HYPOTHESIS TESTING FOR AUTOMATIC COLLAPSE DETECTION OF BUILDINGS USING DSMs

Mehdi Rezaeian*[1] and Armin Gruen[2]
[1] Assistant Professor, Electrical and Computer Engineering Department, Yazd University
Daneshgah Blvd., Safa-ieh, Yazd, Iran; Tel: +98-351-8122610
E-mail: mrezaeian@yazduni.ac.ir

[2] Professor, Institute of Conservation and Building Research, ETH Zurich
Wolfgang-Pauli-Strasse 27, CH-8093 Zuerich Switzerland; Tel: +41-44-6333038
E-mail: armin.gruen@geod.baug.ethz.ch

KEY WORDS: Collapse detection, DSM, Hypothesis test, Earthquake

ABSTRACT: Comparing geometrical data in the form of digital surface models (DSMs) can be an interesting approach for detecting and understanding the extent of demolished areas of buildings. The building changes can be automatically detected by simply subtracting one of the DSM data sets from another. DSMs can be derived from variety of sources requiring different processing methods. In urban areas, using stereoscopic (multiple) aerial photographs and LIDAR technology are prevalent. These remotely sensed DSM production methods provide users with high resolution DSM data that have well-defined vertical and horizontal accuracies. As such, any resultant DSM is subject to both the precision and accuracy of the measurement sensor as well as the quality of image matching and interpolation methods, and thus is subject to errors from multiple sources. The critical value for damage detection is a threshold to which the difference of a point's elevation - before and after the earthquake - is compared to determine whether or not the building is damaged. The optimum threshold values differ considerably between different datasets. This may be as a result of shape and height of buildings together with the arrangement of collapsed buildings. Therefore, an optimum threshold value must be used in an adaptive manner. Considering conventional stochastic theories, we introduce a hypothesis test to improve the collapse detection without knowing optimal thresholds. We present an empirical investigation using two datasets of Kobe and Bam earthquakes. DSMs were created automatically from both pre- and post-earthquake aerial images. Also, a visual inspection of building damages was conducted, based on stereo pairs of aerial photos before and after the earthquake to generate reference data for evaluating the proposed method. The Kobe and Bam datasets contain vast varieties of real collapsed buildings and the results achieved for our datasets are very promising to detect collapsed buildings automatically. The overall accuracies are computed to be 91.8% and 82.7% for Kobe and Bam, respectively.

1. INTRODUCTION

After natural disasters, data obtained from satellites and airborne platforms are useful in providing an understanding of the distribution of damaged buildings. Comparing geometrical data in the form of digital surface model (DSM) can be major approach to detecting the extent of demolished area. DSMs can be derived from variety of sources requiring different processing methods before and after the disasters. In urban areas, using stereoscopic (multiple) aerial photographs and LIDAR technology are prevalent. These remotely sensed DSM production methods provide users with high resolution DSM data that have well-defined vertical and horizontal accuracies. As such, any resultant DSM is subject to both the precision and accuracy of the measurement sensor as well as the quality of image matching and interpolation methods, and thus is subject to errors from multiple sources. DSM-based change detection researches mainly propose to compute simple difference between DSMs from different epochs (Gong et al., 2000, Heller et al., 2001, Hollands et al., 2007). The building changes can be automatically detected by simply subtracting one of the DSM data sets from another (Turker & Cetinkaya, 2005, Maruyama et al., 2010). The critical value for damage detection is a threshold to which the difference of a point's elevation - before and after the earthquake - is compared to determine whether or not the building is collapsed (Rezaeian & Gruen, 2007).

Our overall objective in this paper is the development of reliable technique to detect significant changes of buildings after the earthquake using DSMs. For this purpose, considering conventional stochastic theories, we introduce a hypothesis test to improve the collapse detection without knowing optimal thresholds.

