Urban building detection using object-based image analysis (OBIA) and machine learning (ML) algorithms

Masayu Norman¹, Hanani Mohd Shahar¹, Zuraihan Mohamad¹, Ashnita Rahim¹, Fazly Amri Mohd ¹, Helmi Zulhaidi Mohd Shafri²

¹ Faculty of Architecture, Planning & Surveying, Universiti Teknologi MARA, Perlis Branch, Arau Campus, 02600 Arau, Perlis, Malaysia
² Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia

E-mail: ayunorman@gmail.com

Abstract. The information on building features especially in the urban area is very important to support urban management and development. Nevertheless, the automated and transferable detection of building features is still challenging because of variations of the spatial and spectral characteristics to support urban building classification using remote sensing techniques. Most previous studies utilized high-resolution images to discriminate buildings from other land use in the urban area and indeed it involves a high cost to achieve that purpose. Consequently, this study utilized a medium resolution remote sensing image, Sentinel-2B with a 10-meter spatial resolution to classified the building in Selangor, Malaysia. In order to obtain a good classification accuracy, the suitable segmentation parameters (scale, shape and compactness) and features selection for building detection have been determined. Machine learning (ML) algorithms, namely Support Vector Machine (SVM) and Decision Tree (DT) classifiers have been applied to categorized five different classes which are water, forest, green area, building, and road. The result from these two classifiers was then have been compared and it is obviously showing that the SVM classifier is able to produce 20% better accuracy compared to the DT classifier, with 93% and kappa is 0.92. Thus, by enhancing the classification techniques in OBIA, building extraction accuracy using ML algorithms for medium resolution images can be improved and the expenses can be reduced as well.

1. Introduction

This study is motivated by the building classification in urban areas using high-spatial-resolution images. Previous studies show that the utilization of high-spatial-resolution images usually involved high costs, thus a medium-resolution remote sensing image has been used to discriminate urban buildings among other landuse landcover (LULC) in Klang, Selangor. Buildings in urban area play an important role because it is known as the main types of land cover and it covered nearly half of the urban area. Therefore, [1] have classified the building in Serdang residential area for harvested rainwater system applications. An automatic building detection from remote sensing data is a prerequisite for 3D (three-dimensional) building modeling, urban planning, disaster assessment, digital map updates and GIS databases [2]. According to [3], building detection can be defined for the purpose of identifying the shape of their reconfiguration or approaching each building's height information.
Object-based image analysis (OBIA) technique has been applied in several building studies, as this method able to aggregates image pixels into spectrally homogeneous image objects using an image segmentation algorithm and then classifies the individual objects [4]. So, it has become one of the finest and most common remote sensing [5]. Furthermore, the OBIA method is currently quite important with the image classification in which requires into account the shapes, texture and spectral data [6]. Nevertheless, the accuracy of classification using OBIA method is depending on their segmentation, features selection and determination of classifier [7]. [5] has stated that the image classification method's quality depends mainly on the quality of the segmentation process, which depends on a large choice of parameter values for segmentation. Moreover, the feature selection is considered a significant step in a classification system as it increases the efficiency of the classifier and reduces computational complexity by eliminating redundant data [8].

SVM and DT are an example of algorithms used for features classification. These classifiers used to classify images of low, medium or high spatial resolution [9]. The use of SVM was accomplished for change detection [10] and was rarely used to detect changes in the metropolitan area. However, DT is easy to apply and does not require the data to be standardized [11]. Therefore this study is intending to determine the significant segmentation parameters and features for building detection in an urban area using the OBIA method. Accuracy of SVM and DT classifiers then were assessed to examine the efficiency of the building classification from each classifier. Consequently, machine learning (ML) algorithms have proposed to evaluate the accuracy of building classification results with a feasible cost.

