

## Efficacy of Damage Detection Measures from Satellite Images

Jim Thomas<sup>1</sup>, Joe Jeray<sup>2</sup>, Ahsan Kareem<sup>3</sup>, Kevin Bowyer<sup>4</sup>

<sup>1</sup>Computer Science and Engineering, Univ. of Notre Dame, IN, USA, [jthoma14@cse.nd.edu](mailto:jthoma14@cse.nd.edu)

<sup>2</sup>Civil Engineering, Univ. of Notre Dame, Notre Dame, IN, USA, [jjeray@nd.edu](mailto:jjeray@nd.edu)

<sup>3</sup>Civil Engineering, Univ. of Notre Dame, Notre Dame, IN, USA, [kareem@nd.edu](mailto:kareem@nd.edu)

<sup>4</sup>Computer Science and Engineering, Univ. of Notre Dame, IN, USA, [kwb@cse.nd.edu](mailto:kwb@cse.nd.edu)

### ABSTRACT

*Estimating the extent of damage caused by natural disasters is necessary for implementing effective recovery measures. Aerial images of affected areas are easily obtained through satellite or aerial sensors. A careful analysis of images from before and after an event facilitates rapid detection and assessment of damage. Significant previous research has been done on developing measures to quantify the damage. In this study we evaluate the efficacy of existing change measures used to estimate damage. Determining the efficacy of these damage measures in definitive characterization of damage states is necessary for accurate automated assessment of windstorm damage.*

### INTRODUCTION

Pre- and post-storm aerial images have been used for a rapid and thorough assessment of damage following a hurricane. The first known opportunity for using this technology was with Hurricanes Charley and Ivan in 2004 [6]. The collection and analysis of these images provided opportunities for faster damage estimation. This enabled planning of quick recovery, as well as rescue efforts required to mitigate human suffering and property loss. The development of an automated damage estimation process is required for fast and efficient damage estimation.

A major component of damage assessment is comparing pre- and post-storm images for change detection. The damage in an area is represented by change index values and categorized into defined damage states. However, the problem is complicated by the fact that non-windstorm factors can lead to observed change. The presence of atmospheric disturbances, differences in shadows due to images being taken at different times of the day and inaccuracies in registration are just some of the factors. The efficacy of a detection measure is demonstrated by its ability to correctly detect and quantify the amount of change caused by windstorm damage while at the same time minimizing the effect of noise factors.

Recent studies have investigated the use of modern remote-sensing technology for windstorm damage assessment. Womble [6],[7] suggested the use of measures obtained from spectral bands for a per-building classification of windstorm damage. In this study, a framework for comparing the efficacy of modern change detection algorithms in damage estimation is presented. First, the general steps used for damage detection are explained. Next, a framework that can be used for comparison of change detection algorithms is presented. Then our evaluation of existing algorithms is offered. Finally, the correlation between change detection measures and windstorm damage states [6] is examined.

## DAMAGE DETECTION

Building damage and debris spread can be identified and classified by applying change-detection algorithms to pre- and post-storm image pairs. In this comparison step, building damage appears as changes in shape, lines, colors, texture, or other image properties. Previous research has shown that the severity of damage to buildings can be estimated from the extent of change in the roof structure. We performed our study on 6 before-and-after image pairs, taken from Hurricane Charley (2004), Hurricane Dennis (2005) and Hurricane Ivan (2004). The images were taken from Pensacola, Fort Walton Beach, and Punta Gorda, Florida. Figure 1.1 shows a pair of pre- and post-storm Hurricane Charley images and Figure 1.2 shows all the manually marked damages areas colored in red.



Figure 1.1: Pre- and post-storm images(grayscale) from Womble [2005]. QuickBird imagery from DigitalGlobe, Inc.[[www.digitalglobe.com](http://www.digitalglobe.com)].

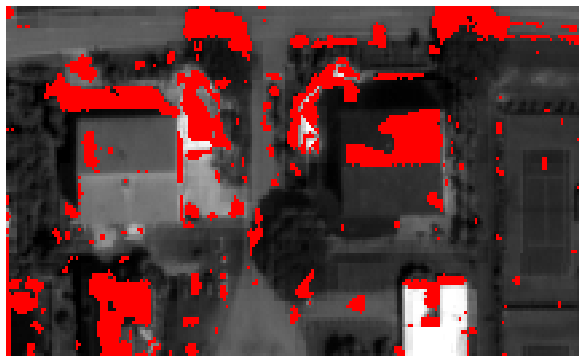


Figure 1.2: Damage regions, in red, manually marked from pre- and post-storm images in Figure 1.1.