2. DSMs COMPARISON

2.1 Earthquake data sets

For empirical investigations, two datasets were obtained from aerial images. In our research, parts of the Kobe and Bam cities are selected as study regions. DSMs were created automatically from both pre- and post-earthquake aerial images using the SAT-PP software, which is efficient in-house developed software of IGP-ETHZ. The root mean square error value for checkpoints in pre- and post-earthquake DSMs are estimated 2.37 and 2.13 meter for Kobe, and 1.6 and 1.5 meter for Bam, respectively. A visual inspection of building damages is conducted, based on stereo pairs of aerial photos before and after the earthquake to generate reference data. If the height is reduced more than one meter, the building labeled as collapsed one.

2.2 Optimum threshold value

The reduction of point's elevation will be a significant cue to detect collapsed objects, however, simple pointwise comparison between pre- and post-event DSMs can't be a reliable evidence for damaged points. The average of height differences appears to be more reliable than a pointwise comparison. The critical parameter for collapse detection is an optimum threshold value for comparing elevation data in DSMs. This cannot be defined generally due to stochastically behaviors of the models in different areas. For all building polygons the difference of its pre- and post-event average elevation could be analyzed and classified as follow:
Average Height Difference (AHD):

$$\begin{array}{l} if \quad \bar{h}_{diff} = \bar{h}_{before} - \bar{h}_{after} > Threshold \\ then \quad \text{"}\textbf{Collapsed building"} \quad else \quad \text{"}\textbf{Uncollapsed building"} \end{array} \tag{1}$$

It closely resembles the method proposed by Turker and Cetinkaya (2005). Similarly, we (Rezaeian & Gruen, 2008) proposed the method by using ratio of building volumes (V_a/V_b) to detect collapsed buildings:
Volumes Ratio (VR):

$$\begin{array}{l} if \quad \frac{V_a}{V_b} < Threshold \\ then \quad \text{"}\textbf{Collapsed building"} \quad else \quad \text{"}\textbf{Uncollapsed building"} \end{array} \tag{2}$$

V_b,V_a denote building volumes before and after earthquake, respectively. The optimum threshold value can be determined by a method proposed by Fung and LeDrew (1988). This threshold is identified as a value that provides the maximum overall accuracy.

Table 1: Optimum threshold values computed from error matrices

Volumes Ratio (VR)	Overall accuracy % (Correct decision)		Average Height Difference (AHD)	Overall accuracy % (Correct decision)	
Threshold	Kobe	Bam	Threshold [m]	Kobe	Bam
0.10	69.4	59.2	0	78.8	55.5
0.15	74.7	64.6	0.5	90.3	57.0
0.20	78.3	70.4	1.0 ← Optimum for Kobe	**91.8**	61.0
0.25	80.1	75.2	1.5	90.1	66.0
0.30	81.5	79.7	2.0	88.2	70.7
0.35	82.9	83.3	2.5	86.3	76.7
0.40←Optimum for Bam	83.7	**83.7**	3.0	81.0	81.7
0.45	84.3	82.8	3.5	73.3	82.7
0.50	85.1	80.9	4.0 ← Optimum for Bam	70.5	**83.0**
0.55	85.7	78.0	4.5	67.5	78.8
0.60	86.7	75.8	5.0	59.5	69.9
0.65	87.4	71.1	5.5	50.4	62.6
0.70	88.9	68.2	6.0	45.8	56.7
0.75	90.1	64.5	6.5	44.3	52.8
0.80	91.8	61.5	7.0	44.0	50.0
0.85←Optimum for Kobe	**92.3**	59.3	7.5	44.0	48.3
0.90	91.2	57.4	8.0	43.8	46.7
0.95	87.6	56.5	8.5	43.8	46.3

Table 1 shows the overall accuracies using different thresholds. The thresholds of 0 for AHD and 0.1 for VR were chosen in the first iteration, and in each stage it is increased. Error matrices are produced and analyzed for each threshold calculating overall accuracies. Using volume ratio (VR), the maximum overall accuracies for Kobe and Bam dataset are appeared 92.3% and 83.7%, respectively. Although, there is not significant difference between AHD and VR methods, the optimum threshold values differ considerably between Kobe and Bam dataset. It may be as a result of shape and height of buildings together with arrangement of collapsed buildings. Therefore, optimum threshold value must be used in an adaptive manner. Setting up and testing hypotheses is an essential part of statistical inference, which can be a general solution. Considering conventional stochastic theories, we introduce following hypothesis test to improve collapse detection without knowing optimal thresholds.

