EARTHQUAKE DISATER AREA EXTRACTION BY MACHINE LEARNING

Jonggeol Park[1] and Youngjoo Kwak[2]

[1]Tokyo University of Information Sciences, 4-1, Onaridai, Wakaba-ku, Chiba 25-8501, Japan

Email: jgparktuis@yahoo.co.jp

[2]International Centre for water Hazard and Risk Management under the auspices of UNESCO, Japan

Email: kwak@icharm.org

KEY WORDS: Earthquake, Machine learning, Object-based analysis, Disaster area, GeoEye-1

ABSTRACT: The final goal of this study is to create data to support emergency efforts in a disaster affected area by locating damaged buildings shortly after the disaster. In this study, prioritizing the practicality of the method for emergency purposes, we designed a method only to use a single satellite image of an affected area, eliminating the use of complex algorithms and auxiliary data. The uniqueness of our method lies in the application of an object-based region segmentation to images and the use of features of objects obtained from texture, hierarchical and other information in order to extract damaged buildings. Out of 26 features resulting from the analysis of objects, we found one feature and three combinations of two different features that are effective in extracting damaged buildings, such as Rectangular fit, Homogeneity, Number of sub-objects/Area, and Length of longest of edge/Area.

1. INTRODUCTION

When a disaster occurs, it is very important to have an overall picture of the devastation as soon as possible. Locally organized voluntary groups for disaster prevention and other groups of volunteers can play a key role in this effort as invaluable information sources. However, in a large-scale disaster, such groups are also likely to suffer damage. Emergency relief and rescue teams will be provided from outside the affected area, particularly if the disaster results in a widespread destruction. Unfortunately, it has been pointed out that firefighters and volunteers from other areas often do not know enough about how to get around the affected area and thus often have difficulty for performing efficiently and effectively their full capacity. In this sense, maps and images that show the current status of the affected area are necessary to maximize rescue and relief operations.

A number of studies have also been conducted on another automatic method for identifying damaged buildings in a disaster affected area, using a single image of the area after a disaster. This method is designed to identify damaged buildings using thresholds such as color information (hue, saturation, brightness) and edge information (intensity, intensity dispersion, frequency values of direction). Among the six thresholds, hue has been found most effective to identify damage. However, a careful adjustment in thresholds is needed because color information varies in different natural settings.

A Pixel-Based Image Analysis (PBIA) approach has also been studied by many experts. This approach has been found to have difficulty in recognizing an area as a single object when the area consists of multiple types of land cover. An Object-Based Image Analysis (OBIA) approach can be a solution to this problem. Some studies have reported that this approach, dividing an image into small segments called image objects, is capable of discerning objects as accurately as visual interpretation by experts. Because objects such as buildings in an image are made of multiple pixels, a PBIA approach sometimes fails to identify a building as a single object. Particularly in urban areas, where many objects of different types exist, an OBIA approach is considered to be more effective than the PBIA approach. In addition, not only spectral information but also object shapes and texture information can be taken into account when an OBIA approach is employed. In this respect, an OBIA approach has more possibility for practical applications, with more options of effective features to help determine affected areas with higher accuracy. Yamazaki et al. (2005) extracted buildings damaged by an earthquake that occurred in Algeria in 2003 using QuickBird images. By visual interpretation, they first outlined the buildings in the images before and after the earthquake and then identified damaged buildings. Janalipour and Mohammadzaheh (2017) proposed a semi-automated fuzzy decision making system using Genetic Algorithm to detect a damaged building.

The purpose of this study was to create data to support near real-time relief and rescue efforts in a disaster affected area by locating damage in a short period of time using only post-disaster satellite images of the area. For this purpose, we used only high-resolution satellite images to ensure the practicality of the proposed method, eliminating the use of complex algorithms and auxiliary data. A segmentation technique was applied to high-resolution satellite images, and the features of buildings damaged by an earthquake (e.g., information gained after segmentation, shape, texture) were collected from training objects for analysis. The Stepwise method was applied to the collected features to select effective features to extract damaged buildings. The selected features were then used to develop an equation for identifying damaged buildings. The equation was finally applied to different types of objects to test its discerning power.