The damage detection process is a sequence of operations. The first step is pre-processing that involves image registration, contrast equalization and cropping of common area. The next step is to identify regions of the image that represent damage. For example, in windstorm damage the objects of primary interest are buildings and damage is indicated by changes in roof structure. We can identify such objects manually or by using image segmentation techniques. Then change detection algorithms are used to identify the degree of change that has occurred between the before and after images. The next step is to compute damage metrics that quantify the extent of damage. Lastly the damage metrics are used to classify regions of the image into categories ranging from no damage to severely damaged.

## FRAMEWORK FOR EVALUATION OF CHANGE DETECTION ALGORITHMS

In order to evaluate and compare different change detection algorithms, we use a set of pre- and post-storm images for which we have manually specified the “ground truth” of damage locations. The results of a given change detection algorithm are evaluated according to how well they agree with the ground truth. This enables a quantitative comparison of algorithms according to their performance on defined metrics.

### IMAGE PRE-PROCESSING

Satellite images of the same location are in general not aligned correctly. Differences may be due to change in position, angle and location from which the images were taken. In our current pre-processing step, control points are manually marked in the pre- and post-storm images. The control points are typically marked at distinguishable locations like corners of buildings, roads, etc. In Figure 2, twelve corresponding control points are shown. A geometric transform is then applied to register the images. The images are then cropped to include only the region belonging to both of images.

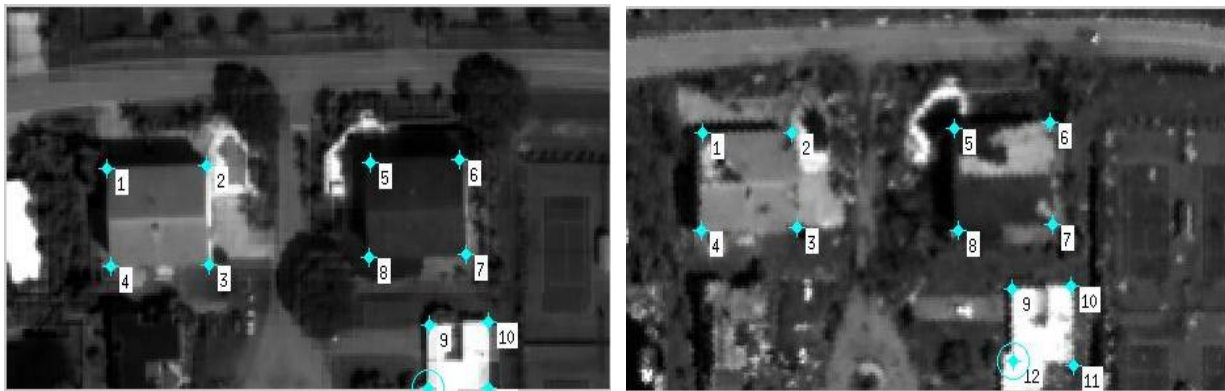


Figure 2: Pre- and post-storm Images with control points marked at corners of buildings.



Figure 3: Pre- and post-storm images after registration and cropping.

## CHANGE DETECTION

Change detection algorithms use different criteria to classify pixels as representing change or not. A survey of existing change detection algorithms, though not from the specific viewpoint of windstorm damage assessment, is presented in [5]. The simplest change detection algorithm is based on a plain differencing of corresponding pixels in the pre- and post-storm images. This technique thresholds a difference image to obtain a change mask. A pixel is colored white in the change mask if there is a change and black otherwise. (See Figure 4.) A more complicated approach uses significance tests in which the decision at each pixel depends on a local neighborhood of pixels. Algorithms based on shading models use intensity at pixels. In this study we compare change detection algorithms of various types.



Figure 4: A change mask produced by performing Simple Differencing algorithm on images from Figure 3.