3. HYPOTHESIS TEST

The question of interest is simplified into two competing claims/hypotheses between which we have a choice; the null hypothesis denoted H0: "*Building is not changed*", against the alternative hypothesis, denoted H1: "*Building is changed*". The outcome of a hypothesis test is 'reject H0' or 'do not reject H0'. Despite the fact that rejecting the null hypothesis does not imply accepting the alternative, the buildings are classified according to Table 2. The operations are carried out in an attempt to disprove or reject the null hypothesis and it cannot be rejected unless the evidence against it is sufficiently strong.

Table 2: Decision using hypothesis test

		Actual condition	
		H0 True (Changed)	**H0 False (Unchanged)**
Decision	**Do not reject H0**	Correct decision	False negative error
	Reject H0	False positive error	Correct decision

Statistical tests always involve a trade-off between the acceptable level of false positives (in which an uncollapsed is declared to be collapsed) and the acceptable level of false negatives (in which an actual damaged is not detected). It should also be noted that in damage detection false elimination of damaged buildings is typically much more costly than a false addition of uncollapsed buildings as collapsed ones.

The DSM elevation of given point P_n can be represented as a random process $\tilde{Z}_n$:

$$\forall P_n : (X_n, Y_n) \qquad \tilde{Z}_n = Z_n + e_n \tag{3}$$

Where, Z_n denotes actual elevation (deterministic value) and e_n denotes DSM error (stochastic process). Usually, both the magnitude and spatial distribution of the error at any particular location are unknown. Different types of error are often listed: blunders, systematic and random errors as being typical in DSMs. Blunders are gross errors, which occur less frequently in DSM products. The avoidance and detection of blunders in the automatically generated DSM by image matching are critical issues for current researches. Systematic errors show a common trend or dependency, and can be the results of processing or recording procedures. Random errors originate from a variety of sources, and no trend can be observed (Fisher & Tate, 2006). We define a new variable of difference between before and after heights:

$$\begin{aligned} &\forall P_n : (X_n, Y_n) \qquad d_n = \tilde{Z}_n^b - \tilde{Z}_n^a = (Z_n^b + e_n^b) - (Z_n^a + e_n^a) \\ &\textit{if } (P_n \in \text{Unchanged Object}) \textit{ then: } \quad Z_n^b = Z_n^a \quad \rightarrow \quad d_n = e_n^b - e_n^a \end{aligned} \tag{4}$$

Where, $\tilde{Z}_n^b$ & $\tilde{Z}_n^a$ denote random variables of DSMs elevations, Z_n^b & Z_n^a denote true elevations of P_n and e_n^b & e_n^a represent DSM errors, before and after earthquake respectively. The central limit theorem is critical to applying inferential statistics and hypothesis testing. Let d_1, d_2, d_3 ... d_N be a sequence of N independent and identically distributed (i.i.d) random variables each with finite mean and variance. The central limit theorem states that as the sample size N increases, the distribution of the sample average of these random variables approaches the normal distribution irrespective of the shape of the original distribution. There are many versions of the central limit theorem. Several of these place additional restrictions but do not require being identically distributed. Generally the additional restrictions are designed to prevent one or a handful of random variables from dominating the average, which might happen if one random variable has a standard deviation far greater than the rest (Lyapunov or Lindeberg conditions) (Durrett 1996). Roughly speaking, a sum of many small independent random variables will be nearly normally distributed. Although, this assumption might be invalidated locally due to systematic errors, we assume that the differences between systematic errors for DSMs (before and after) are negligible if both DSMs are generated by one system through similar procedures. We define following one-sided hypothesis test:

$$
\begin{aligned}
&\forall P_n \;\; within \;\; building \;\; polygon: \quad \bar{d} = \frac{1}{N}\sum_n d_n \\
&\begin{cases} H0: building \;\; is \;\; unchanged & \bar{d} = \mu_0 + \delta \\ H1: building \;\; is \;\; changed & \bar{d} > \mu_0 + \delta \end{cases}
\end{aligned}
\tag{5}
$$

