ABSTRACT

Classification of building into different typology is one of the significant methods to assess the risk and vulnerability of an area. Manual survey of buildings or field visits to classify buildings into different typology is a time and manpower consuming method involving high cost. With the advancement of different techniques and methods now it is possible to know about the building typology in a much quicker way and at a much lower cost. This paper makes an attempt to extract the information of buildings from satellite images to use them for hazard assessment. Meghalaya, a regional state with hilly terrains in the North Eastern region of India has been selected for the study purpose in this paper. For an accurate and reliable building extraction operation, an ideal framework will be proposed to detect the building from the high resolution satellite/aerial images. The proposed framework along with the domain knowledge of spatial and spectral features will provide various characteristics such as the nature of the building, information on the type of building, location of building and surrounding area. The nature of the building might include edges, planes or lines of the building, its roof types, etc. The methodology will include edge detection, line generation and clustering of the building whichever is most suitable for the study purpose. The extracted built-up area of the buildings from the satellite images will be used for seismic hazard assessment of the area. Buildings extracted will be classified into different types based on the architectural layout, roof types, usage, occupancy, etc.

Keywords: Typology; building extraction; satellite image; clustering; roof type; hazard assessment.

1. INTRODUCTION

Satellite image classification can be referred to as extracting information from satellite images. Satellite images are rich in geographical information and play a major role in acquiring information about objects on the surface of the earth. Satellite images also play an important role in ground features extraction. In comparison to field study, it is easier to extract quantitative as well as qualitative data from satellite images at a much lower cost and time.

Image classification is an important technique through which we can extract information from large number of satellite images. It is a process of grouping similar pixels into meaningful classes (Sunitha et al. 2015) to give more detailed information for that class. Feature extraction from satellite images, however, is a great challenge because of the high density of buildings and complexity of type of housing especially in urban areas. For example, In an area like Shillong, the Capital of Meghalaya, where there is mix housing typology, carrying out building extraction process and clustering the buildings into various types is a very difficult task.

Since from late 1980s, the identification of man-made structure and buildings has been considered as an active field of interest. There are many important uses of building detection from satellite imagery. Some of them are urban mapping, urban planning, urban change detection analysis, target detection, geographic information system (GIS) (Shorter & Kasparis, 2009). However, Image extraction methods are difficult operations to perform, because of the presence of natural texture of vegetation, area

---

1Research Scholar, Earthquake Engineering Department, IIT Roorkee, India, devilata.pegu15@gmail.com
2Associate Professor, Earthquake Engineering Department, IIT Roorkee, India, josodequiitr@gmail.com
3Professor, Earthquake Engineering Department, IIT Roorkee, India, sarmamukat@gmail.com
occupied by water and other different kinds of elements which are generally present on or near the object of interest in the aerial or satellite image. The similarity in appearance of buildings and other surrounding features is also an issue to be considered while extracting the building from the satellite image. The main objective of this study is to extract the built-up area of the buildings from the selected study area and then used them for seismic hazard assessment. The methodology to be adopted in this study is to identify and isolate the built up area by image extraction methods.

2. BACKGROUND

The extraction of buildings from satellite imagery is a challenging task. However, recent improvements in the spatial and spectral resolution of satellite imagery has driven researchers to develop different algorithms (i.e. automatic and semi-automatic) for the extraction of buildings from very high resolution satellite imagery. Many features that appear in the satellite imagery make the process more complex for image extraction of buildings, even though the new sensors provide satellite imagery with improved resolutions. Factors such as scene complexity, building variability and sensor resolution (Mayer, 1999) affect the overall accuracy for the detection of buildings. The man-made feature (i.e. building) is one of the most significant features among the other features, for the reason of their variability, complexity and abundance in urban areas (Chaudhuri et al. 2016). Generally, very high resolution satellite imagery is necessary to extract detailed spatial and spectral information of buildings. In order to extract building features from satellite imagery, numerous algorithms were proposed by various researchers, which are as follows. Attarzadeh and Momeni (2012) proposed object based algorithm, in which stable and variable features were utilized jointly, obtained from inherent qualities and threshold analysis. The visual analysis indicates that the algorithm can detect major rectangular buildings of Quickbird imagery. Wang et al. 2013 detected the rectangular buildings using mean shift segmentation, scale invariant feature transform (corner detection) and adaptive windowed Hough transform. Wang et al. 2013 adopted bilateral filter, line segment detector and perceptual grouping approach. The entire algorithm mentioned above detects only the rectangular buildings from the RGB image. Ghaffarian and Ghaffarian (2014) used double threshold method, parallelepiped supervised classification and morphological operators for building detection from Google earth image. Tournaire et al. (2010) used point processes on digital elevation models. Brunn and Weinder (2013) separated buildings and vegetation areas using height data and geometric information on Digital Surface Models (DSM) data. Abdullah et al. (1997) proposed two novel methods to detect buildings by combining panchromatic and DSM data. The major drawback of DSM based scheme is that a group of trees may look like a building and there is no easy way to separate them. Studies have been done where buildings are extracted using google earth image, panchromatic image, and multispectral image with the combination of R, G, and B color mode. Many algorithms have been developed for detecting buildings with the same shape (i.e. rectangular, square), colour and size but only a few algorithms are available for detecting buildings with arbitrary shapes and sizes.