## GROUND TRUTH

Damage areas that are manually marked in our ground truth change masks include rooftops, missing trees, missing pools, and debris. Changes such as shadows due to buildings are not marked, as these changes could not represent storm damage.

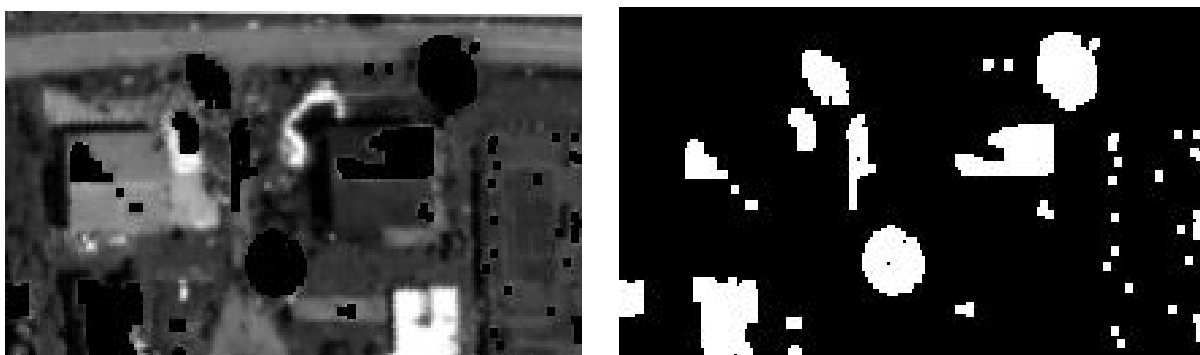


Figure 5: A manually marked ground truth and corresponding change mask.

## COMPARISON

The goal is to compare detected change with ground truth. This comparison requires an algorithm that tells us about the accuracy of the detected change in an objective, quantitative manner. We propose to use the measures suggested in [4]. For the measures to be computed, the

change mask and ground truth should be separated into different regions. For example, damage to building 1 would be a region, damage to building 2 would be another region etc. In order to classify a change mask into these regions, we use a connected components algorithm [3]. Then the quantitative measures are obtained from the comparison as in [4].

The comparison process is summarized as follows. One, apply a connected components algorithm to label each damage region as a separate segment. Two, the labeled images (Ground truth & Change Detection) are passed to the comparison algorithm [4]. Three, the comparison algorithm produces a set of quantitative measures. Four, steps 2 and 3 are repeated for different threshold values in change detection until the best possible match is found.

Firstly the connected components labeling scans an image and groups its pixels into components based on connectivity. Pixels in a connected component share similar pixel intensity values and are spatially connected with each other. In our evaluation each connected component can be considered to be a separate damage region. Each damage region is assigned a unique number. This process is shown in Figure 6. The labeled regions in the ground truth (GT) and the change detected (CD) images are then classified as in [4].

We consider five types of result for the comparison of detected damage regions to ground-truth damage regions: correct detection, over-segmentation, under-segmentation, missed, and noise [4]. The formulas for deciding classifications are based upon threshold  $T$ , for the mutual overlap between two regions, where  $0.5 < T \leq 1.0$ . The value of  $T$  can be set to reflect the strictness of definition desired. The following metrics define each classification:

1) An instance of **correct detection**.

A pair of regions  $R_n$  in the GT image and  $R_m$  in the CD image are classified as an instance of correct detection if a) At least  $T$  percent of the pixels in region  $R_m$  in the CD image are marked as pixels in region  $R_n$  in the GT image, and b) At least  $T$  percent of the pixels in region  $R_n$  in the GT image are marked as pixels in region  $R_m$  in the CD image.

2) An instance of **over-segmentation**.

A region  $R_n$  in the GT image and a set of regions in the CD image are classified as an instance of over-segmentation if a) At least  $T$  percent of the pixels in each region in the CD image are marked as pixels in region  $R_n$  in the GT image, and b) At least  $T$  percent of the pixels in region  $R_n$  in the GT image are marked as pixels in the union of regions in the CD image.

3) An instance of **under-segmentation**.

A set of regions in the GT image and a region  $R_m$  in the CD image are classified as an instance of under-segmentation if a) At least  $T$  percent of the pixels in region  $R_m$  in the CD image are marked as pixels in the union of regions in the GT image, and b) At least  $T$  percent of the pixels in each region in the GT image are marked as pixels in region  $R_m$  in the CD image.