The p-value provides an objective measure of the strength of evidence, which the data supplies in favor of the null hypothesis. For all building polygons, the following p-value is calculated and the value of δ which cause to reject null hypothesis is computed:

$$
\begin{aligned}
&p - value = \frac{\bar{d} - \mu_0 - \delta}{\sigma_0 / \sqrt{N}} \\
&\alpha = 0.05 \qquad if \quad p\text{-}value > 1.645 \Rightarrow Reject\ H0
\end{aligned}
\tag{6}
$$

Here, μ_0 and σ_0 represent the sample mean and standard deviation of $\{d_n\}$ for undamaged buildings, respectively, and δ denotes an estimation of height reduction for collapsed buildings. N denotes total number of points surrounded by building polygon, which has to be sufficiently large. Therefore, the hypothesis test could not be reliable for polygons with small area.

4. EMPIRICAL INVESTIGATIONS

4.1. Automatic collapse detection using pre- and post-event DSMs

After automatically generating the DSMs from both pre- and post-earthquake aerial photographs, in order to obtain μ_0 and σ_0, some uncollapsed buildings are selected. Each building polygon encompasses several points of DSMs (before and after) and so $\{d_n\}$ can be calculated for uncollapsed sample buildings. μ_0 and σ_0 are estimated using following estimators:

$$
\mu_0 = \frac{1}{M}\sum_{n=1}^{M} d_n \quad , \; \sigma_0 = \sqrt{\frac{\sum_{n=1}^{M}(d_n - \mu_0)^2}{M-1}}
\tag{7}
$$

Where, M denotes total number of points belong to fifteen samples of uncollapsed buildings. We applied the above-mentioned hypothesis test for any building of both datasets. In a conventional test, it is assumed that δ equals zero and H0 is rejected if p-value would be greater than 1.645 (for $\alpha = 5\%$). We determine the boundary value of δ for rejecting H0 and make a decision based on the δ value. Obviously, it is anticipated that the greater the demolition, the larger the value of δ. For bi-level classification assuming that if $\delta < 1$m then the building is classified as "Uncollapsed" and $\delta \geq 1$m the building is classified as "Collapsed". This value is selected according to our criteria to interpret collapsed buildings (c.f section 2.1). Table 3 shows the numerical results. The overall accuracies are computed to be 91.8% and 82.7% for Kobe and Bam, respectively. Producers' accuracy (i.e. 1 - false negative error) for collapsed buildings is computed 90.3% and 72.0% for Kobe and Bam, respectively. In Kobe, a concentration of heavily collapsed structures plus a few numbers of partly collapsed buildings may be the cause of the better producer's accuracy in comparison with the Bam dataset.

Table 3: Results of hypothesis tests for damage detection using pre- and post-event DSMs

Hypothesis test		Visual interpretation			
		Kobe		Bam	
Decision	δ [m]	Uncollapsed	Collapsed	Uncollapsed	Collapsed
Uncollapsed	$\delta < 1$	260	35	389	135
Collapsed	$1 \leq \delta$	17	325	19	347
Overall accuracy		91.8%		82.7%	

One important result is that using the proposed hypothesis test, the computed overall accuracy will be very close to overall accuracy of using optimum threshold values of both Kobe and Bam data set (Table 1 and Table 3). In addition, false positive errors that are calculated for Kobe: (17÷277)×100 = 6.1% and for Bam (19÷408) ×100 = 4.7%, appear to be remarkably close to $\alpha = 5\%$. These numerical results indicate to us that performed hypothesis test is supported by an appropriate statistical model. This approach is promising for detection of collapsed buildings using pre- and post-event DSMs of any dataset.