2. HIGH RESOLUTION SATELLITE DATA

This study used GeoEye-1 data, a type of high resolution satellite data, collected before the earthquake (July 27, 2009) and after the earthquake (January 13, 2010). The pre-quake data were used to verify the analysis results. GeoEye-1, a commercial satellite launched on September 6, 2008, is mounted with panchromatic and multispectral sensors with four bands covering the wavelength range from blue to near infrared. The satellite is capable of collecting images with a ground resolution of 0.41 m in the panchromatic mode and 1.65 m in the multispectral mode, but actually provides images with 0.5 m panchromatic resolution and 2 m multispectral resolutions. The nominal swath width is 15 km, and the data is encoded in 2-byte, or 11-bit, format. In this study, we used pan-sharpened color images at a 0.5 m resolution that were created by merging panchromatic monochrome images with multispectral images.

3. METHOD

3.1 Two-step hierarchical region segmentation

The image object levels of an image object hierarchy range from fine resolution of the lowest image object level to coarse resolution of the highest image object level. Image objects have multiple sub-objects (figure 1). The area of an image object is the number of pixels forming it, that is the number of sub-objects that are located on the next lower image object level in the image object hierarchy. Segmentation was conducted twice on training images because debris of walls and roofs scatter around damaged buildings in disaster affected areas. The first segmentation was to extract each small piece of debris as a sub-object, while the second segmentation was to extract a building with pieces of debris as an object. This two-step segmentation helps determine the number of sub-objects from the first segmentation that is included in an object from the second segmentation. Because each piece of debris around a building is recognized as a sub-object in this segmentation technique, the number of sub-objects increases as the number of pieces of debris increases. The feature values of Number of sub-objects and Area were obtained from the hierarchical region segmentation. Figure 1 shows image object 1, which has three sub-objects and an area of eight pixels; and image object 2, which has two sub-objects and an area of three pixels.

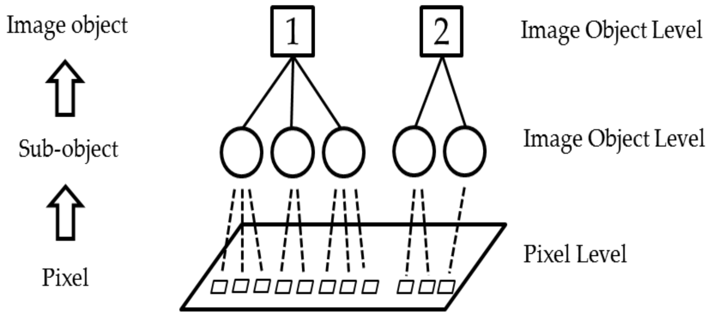


Figure 1. Two-step hierarchical region segmentation

3.2 Support Vector Machines (SVMs)

The size of an image becomes larger as spatial resolution becomes higher and observation covers a wider area. A larger size of image sometimes requires dozens of hours to process if we plan to calculate all features for each object. To identify damaged buildings as soon as possible, significant features should be selectively processed for higher efficiency. In this study, we used the Support Vector Machines method to determine damaged buildings. SVMs are based on statistical learning theory and have the aim of determining the location of decision boundaries that produce the optimal separation of classes. In a two-class pattern recognition problem where classes are linearly separable, the SVMs select the one linear decision boundary that leaves the greatest margin between the two classes. SVMs were initially designed for binary (two-class) problems. When dealing with multiple classes, an appropriate multi-class method is needed. Techniques such as 'one against one' and the 'one against the rest' are in frequent use for the multi-class problems

3.3 Random forest classifier

The random forest classifier consists of a combination of tree classifiers where each classifier is generated using a random vector sampled independently from the input vector, and each tree casts a unit vote for the most popular class to classify an input vector. There are many approaches to the selection of attributes used for decision tree induction and most approaches assign a quality measure directly to the attribute. The number of features used at each node to generate a tree and the number of trees to be grown are two userdefined parameters required to generate a random forest classifier.

4. RESULT

4.1 Selection of significant feature by the stepwise method

The damaged buildings identified by visual interpretation were analyzed for the 26 features of the three different feature groups. In this section, we explain the process of selecting significant features out of the 26 features using the Stepwise method for efficiently extracting damaged buildings.

In most OBIA the objects are constructed using a single segmentation. A characteristic of this research is that areas covered with fragments of debris from collapsed walls and roofs were treated as image-objects; but the individual debris fragments were treated as sub-objects. The number of sub-objects was low in areas characterized by roadways and intact roofs, but high in areas where roofs and walls had collapsed. The Stepwise method was used three times to investigate the effectiveness of the two-step hierarchical region segmentation and the feature values created to highlight the characteristics of damaged buildings. Three groups were prepared for this analysis: Group I consists of only Haralick’s eight texture features; Group II comprises features from three feature types of Haralick’s, the two hierarchical region classification, and the figure classification; and Group III is made of the four feature types and the additional features highlighting the characteristics of damaged buildings (i.e., Dissimilarity/Homogeneity, Contrast/Homogeneity, Number of Sub-Objects/Area, Length of Longest of Edge/Area, Length/Width). We analyzed these three groups for accuracy rates of damaged building identification to find out significantly effective features.