The proposed methodology in this paper detects the buildings without any influence from their geometric characteristics and it also provides the training data sample automatically to the supervised classification. However, the method classifies the non-building features as building features when they have equivalent spectral values. A few limitations were observed during the proposed image classification process. Some non building objects are shown as buildings. Further, buildings which are close to each other were classified as a single building. The roof types of many buildings in the study area matches with the vegetation which make the process difficult to differentiate between them. But on adjusting the spectral values in the image classification process, better results are obtained for these errors.

Two approaches were adopted to extract the buildings: (1) automatic method, (2) manual method. After processing the image by the two methods, their results were compared qualitatively. In the automatic approach; firstly, the vegetation portion was removed from the input image. Secondly, adaptive K-means clustering algorithm was adopted to cluster the different pixels into different classes. And then morphological operator fill and open was implemented to extract the buildings. In the manual approach, area of interest (AOI) was created from the input image. Later, the generated
AOI was used to subset the interested features (i.e. buildings, vegetations, open areas) from the image. The next important objective was to find the effectiveness of high spatial resolution google earth image, for the extraction of buildings using automatic and manual methods. The results of both the methods were compared qualitatively and discussed.

3. DATA USED AND METHODOLOGY

High resolution Google earth images of Meghalaya were downloaded for the study purpose. Taking the various pattern of building settlement of major towns in the state and their vulnerability to seismic risk an attempt is made to extract the buildings from the satellite images.

Both manual and automated methods are used for the extraction of buildings from the google earth images. Automated image classification is further classified into two categories viz, Supervised Classification Methods and Unsupervised Classification Methods. In Supervised Classification Methods an input is required from an analyst which is known as training set. Training sample is the most important factor in the supervised satellite image classification methods. Supervised classification includes additional functionality such as analyzing input data, creating training samples and signature files, and determining the quality of the training samples and signature files. Accuracy of the methods depends on the samples taken for training and its quality. Unsupervised classification method involves in clustering mechanisms to group satellite image pixels into unlabelled classes/clusters. By comparing with the original image labels are assigned to the clusters. Most common unsupervised satellite image classification is ISODATA and K-Means.

Manual satellite image classification avoids loss of useful information of the buildings from the satellite image. This approach is time consuming but effective. In this approach, an AOI is created from the input image with the building features. Later, the generated AOI is used to subset the interested features (i.e. buildings) from the image. Accuracy depends on the knowledge of the analyst and the familiarity towards the area of interest.

For the purpose of this study the area selected is divided into grids of 1 km x 1 km. Then a point is georeferenced on the selected grid. The selected grid is the input image for the image classification process. By both supervised and unsupervised method the selected grid is processed by combination of different clustering method. The approach presented in this paper is designed to effectively extract the building features from the Google earth image so that only the built-up area can be separated from the other features. The approach to create building typology from the extracted building features will be
used to assess the seismic hazard and vulnerability of the built infra structure for seismic risk.

4. COMPARISON OF METHODS

The input image is a 1 km x 1km grid from an urban area of Meghalaya. Colour codes are set for different objects to be detected from the input Google earth image and from this building features are to be extracted. High resolution google earth image with high pixel values are used for this study. In both supervised and unsupervised classification method some errors were observed in the output images. By comparing the output image with the input image, it is observed that some of the vegetation feature and buildings are shown in the same colour code. Again it is also observed that some portion of vegetations have the same colour code of buildings. In order to overcome these, the threshold values of the spectra are adjusted in such a way to remove the error in the output image. These values are manually adjusted by trials.

*Adaptive K-means clustering.* The Adaptive K-means clustering algorithm functions by automatically choosing the appropriate K elements from the input image (Bhatia, 2004). The algorithm automatically determines the K elements and generates the group of clusters (i.e. the features with the same intensity value are grouped together). Generally, the algorithm classifies each pixel into the clusters, based on their intensity values. Firstly, the algorithm computes the distance between the selected element and the number of clusters. This process also helps to determine the distance between the two elements. In order to compute the distance, it is important to normalize the distance properties, so that the domination of distance from one property (or) certain properties is not omitted from the computation of distance.