4.2. Automatic collapse detection using pre–event building models and post-event DSM

The average height difference (AHD) between post-event DSM and pre-event model of buildings can be used for detecting damaged buildings as well. It is assumed that the pre-event building models have already been generated and are available after the earthquake. In this research, the prismatic models (building polygons with their heights) are generated manually using a digital photogrammetry workstation. Figure 1 depicts the distribution of uncollapsed buildings with respect to AHD between pre- and post-event DSMs in comparison with the AHD between prismatic models and post-event DSM. It shows that for both datasets the histograms (probability density functions) are similar in appearance to Gaussian distributions, however, the mean and variance are changed. Therefore, the proposed hypothesis test can be performed using the new values of mean and variance of AHD for DSM and prismatic model of uncollapsed buildings.

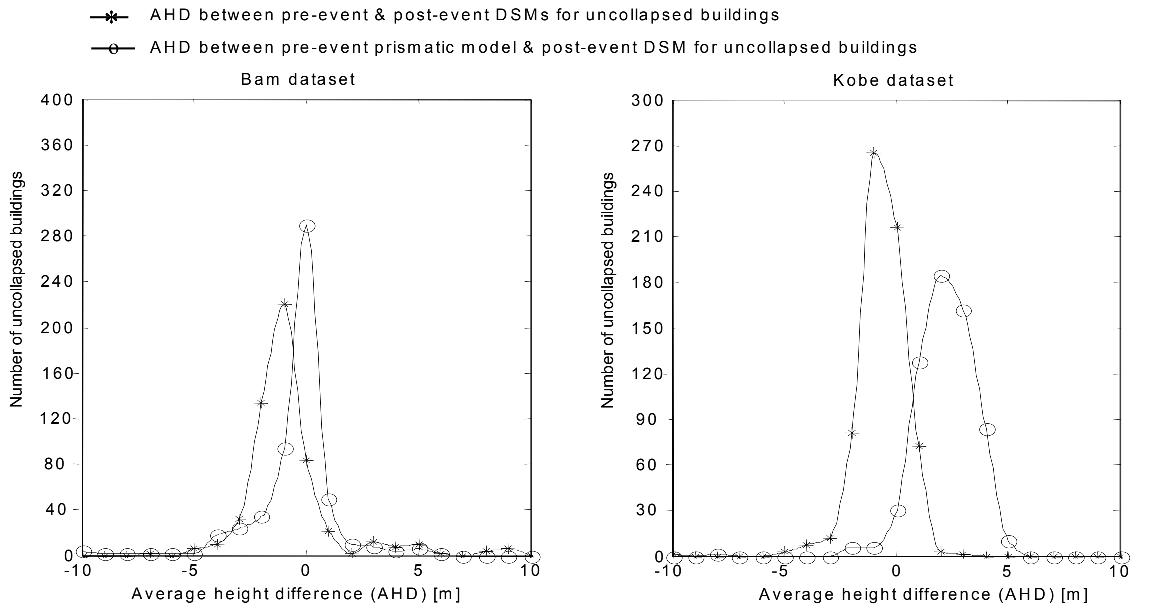

Figure 1: Histogram of uncollapsed buildings vs. average height difference

Once again, the mean (μ_0) and standard deviation (σ_0) of AHD between building models and DSM points for fifteen uncollapsed buildings are estimated. The hypothesis test is performed for all buildings in the Bam and Kobe datasets. Table 4 presents numerical results. The overall accuracy is computed to be 80.3% for the Bam dataset, which is close to the result of comparing pre- and post-event DSMs (82.7% in Table 3). However, the accuracy of the classifier significantly has decreased for Kobe dataset (79.6% in Table 4 vs. 91.8% in Table 3).