4) An instance of **missed** classification. A region  $R_n$  in the GT image that does not participate in any instance of correct detection, over-segmentation or under-segmentation is classified as missed.

5) An instance of a **noise** classification.

A region  $R_m$  in the CD image that does not participate in any instance of correct detection, over-segmentation or under-segmentation is classified as noise.

Although these definitions result in a classification for every region in the GT and CD images, they are not unique for  $T < 1.0$ . However, for  $0.5 < T < 1.0$  any region can contribute to at most three classifications, one each of correct detection, over-segmentation and under-segmentation. An example of these classifications corresponding to GT and CD images of Figure 6 is shown in Table 1.

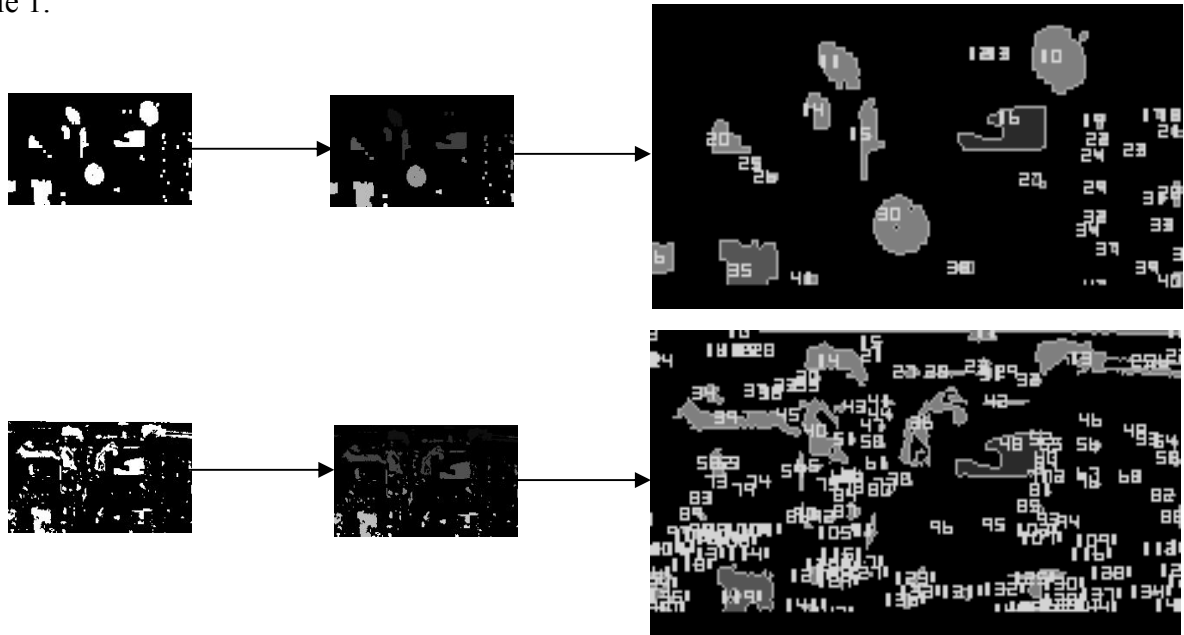


Figure 6: The pictures show the comparison process for ground truth (GT) and change detection (CD). The different connected components in the change masks are computed and labeled by using different pixel intensities. Then the connected components are numbered and compared to obtain the quantitative measures.

Table 1: An example of region classifications for GT and CD images of Figure 6 is shown here. GT region and CD region are the label numbers of the corresponding regions in the images. Measure 1 and measure 2 are the confidence measures for GT and CD respectively.

GT region	Category	CD Region	Measure 1	Measure 2
16	Correct	48	0.935	0.980
18	Correct	54	0.778	0.100
21	Correct	58	0.750	0.667
23	Correct	68	0.667	1.000
27	Correct	85	0.840	0.840
37	Correct	128	0.778	1.000
38	Correct	131	0.926	0.962
44	Correct	155	0.778	1.000
35	Over segmented	119,148,158	0.775	0.980

It should be noted here that the number of correct regions found also depends on the threshold parameter of each change detection algorithm. For a fair comparison process each algorithm should use the optimal threshold value. In order to tune each algorithm to the right threshold for it we need to try a range of threshold values and choose the value corresponding to the largest number of correct regions detected.