Since the study area is located in densely inhabited area with different buildings types, other methods used by various researchers is not feasible in the current area. The current methodology is comparatively one of the best and easiest methods to extract buildings features with arbitrary shape and size from Google earth images.
Figure 2. A grid of 1km x 1km selected from an urban area of the state Meghalaya with georeferenced point Lat 25.3548N and Long 91.5296E with coordinates 25.601302N, 91.884649E
In Figure 3 and Figure 4 buildings are shown in red colour while forest and open areas are shown in green and blue colour. Here some forest and open areas which have the same spectral values are also shown as building features.

After adjusting the spectral values of all the features in the satellite image only building features are extracted from the google earth image. The extracted building features are shown in Figure 5.
Figure 6. Supervised classification (Minimum distance) showing only the built-up features. The building features can be identified easily from the extracted image as shown in the figure above.

5. RESULTS AND DISCUSSION

Satellite imagery contains important information and it is also source of knowledge for many different applications. Hence, extracting the main characteristics from these images is a challenging task. Therefore, there is a need of addressing different issues before performing operations like isolating the characteristics and identifying them. Building being an important object in a satellite image, building detections in has a large number of applications in civilian, commercial and military fields. Building detection can also be handy and useful in cases of natural or manmade disasters, where operations like finding out the existence of a building can be performed.

In order to extract the buildings without any loss of information, both automated and manual extraction methods are adopted using ERDAS Imagine 2014 software. Only building features are extracted from the input Google earth image. Since the area selected is in varied built-up location, clustering process and classification of buildings into different typology was a difficult process. Hence the assessment of total number of building in the output image had to be done by approximation. This was done by removing the vegetation and other features from the output image and retaining only the manmade features from the output image. For verifying the final results of building extraction and classification of buildings, ground survey of few buildings has been carried out to validate the results. Here, the manual method extracts all the buildings without much loss of information and the google earth image helps to extract the building information very effectively. Nevertheless, manual extraction of building is time consuming and also depends on the experience of the user to digitize the boundary of buildings for effective extraction of buildings.

By comparing the results of both automatic and manual methods, it is understood that the automatic algorithm works efficiently only when the features of interests in the image are recognized to be in the same pattern and size. Generally, the loss of information in the output image is common regardless of input image (i.e. features with the same size or different size, same shape or different shape and same pattern or different pattern). Here, the loss of information in the automated algorithm is noted for various reasons such as different building size, shape, pattern and color. In the case of manual method, the loss of information for extracting buildings is less, but the method requires more time to complete the process.
6. CONCLUSIONS

In this paper, building extraction processes of Meghalaya are presented. The methods used are both automatic and manual method. Both the methods are adopted for the 1km x 1km grid taking Google earth image as the input image. Buildings are extracted by taking different combination of clustering methods. The extracted buildings are compared qualitatively and quantitatively. The best extracted building features are further analyzed to obtain the typology of the extracted buildings. Since the study area is located in a hilly region, high-rise buildings are very few. Mostly Type-A (Mud/Random Rubble Masonry, Adobe), Type-B (Brick Masonry Buildings), and Type-C (Reinforce Cement Concrete Buildings) types of buildings are found in the region. For validating the results one of the extracted buildings are verified by manual ground survey of that building. However, in both the methods it was observed that there is loss of information in the output images. The results of manual method indicate that the extraction of buildings is achieved with minimum loss of information in comparison with the automated method. The results from both the automated and manual methods indicate that the spatial and spectral information of buildings are clearly identifiable. Therefore, improving the spatial resolution of the original image increases the spatial and spectral information of buildings. In the case of input image, (i.e. if the interested features are identified to be different from each other in terms of shape, size and color) the manual method is recommended, in order to reduce loss of information. It is to be noted that the performance of automatic algorithm is very effective when all buildings are in rectangular shape. In this case, the area selected is densely built-up area and the building types are highly varied. Hence the clustering of building was a difficult task. However, area with regular shaped buildings like group housing complexes could be classified into one cluster and same typology.

7. ACKNOWLEDGEMENT

The author would like to extend her sincere thanks to the Department of Earthquake Engineering, IIT Roorkee and Ministry of Human Resource, Government of India for providing assistantship during the period of carrying out this work.

8. REFERENCES


Samadzadegan, F. and Rastiveisa, H. A Clustering-based technique for automatic damage map generation using high resolution satellite imagery, Department of Surveying and Geomatics Engineering, University College of Engineering, University of Tehran, Tehran, Iran.