Table 4: Results of hypothesis tests for collapse detection using pre-event building prismatic models and post-event DSM

Hypothesis test		Visual interpretation			
		Kobe		Bam	
Decision	δ [m]	Uncollapsed	Collapsed	Uncollapsed	Collapsed
Uncollapsed	$\delta < 1$	250	103	396	163
Collapsed	$1 \le \delta$	27	257	12	319
Overall accuracy		79.6%		80.3%	

In Bam city, the majority of buildings are simple with onefold rooftop, therefore, the prismatic model perfectly represents most of the buildings. On the other hand prismatic models are unable to exhibit detailed surfaces of modern buildings and complex rooftops in Kobe city. Nevertheless, it indicates to us that 3D city modeling is an essential pre-event task for disaster management and generates extremely valuable data for damage assessment.

5. CONCLUSIONS

In this research, approaches for detecting "Collapsed" and "Uncollapsed" buildings using surface models were presented. We focused on three-dimensional information extracted before and after the earthquake. Although the reduction of point's elevation will be a significant cue to detect collapsed objects and we are utilizing DSMs generated automatically by SAT-PP, which is one of the best image matcher software, the systematic errors and blunders especially in densely constructed regions will be unavoidable. To model the statistical behavior of DSM points we applied hypothesis tests for the mean value of height difference between clustered points, which are enclosed by the building polygon. The hypothesis test suggested in this paper shows that normalized value of "average height differences" (AHD) could provide the optimum overall accuracy for bi-level classification. However, the mean and variance should be estimated from some undamaged sample buildings. Sample buildings selected for evaluating the mean and variance (Equation 7) need to be distributed almost uniformly in a test area including variety of the buildings. The DSM blunders and error fluctuations especially in steep slopes of buildings are the main reasons for the misclassifications. The accuracy of the DSMs - generated automatically - becomes worse in these areas with densely manmade objects. Considering rooftops details, we may replace pre-event DSM with 3D models of buildings. The Kobe and Bam datasets contain vast varieties of real collapsed buildings and the results achieved for our dataset are very promising. However, airborne LIDAR data can be used allowing a rapid and extensive acquisition of height data and the advantages of such DSMs is that the height component is usually better than the height component of DSMs acquired by matcher software.

6. REFERENCES

Durrett, Richard (1996). Probability: theory and examples (Second ed.)

Fisher, P. F., Tate, N. J., 2006. Causes and consequences of error in digital elevation models. Progress in Physical Geography 30, pp. 467–489.

Fung, T., Ledrew, E., 1988. The determination of optimal threshold levels for change detection using various accuracy indices. Photogrammetric Engineering and Remote Sensing, 54(10), pp 1449–1454.

Gong, P., Biging, G. and Standiford, R., 2000. Use of Digital Surface Model for Hardwood Rangeland Monitoring. Journal of Range Management 53(6), pp. 622–626.

Heller, A., Leclerc, Y. and Luong, Q., 2001. A Framework for Robust 3-D Change Detection. Proc. of SPIE, Toulouse, France.

Hollands, T., Bostrm, G., Goncalves, J. G., Gutjahr, K., Niemeyer, I. and V.Sequeira, 2007. 3D Scene Change Detection from Satellite Imagery. Proc. 29th Symp. on Safeguards and Nuclear Material Management, Aix-en-Provence, France pp. 1–6.

Maruyama, Y., Tashiro, A., Yamazaki, F., 2010. Detection of Collapsed Buildings after the 2007 Niigata Chuetsu-oki Earthquake based on Digital Surface Model Constructed from Aerial Images, Proceedings of the Second International Symposium on Advanced in Urban Safety, pp. 319-324.

Rezaeian, M. and Gruen, A., 2007. Automatic classification of collapsed buildings using object and image space features. In: J. Li, S. Zlatanova and A. Fabbri (eds.), Geomatics solutions for Disaster Management. Springer, pp. 135-148.

Rezaeian, M., Gruen, A., 2008. Automatic classification of collapsed buildings using digital surface models. Presented at Second Disaster Management Conference, Tehran, Iran, On CDROM, 8p.

Turker, M, Cetinkaya, B., 2005. Automatic detection of earthquake damaged buildings using DEMs created from pre- and post-earthquake stereo aerial photographs. International Journal of Remote Sensing, vol 26(4), pp. 823–832.