## EVALUATION OF CHANGE DETECTION ALGORITHMS

We use the comparison framework outlined above, based on [4], to evaluate the change detection algorithms discussed in [5]. The simplest class of algorithms just thresholds the difference between two images. That is, the change mask  $B(x)$  is generated according to the following decision rule:

$$B(x) = \begin{cases} 1 & \text{if } |D(x)| > T \\ 0 & \text{otherwise} \end{cases}$$

$D(x)$  is the difference image and  $T$  is an empirically chosen threshold such that  $0 < T < 1.0$ . This process is known as **simple differencing**.

Another class of change detection algorithms uses a **statistical hypothesis test** to make a decision. The decision as to whether or not a change has occurred at a given pixel  $x$  corresponds to choosing one of two competing hypotheses: the null hypothesis  $H_0$  or the alternative hypothesis  $H_1$ , corresponding to no-change and change decisions, respectively. The image pair  $(I_1(x), I_2(x))$  is viewed as a random vector. Knowledge of the conditional joint probability density functions (pdfs)  $p(I_1(x), I_2(x)|H_0)$  and  $p(I_1(x), I_2(x)|H_1)$  allows us to choose the hypothesis that best describes the intensity change at  $x$  using the classical framework of hypothesis testing. In our evaluation we use a statistical change detection algorithm that uses Gaussian model for the noise. Under the null hypothesis, unchanged regions in the difference image are modeled as zero mean Gaussian and the noise variance is estimated from the images. Characterizing the alternative (change) hypothesis  $H_1$  is more challenging, since the observations consist of change components that are not known a priori or cannot easily be described by parametric distributions. When both conditional pdfs are known, a likelihood ratio can be formed as:

$$L(x) = p(D(x)|H_1)/p(D(x)|H_0)$$

This ratio is compared to a threshold defined as:

$$T = P(H_0)(C_{10} - C_{00}) / (P(H_1)(C_{01} - C_{11}))$$

where  $P(H_i)$  is the a priori probability of hypothesis  $H_i$ , and  $C_{ij}$  is the cost associated with making a decision in favor of hypothesis  $H_i$  when  $H_j$  is true. In particular,  $C_{10}$  is the cost associated with “false alarms” and  $C_{01}$  is the cost associated with “misses”. This algorithm is referred to as **likelihood ratio test**.

Several change detection techniques are based on the shading model for the intensity at a pixel. Such algorithms generally compare the ratio of image intensities  $R(x) = I_0(x)/I_1(x)$ . This uses a null hypothesis that assumes linear dependence between vectors of corresponding pixel intensities, giving rise to the test statistic  $R(x)$ . In our evaluation we use the **linear dependence algorithm** suggested by Skifstad and Jain [2] to assess the linear dependence between two blocks of pixel values.

More sophisticated change detection algorithms result from exploiting the relationships between nearby pixels. These algorithms are classified under predictive models. In this approach to

change detection we fit the intensity values of each block to a polynomial function of the pixel coordinates. Hsu et al. [1] discussed generalized likelihood ratio tests using a **constant**, linear, or **quadratic** model for image blocks. The null hypothesis in the test is that corresponding blocks in the two images are best fit by the same polynomial coefficients, whereas the alternative hypothesis is that the corresponding blocks are best fit by different polynomial coefficients. In each case, the various model parameters are obtained by a least-squares fit to the intensity values in one or both corresponding image blocks. An extension of this algorithm uses **mean derivatives** of the quadratic models at each pixel in the block.

## EXPERIMENTAL RESULTS

We compared the change detection algorithms discussed above using six pairs of pre- and post-storm images. We computed the region classifications for (1) simple differencing algorithm, (2) statistical hypotheses test using a Gaussian noise model, (3) linear dependence algorithm, and (4) polynomial algorithms- constant, quadratic and mean derivative. The change-detected regions were classified as correct, under-segmented, over-segmented, missed and noise. For each algorithm the change masks were computed for 40 detection threshold values ranging from 0 to 1, and for each algorithm the value giving the best result for that algorithm was kept. This is to ensure that each algorithm is well tuned. The change mask corresponding to the maximum number of correct classifications was chosen for calculating the result. Table 2 summarizes the results by averaging the region classifications for the 6 pairs of images. Figures 7 and 8 show detailed results for correct and noise classifications, respectively.