Table 1 shows the comparison of accuracy assessment result of three groups using the stepwise method for the identification of damaged buildings and non-damaged items. All eight texture features of Group I were selected as effective. Twenty-four out of 29 damaged buildings (82.8%) and 41 out of 60 non-damaged items (68.3%) were identified correctly. The overall accuracy rate was 73%.

The analysis with Group II found Homogeneity, Standard deviation length of edges, Number of edges, and Number of sub-objects as effective discriminating features. Twenty-six out of 29 damaged buildings (89.7%) and 48 out of 60 non-damaged items (80.0%) were identified correctly. The overall accuracy rate was 83.1%, which was a 10% improvement compared with the accuracy rate of Group I. The analysis with Group III found Rectangular_fit, Homogeneity, Number of sub-objects/Area, and Length of longest of edge/Area as significantly effective features. Twenty-six out of 29 damaged buildings (89.7%) and 51 out of 60 non-damaged items (85%) were identified accurately. The overall accuracy rate was 86.5%. The results indicated that the accuracy rate improved by 3.4% by the introduction of the features that underline the qualities of damaged buildings.

Table 1. Comparison of accuracy assessment result of three Groups on damaged buildings and non-damaged buildings.

		Group I				Group II				Group III			
		ID	Expected group 1	Expected group 2	sum	ID	Expected group 1	Expected group 2	sum	ID	Expected group 1	Expected group 2	sum
Original Data	frequency	1	24	5	29	1	26	3	29	1	26	3	29
		2	19	41	60	2	12	48	60	2	9	51	60
	%	1	82.8	17.2	100	1	89.7	10.3	100	1	89.7	10.3	100
		2	31.7	68.3	100	2	20.0	80.0	100	2	15.0	85.0	100
Overall accuracy:					73.0%				83.1%				86.5%

ID 1: Damaged Buildings

ID 2: Non-damaged items (buildings, tree, roads)

Table 2. Coefficients table

	Unstandardized Coefficients		Standardized Coefficients	t	P
	B	Std.Error	Beta		
(constant)	0.386	0.504		0.767	0.445
Number of sub-object / Area	-19.980	3.459	-0.427	-5.777	0.000
Homogeneity	14.088	3.326	0.317	4.236	0.000
Length of longest Edge / Area	6.241	1.923	0.236	3.246	0.000
Rectangular fit	1.112	0.578	0.143	1.924	0.057

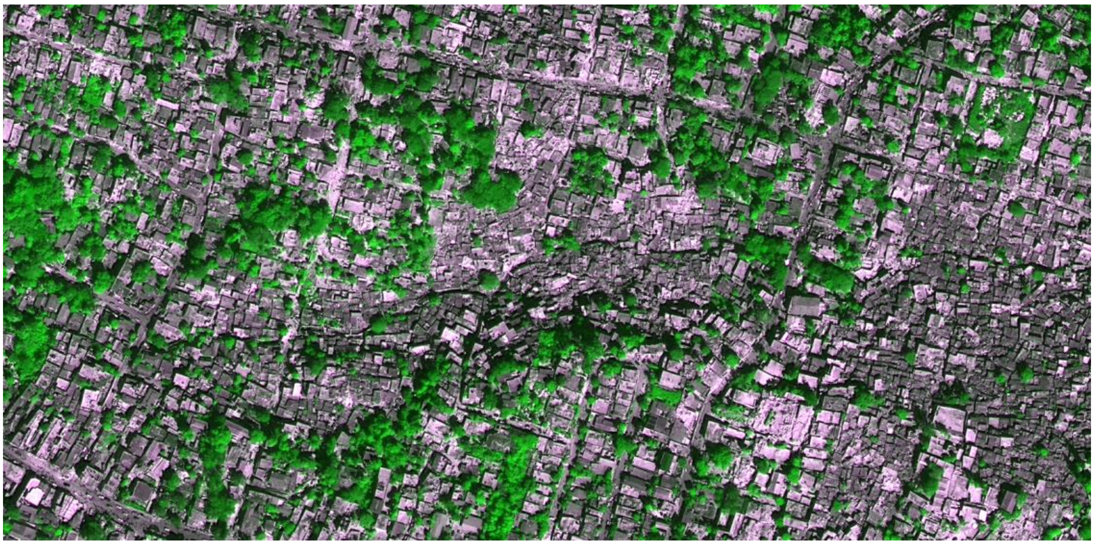
Table 2 shows the most useful variables extracted from many explanatory variables that could explain the differences in the objective variable (damaged or not-damaged) as generated by the Stepwise Method, which is a regression model. The values enclosed in dotted lines are those indices obtained when explaining the objective variable utilizing the extracted explanatory variables. Standardized Coefficients express the degree of influence on the objective variable for the various standardized explanatory variables (with influence of measurement units removed). The higher the absolute value the greater the influence of the variable.