2. Study area

The study area located in Klang, Selangor, Malaysia. The area is approximate to 75.67 km$^2$ between latitudes 2°58’ and 3°01’ N and longitudes 101°24’ and 101°29’ E. Selangor is a state that surrounds the capital Kuala Lumpur on the west coast of Peninsular Malaysia. Selangor is the most advanced indicated in Malaysia with excellent infrastructures such as roads and transportation and has Malaysia's biggest population with a high standard of love and the highest poverty rate in the nation.
3.0 Data used and Methodology
The Sentinel-2B image was acquired on February 03, 2016. The dataset has 10 m spatial resolution for visible and wide NIR bands, 20 m for the red edge, narrow NIR and SWIR bands, and 60 m for atmospheric bands with a radiometric resolution of 16 bits.

| Spectral band                  | Center wavelength (nm) | Bandwidth (nm) | Spatial resolution (nm) |
|-------------------------------|------------------------|----------------|------------------------|
| B1 Coastal aerosol            | 443                    | 20             | 60                     |
| B2 Blue (B)                   | 490                    | 65             | 10                     |
| B3 Green (G)                  | 560                    | 35             | 10                     |
| B4 Red (R)                    | 665                    | 30             | 10                     |
| B5 Red-edge 1 (Re1)           | 705                    | 15             | 20                     |
| B6 Red-edge 2 (Re2)           | 740                    | 15             | 20                     |
| B7 Red-edge 3 (Re3)           | 783                    | 20             | 20                     |
| B8 Near infrared (NIR)        | 842                    | 115            | 10                     |
| B8a Near infrared narrow (NIRn)| 865                   | 20             | 20                     |
| B9 Water vapor                | 945                    | 20             | 60                     |
| B10 Shortwave infrared/Cirrus | 1380                   | 30             | 60                     |
| B11 Shortwave infrared 1 (SWIR1) | 1910             | 90             | 20                     |
| B12 Shortwave infrared 2 (SWIR2) | 2190               | 180            | 20                     |

The methodology of this study was divided into three phases as shown in Figure 2. The initial phase is the acquisition of data from open sources United States Geological Survey (USGS). The image was interpreted using band combination 4, 3, 2 with Sentinel Application Platform (SNAP) software. The image was projected to the Rectified Skew Orthomoporic (RSO) datum using ERDAS Imagine software. Then, an OBIA technique was applied which involves image segmentation, features selection, image classification, and accuracy assessment. Multi-resolution segmentation was used and the parameters such as scale, shape and compactness have been optimized to select the significant value using trial and error method. Next, suitable features are also determined and the selected features were used to classified the buildings using SVM and DT. Finally, their accuracy assessment was compared in order to obtain the most accurate classifier for building detection.

3.1 Determination of optimal segmentation
Multi-resolution segmentation (MRS) has been used and applied in this study in which the multi-resolution algorithm can reduce the average heterogeneity of the image object. El-Naggar [5] has stated that the multi-resolution algorithm known as one of the most widespread models for extracting buildings. By using MRS to segment the images of remote sensing, the parameter that has a very large impact on this process of segmentation is the scale parameter. The trial method of segmentation to identify the optimum parameter's value for urban classification has been analyzed. MRS uses three parameters to partition an image into objects: scale, shape, and compactness. The value of scale was set to 30 and 40, respectively, while shape and compactness were set to different values for 5 samples of each scale.
Table 2. The values of used segmentation parameters

| Scale | 30 | 40 |
|-------|----|----|
| Level | 1  | 2  | 3  | 4  | 5  | 1  | 2  | 3  | 4  | 5  |
| Shape | 0.2| 0.4| 0.6| 0.7| 0.8| 0.2| 0.4| 0.6| 0.7| 0.8|
| Compactness | 0.1| 0.6| 0.5| 0.7| 0.4| 0.1| 0.6| 0.5| 0.7| 0.4|

Figure 2. Research methodology flowchart
3.2 Determination of feature selection
There were 27 features that were selected within eCognition for each object, including spectral, extent, shape and texture features, to subsequently implement in the feature selection algorithms [12]. Table 3 presents the details of the selected features. The spectral features comprised the mean and standard deviation of the object spectrum, along with the maximum difference and feature brightness. The shape measures consisted of the geometrical features provided by each segmented object, such as Asymmetry, border index, compactness, density, elliptic, a radius of the largest enclosed ellipse, radius of smallest enclosing ellipse, rectangular fit, roundness, and shape index. Furthermore, the texture features of this study are based on the Haralick analysis with the grey-level co-occurrence matrix (GLCM), are dependent upon all directions.