From Table 2 it is evident that the simple differencing algorithm and the statistical hypothesis test algorithm outperformed the other algorithms. The statistical hypothesis test algorithm found a slightly larger number of correct classifications in comparison to the simple differencing algorithm. However from Figure 8 we can conclude that the frequency of noise regions is the highest for the simple differencing algorithm. Among polynomial algorithms, the quadratic model performed better than the constant model. The extension of the quadratic model performed better than other two models. The constant model had the least number of noise classifications among the polynomial models. The shading model approach found the least number of correct classifications, showing that measure based on linear dependence is least desirable.

**Table 2: Average region classifications for 6 pair of before-and-after images. The average of regions classified as correct, under-segmented, over-segmented, missed and noise were computed for each change detection algorithm evaluated.**

	Correct	Under-segmented	Over-segmented	Missed	Noise
Simple Diff.	7.67	0.34	0.16	16.00	224.00
Stat. Hyp. Test	7.83	0.16	0.00	16.16	65.34
Linear Dep.	0.50	0.00	0.00	23.67	74.16
Polynomial-const.	1.00	0.00	0.00	23.17	98.50
Polynomial-quad.	2.16	0.00	0.00	22.00	102.50
Mean derivative	2.34	0.16	0.16	21.50	129.00

Using best performance result, most algorithms seem to have relatively fewer regions classified as over- or under-segmented. The largest number of under-segmented regions is found in simple differencing. Figure 7 shows that the statistical hypothesis test approach finds the highest number of correct classifications for almost all image pairs. As it uses a Gaussian model for noise variance, statistical hypothesis test also finds the least number of noise classifications. This can be seen in Figure 8. The number of missed classifications is considerably larger than the number of correct classifications, as we do not mark any change that does not represent storm damages in the ground truth.

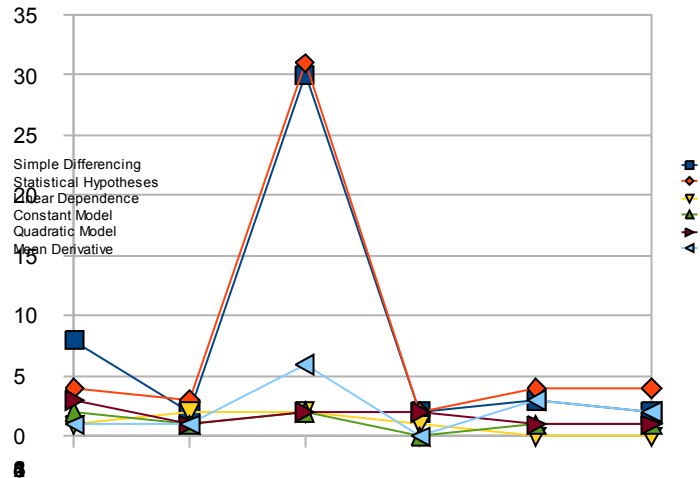


Figure 7: The chart shows the number of correct classifications for each of the image pairs numbered from 1 to 6. The number of correct classifications for each algorithm is shown separately.