The t-value and P-value were employed to evaluate the utility of the partial regression coefficient. The higher the absolute t-value the greater the influence of the variable, and the lower the P-value the greater the usefulness of the variable for explaining the objective variable. Both the t-value and P-value indicate that the number of sub-objects/area is the best explanatory variable for distinguishing damaged areas from non-damaged areas. These results suggested that to identify damaged buildings with high accuracy, the combined use of five features (Length of longest of edge, Area, Homogeneity, Length of the longest of edge and Number of sub-objects) set off the characteristics of damaged buildings. A regression model to predict target variables can be established based on the results of multiple regression analysis. Equation (1) was devised using the regression coefficients of the four selected explanatory variables. The equation was defined as Earthquake_damage (EQDG), and a building was considered as damaged if its EQDG was calculated to be 1.5 or lower. Unstandardized coefficients indicate how much the dependent variable varies with an independent variable when all other independent variables are hold constant.

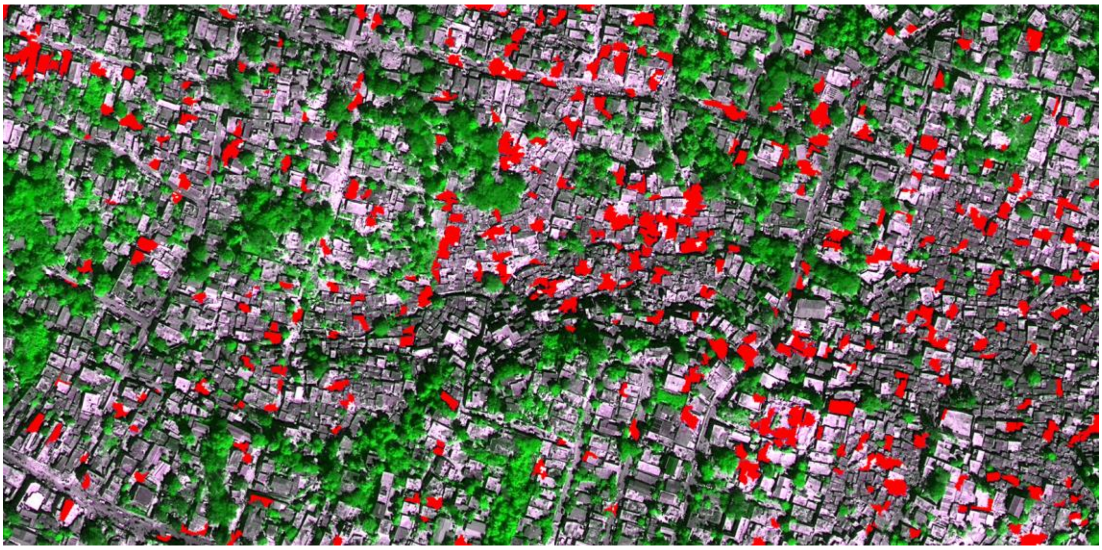
$$\begin{aligned}\text{Earthquake\_damage (EQDG)} = & -19.980\times(\text{Number of sub-objects/Area}) + 14.088\times\text{Homogeneity} \\ & + 6.241\times(\text{Length of the longest of edge/Area}) + 1.112\times\text{Rectangular\_fit} + 0.386\end{aligned} \quad (1)$$

4.2 Result of identification of damaged buildings using EQDG

We identified damaged buildings using equation (1) and the features obtained from the two-step hierarchical region segmentation. Figure 2 shows the images of three different affected areas before and after the earthquake including the images with objects having an EQDG of 1.4 or lower.



a)



b)

c)