| Feature Type                        | Feature Names                                                                 |
|-------------------------------------|-------------------------------------------------------------------------------|
| Mean                                | Mean Blue, mean green, max. diff, mean red, brightness                        |
| Standard deviation (std. dev.)      | Std. dev. blue, std. dev. green, std. dev. red                               |
| Extent                              | Area, border length, length, length/width, number of pixels, rel. border, volume, width |
| Shape                               | Asymmetry, border index, compactness, density, elliptic, a radius of the largest enclosed ellipse, radius of smallest enclosing ellipse, rectangular fit, roundness, shape index |
| Texture after haralick              | GLCM homogeneity - all directions                                             |

3.3 Image classification using SVM and DT classifiers
The classification for urban features was using machine learning (ML) algorithms. The image was classified into five classes which are building, road, forest, water, and green area. In this study, two classification methods were applied, which are support vector machine (SVM) and decision tree (DT). The classification utilized the features selection algorithms and segmentation parameter values that have been selected using the trial and error method previously. Next, the significant classification method has been determined based on their accuracy assessment result.

3.4 Accuracy assessment
The quality of a classification image depends on its accuracy. Confusion matrix used to examine the efficiency of the building classification [13]. In this study, the mean overall accuracy of five classification iterations with a fixed number of features and the same training set size was calculated for the different feature-importance-evaluation methods. The mean overall accuracy initially tended to increase rapidly with an increasing number of features used. Furthermore, slightly different classification performance was observed between both classifiers for the same training set sizes, even the use of different feature selection methods.

4.0 Result and analysis
This study evaluated the suitable value of segmentation parameter and features selection since previous studies have determined some specific importance of segmentation and feature selection for building extraction [5, 12]. The comparison of 24 segmentation samples obtained from the segmentation process due to the different results for building detection from medium-spatial resolution Sentinel-2B imagery. Moreover, three samples of determining suitable features selection for building classification was obtained. Regarding the image classification, two algorithms (SVM and DT) were used to classify the image and assess the effect of training set size and both classifiers.
4.1 Suitable segmentation parameters value for Sentinel-2B imagery

Parameters of segmentation should be set prior to image segmentation and the choice of these parameters can have a significant impact on the accuracy of classification of remote sensing images [5]. For the OBIA, multiresolution segmentation was used and performed in eCognition software. The segmentation depends on a process of district growth that places seed pixels over a whole image and gathers neighboring pixels to the local seeds, on the odds of meeting specific criteria.

The set of segmentation was performed using the Sentinel-2B image for the segmentation parameters with different values. The definition of ideal or near-ideal segmentation values for segmentation parameters scale, shape and compactness are based on the combination values shown in Table 2. Therefore, the sizes of segments are adjusted by increasing or decreasing the corresponding parameter in the process of each segmentation algorithm. Figure 3 and Figure 4 show the results of the segmentation parameters value using the trial and error method. The result shows that the 30 scale parameter is appropriate and reasonable for the five selected classes to be segmented against other undersegmented objects. The best segmentation result with the shape of parameters is therefore 0.4 and compactness is 0.6. A total of 15, 171 image objects obtained from the segmentation process.

Figure 3: Segmentation optimization with scale 30. a) shape:0.2, compactness:0.1; b) shape:0.4, compactness:0.6; c) shape:0.6, compactness:0.5; d) shape:0.7, compactness:0.7; e) shape:0.8, compactness:0.4.
According to Gorelick et al. [14], this result is applicable to direct the tile size set-up in a distributed computing environment such as the Google Earth Engine (GEE) platform. In addition, due to the image tiling that incorporates artifacts into the result, conventional image clustering procedures such as region-widening segmentation are poorly performed on distributed computing platforms [15]. However, the result of the generated segmentation does not support this assumption where segmentation is not only insensitive to the lack of information about the global image's heterogeneity, but segmentation adjustment to local conditions actually improves performance.

