, G., Salakhutdinov, R.: Reducing the dimensionality of data with neural networks. Science **313**(5786), 504–507 (2006)
27. Hinton, G., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
28. Salakhutdinov, R., Hinton, G.: An efficient learning procedure for deep Boltzmann machines. Neural Comput. **24**(8), 1967–2006 (2012)
29. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A.: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Mach. Learn. Res. **11**, 3371–3408 (2010)
30. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: OverFeat: integrated recognition, localization and detection using convolutional networks. In: Proceedings of the International Conference on Learning Representations, pp. 1–16 (2014)
31. Simonyan K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proceedings of the International Conference on Learning Representations, pp. 1–13 (2015)
32. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. In: Proceedings of the Conference on Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
33. Luus, F., Salmon, B., Van Den Bergh, F., Maharaj, B.: Multiview deep learning for land-use classification. IEEE Geosci. Remote Sens. Lett. **12**(12), 2448–2452 (2015)
34. Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L.: Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. Cornell University, Ithaca (2015)
35. Zhong, Y., Fei, F., Zhang, L.: Large patch convolutional neural networks for the scene classification of high spatial resolution imagery. Appl. Remote Sens. **10**(2), 025006–025006 (2016)
36. Marmanis, D., Datcu, M., Esch, T., Stilla, U.: Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geosci. Remote Sens. Lett. **13**(1), 105–109 (2015)
37. Jingbo, C., Chengyi, W., Zhong, M., Jiansheng, C., Dongxu, H., Stephen, A.: Remote sensing scene classification based on convolutional neural networks pre-trained using attention-guided sparse filters. Remote Sens. **10**(290), 1–16 (2018)
38. Blaschke, T.: Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. **65**(1), 2–16 (2010)
39. GIS Geography. http://gisgeography.com/image-classification-techniques-remote-sensing/. Accessed Feb 16 2018
40. Tahoun, M., Nagaty, K., El-Arief, T., A-Megeed, M.: A robust content-based image retrieval system using multiple features representations. In: Proceedings of IEEE Networking, Sensing and Control, pp. 116–122 (2005)
41. Mohammed, A-M.: Multiresolution Image Segmentation. Ph.D. Thesis, Department of Computer Science, Humboldt-Universitaet zu Berlin, Germany (2008)