Figure 2. Overall image of the disaster area

a) overall image b) disaster area detection (red area) of figure 11a by equation (1) c) zoom image of road with numerous people and parking lot d) miss identified as a damaged building of figure 2c

Figure 2a is an overall image of the disaster area. The disaster area was near the coast in the city of Port-au-Prince, the capital of Haiti. Spaces between buildings were narrow, and the old town district consisted of many small buildings. Many people had collected on the streets following the quake. Figure 2b shows an area determined by Equation (1) to be a damaged area （EQDG > 1.5）. This indicates that the analysis correctly identified an area where concrete buildings had collapsed. Figure 2(c and d), on the other hand, shows an example where the analysis incorrectly identified a damaged area. Numerous people on the road registered as a large number of sub-objects that the system misinterpreted as debris.

5. FUTURE WORKS

In this study, we visually interpreted images before and after the earthquake for validation purposes. However, it is sometimes difficult to confirm damage through visual interpretation even with high-resolution images particularly when small buildings exist next to each other in a small area. For more accurate validation of the proposed method, ground investigation is necessary. In case of an earthquake, damage may totally vary at different locations, depending on the magnitude of the earthquake, the spatial characteristics of the locations (e.g., urban or mountainous, structures made of brick, concrete or wood), and a secondary disaster that may occur after the earthquake (fire, tsunamis, etc.). More investigation is needed to collect additional data and information on various locations for wider application of the proposed method.

Spatial resolution is another issue. If images with different spatial resolutions (e.g., aerial photos and satellite data) are used to process buildings of the same size, different parameters are necessary to perform the region segmentation for preparing objects. More research should be conducted to develop a method for quantitatively setting segmentation parameters according to the resolution of an image and the size of a building, instead of the current practice in which such parameters are set based on empirical knowledge.

Acknowledgments: This research was supported by JSPS KAKENHI Grant Numbers JP25340014 and JP26350403.

References

Hasegawa, H., H., Aoki, F., Yamazaki, M., Matsuoka, and I., Sekimoto, 2000, Automated Detection of Damaged Buildings Using Aerial HDTV Images, Proceedings of the IEEE 2000 International Geoscience and Remote Sensing Symposium, IEEE, CD-ROM, 2000.

Mitomi H, M. Matuoka and F. Yamazaki, 2002, Attempt for automated detection of building damage areas due to recent earthquakes using aerial television images, Japan society of civil engineers, pp. 267-278 (in Japanese).

Kamagata N., and K. Hara, 2010, Vegetation mapping by object-based image analysis: evaluation of classification accuracy and boundary extraction in a mountainous region of central Honshu, Japan, vegetation science, pp. 83-94(in Japanese).

Gamba, P. and F. Casciati, 1998, GIS and image understanding for near-real-time earthquake damage assessment. Photogramm. Eng. Remote Sens Vol. 64, pp. 987-994.

Hay, G.J., G. Castilla, M.A. Wulder and J.R. Ruiz, 2005, An automated object-based approach for the multiscale image segmentation of forest scenes. International Journal of Applied Earth Observation and Geoinformation, Vol.7, pp.339-359.

Xie, Z., C. Roberts and B. Johnson, 2008. Object-based target search using remotely sensed data: A case study in detecting invasive exotic Australian Pine in south Florida. ISPRS Journal of Photogrammetry & Remote Sensing, Vol.63, pp. 647-660.

Yamazaki F, M., Y. Yano and M, Matsuoka, 2005, Visual damage interpretation of buildings in Bam city Using QuickBird images following the 2003 Bam, Iran, Earthquake, Earthquake Spectra, Vol.21, pp. pp.329-33

Janalipour M., and A. Mohammadzadeh, 2017, A Fuzzy-GA Based Decision Making System for Detecting Damaged Buildings from High-Spatial Resolution Optical Images, Remote Sens. Vol.9, 349; doi:10.3390/rs9040349

Yu, Q., P. Gong, N. Chinton, G. Biging, M. Kelly and D. Schirokauer, 2006, Objectbased detailed vegetation classification with airborne high spatial resolution remote sensing imagery. Photogrammetric Eng. Remote Sensing Vol. 72, pp.799-811.

Kouchi, K., and F., Yamazaki, 2005, Damage detection based on Object-based segmentation and classification from high-resolution satellite images for the 2003 Boumerdes, Algeria earthquake, Proc. 26th Asian Conference on Remote Sensing, CD-ROM, 6p