4.2 Determination of suitable features selection for building classification

The features algorithms involved are spectral, extent, texture and shape. These features were then have been evaluated using SVM classification. Figure 5(a) illustrates the results of building classification using mean, extent and texture. Therefore, Figure 5(b) illustrates the results of building classification using mean, std. dev., extent, shape and texture.
Figure 5. Classification using SVM with different feature selection. (a) mean, extent and texture; and (b) mean, std. dev., extent, shape and texture

Accuracy of the classified image using a different number of object features was assessed. Table 4 shows the accuracy assessment of building classification using mean, std. dev., extent, shape and texture is 13% higher with OA and kappa were 93% and 0.92, respectively. The accuracy was assessed using a minimum of 54 training samples and 15 test samples. Therefore it is obviously shown that the number of features applied will affect the classification accuracy. Furthermore, the application of shape features also contributes to good accuracy because it can easily discriminate against the building from other features.

| Feature selection                  | OA (%) | Kappa | UA (%) | PA (%) |
|-----------------------------------|--------|-------|--------|--------|
| Mean, extent, texture             | 80     | 0.75  | 100    | 67     |
| Mean, std. dev., extent, shape, texture | 93     | 0.92  | 100    | 100    |
The number of training samples for large areas is very low and not relevant since the previous study showed that using the small number of training samples (e.g. 20, 40 samples), classification algorithms on 27 features did not perform well [16]. Since the selection of features in a classification system is considered an important step as it improves efficiency, therefore this study has attempted to increase the number of training set using the same selection of features. The result shows a decrease in precision and is not significant. Nonetheless, the previous study using the test method reported that the outcome of accuracy improved and became stable by adding the training set size. Furthermore, when the training samples were representative enough, ML algorithms reach their highest accuracies. It would be said in this study, the image of medium-spatial resolution is not feasible for the application of building detection. The brightness value for building and road features, as known, is identical and hard to differentiate by computerization.

4.3 Classification using SVM and DT classifiers

The SVM classifier result that employed the features selection of mean, std. dev., extent, shape and texture that produced high classification accuracy were used. In order to assess the best classifier for building detection, SVM and DT classifier were applied and compared. Figure 6 illustrated the classification result using SVM and DT classifiers.

![Figure 6](image)

**Figure 6.** Building classification using features selection mean, std. dev., extent, shape and texture. (a) SVM classifier; and (b) DT classifier

The result of the classified image from both classifiers shows that there have misclassified between the road and building features. Consequently, it might happen when the brightness of road and building are difficult to differentiate by using eCognition software. The accuracy of the classification was assessed and compared. Table 5 shows the comparison result and defined that SVM provides 20% higher accuracy compared to DT. The overall accuracy of SVM is 93% with a kappa coefficient of 0.92.
### Table 5. Comparison of accuracy assessment between different classifiers

| Classifiers | OA (%) | Kappa | UA (%) | PA (%) |
|-------------|--------|-------|--------|--------|
| SVM         | 93     | 0.92  | 100    | 100    |
| DT          | 73     | 0.67  | 50     | 33     |

Moreover, the result of the classification of SVM has been contrasted with Google Earth. Figure 7 presents three examples of classified buildings overlay with Google Earth imagery. This indicates that the result between the classified image and Google Earth is slightly different. Nonetheless, the result of the classified image is relevant for medium-spatial resolution images.

![Figure 7. Example of building classification](image)

**5.0 Conclusion**

Building detection is challenging in the urban application of remote sensing since the classification accuracy depending on the segmentation and features selection process. Medium resolution satellite imagery, Sentinel-2B has been utilized for building detection in Klang, Selangor. Regarding the segmentation optimization, the three parameters of segmentation are scale, shape, and compactness since MRS segmentation in eCognition is only considering those parameters as the algorithm. Nevertheless, other parameters such as slope, texture and intensity could be included in future works. In this study, the small value of scale parameter provides better quality compared to the large scale for medium-spatial resolution Sentinel-2B imagery. Furthermore, the selection of object features for building classification has defined that mean, std. dev., extent, shape, and texture indicate high accuracy. SVM classifier benefits more from a features selection analysis regarding the accuracy, especially for small training set sizes. The consequence of the confusion matrix construct classification reveals that the SVM classifier is 20% higher than the DT classification with overall accuracy is 93% with a kappa coefficient of 0.92.
6. References

[1] M. Norman, H.Z.M. Shafri, M.O. Idrees, S. Mansor, & B. Yusuf, (2019). “Spatio-statistical optimization of image segmentation process for building footprint extraction using very high-resolution WorldView 3 satellite data.” Geocarto International, 6049. https://doi.org/10.1080/10106049.2019.1573853

[2] I. Tomljenovic, B. Höfle, D. Tiede, & T. Blaschke, (2015). “Building Extraction from Airborne Laser Scanning Data: An Analysis of the State of the Art”. 3826–3862. https://doi.org/10.3390/rs70403826

[3] T. Sritarapipat, & W. Takeuchi, (2017). “Building classification in Yangon City, Myanmar using Stereo GeoEye images , Landsat image and night-time light data.” Remote Sensing Applications: Society and Environment, 6(January), 46–51. https://doi.org/10.1016/j.rsea.2017.04.001

[4] D. Liu, & F. Xia, (2010). “Assessing object-based classification: advantages and limitations”. Remote Sensing Letters, 1(4), 187–194. https://doi.org/10.1080/143161003743173

[5] A.M. El-naggar, (2018). “Determination of optimum segmentation parameter values for extracting building from remote sensing images.” Alexandria Engineering Journal, 57(4), 3089–3097. https://doi.org/10.1016/j.aej.2018.10.001

[6] H.I. Sibaruddin, H.Z.M. Shafri, M., B. Pradhan, & N.A. Haron, (2018). “UAV-based Approach to Extract Topographic and As-built Information by Utilising the OBIA Technique,” 6(3), 103–123. https://doi.org/10.12691/jgg-6-3-2

[7] T. Blaschke, (2010). "Object based image analysis for remote sensing." ISPRS Journal of Photogrammetry and Remote Sensing, 65(1), 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

[8] M. Pedergnana, P. Marpu, M. Dalla Mura, J. Benediktsson, & L. Bruzzone, (2013). "A Novel Technique for Optimal Feature Selection in Attribute Profiles Based on Genetic Algorithms." Geoscience and Remote Sensing, IEEE Transactions On, 51, 3514–3528. https://doi.org/10.1109/TGRS.2012.2224874

[9] P.A. Tavares, N. Ely, S. Beltr, U. Silva, (2019). "Integration of Sentinel-1 and Sentinel-2 for Classification and LULC Mapping in the Urban Area of Belém, Eastern Brazilian Amazon." https://doi.org/10.3390/s19051140

[10] P. Grif, P. Hostert, O. Gruebner, & S. Linden, (2010). "Remote Sensing of Environment Mapping megacity growth with multi-sensor data". 114, 426–439. https://doi.org/10.1016/j.rse.2009.09.012

[11] E. Szczerbiicki, (2001). "Management of Complexity and Information Flow. In Agile Manufacturing: The 21st Century Competitive Strategy."

[12] L. Ma, T. Fu, T. Blaschke, M. Li, D. Tiede, & Z. Zhou, (2017). "Evaluation of Feature Selection Methods for Object-Based Land Cover Mapping of Unmanned Aerial Vehicle Imagery Using Random Forest and Support Vector Machine Classifiers." International Journal of Geo-Information. https://doi.org/10.3390/ijgi6020051

[13] B. Pradhan, M.N. Jebur, H.Z.M. Shafri, & M. S. Tehrany, (2015). "Data Fusion Technique Using Wavelet Transform and Taguchi Methods for Automatic Landslide Detection From Airborne Laser Scanning Data and QuickBird Satellite Imagery." IEEE T. Geosci. Remote Sens., 1–13.

[14] N. Gorelick, M. Hancher, M. S. Ilyushchenko, D. Thau, & R. Moore, (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone." Remote Sensing of Environment, (2016). https://doi.org/10.1016/j.rse.2017.06.031

[15] P. Lassalle, J. Ingliad, J. Michel, M. Grizonnet, & J. Malik, (2015). "A Scalable Tile-Based Framework for Region-Merging Segmentation." IEEE Transactions on Geoscience and Remote Sensing, (October). https://doi.org/10.1109/TGRS.2015.2422848

[16] L. Congcong, W. Jie, and W. Lei, (2014). "Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery." Remote Sensing, 6, 964–983. https://doi.org/10.3390/rs6020964