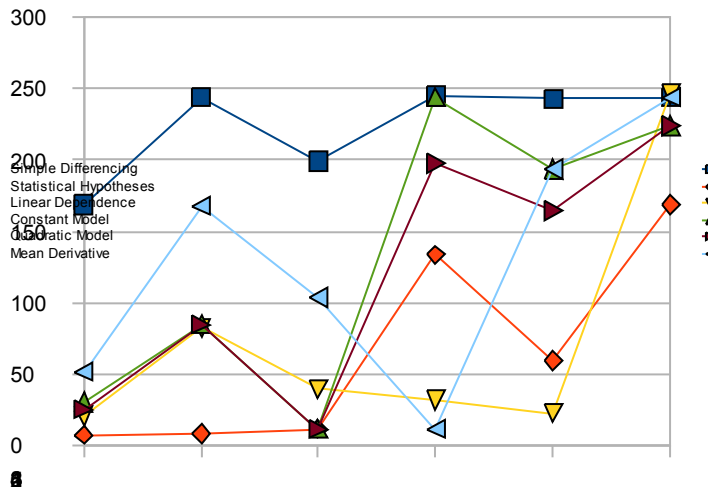
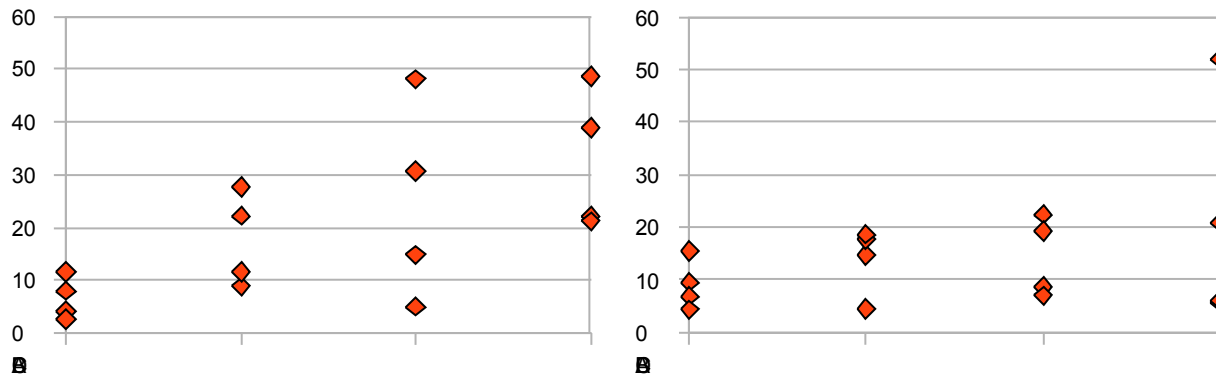


Figure 8: The chart shows the number of noise classifications for each of the image pairs numbered from 1 to 6. The noise classifications for each algorithm are shown separately.

## CHANGE DETECTION MEASURES IN DAMAGE CATEGORY ESTIMATION

Studying the correlation between the quantitative measures obtained from change detection and the level of damage actually present is important for using them in classifying damage in an automated system. For qualitative characterization of building damage, Womble suggested a classification based on roof structure damage [6]. He identified certain visual signatures of wind damage that could be used to categorize damage into four categories: RS-A (no damage), RS-B (tiles removed), RS-C (Decking removed) and RS-D (roof structure collapsed).



**Figure 9: A windstorm damage profile for damage metrics obtained from simple differencing (left) and for statistical hypothesis testing (right). The damage metrics are based on our own assigned ratings for the Remote-Sensing Damage Scale suggested in Womble [6] based on pre- and post-storm comparison of 16 buildings in our image dataset.**

Figure 9 show plots of windstorm damage profile categories as suggested by Womble, versus damage metric values from the simple differencing and the statistical testing change detection, respectively. For these plots, 16 residential buildings were selected from our image dataset, with Remote-Sensing Damage Scale values distributed as: RS-A (4), RS-B (4), RS-C (4), and RS-D (4). The change detection measure used here is expressed as a percentage of pixels that have been detected as changed in the pre- and post-storm roof top images. These damage profiles show a general correlation of the change detection metrics with damage states. The damage profile for simple differencing exhibits a stronger correlation as compared to statistical testing. While there is a clear distinction between no damage and other damage states, distinction between mid-level damage states (RS-B or RS-C) is difficult. However, the data contains too much scatter to accurately assign one of the four damage states based on a given value of damage metric. It may be possible to achieve greater accuracy by taking into consideration other factors such as the changes detected due to debris.

## CONCLUDING REMARKS

A systematic technique for comparing performance of modern change detection algorithms in identifying damage from a pre- and post-storm image pair is proposed. Changes detected by these algorithms included roof top damage, debris, missing trees and other displaced structures. Statistical hypotheses testing outperformed other algorithms by detecting maximum number of correct classifications with a low noise level. The change measures obtained from these algorithms can be used to identify damage in any part of a satellite image in a quick and efficient

manner. Our study on measuring the correlation between damage measures and different levels of roof top damage revealed that while there are strong trends for distinction between damage states, an accurate classification based simply on the given value of damage metric is difficult. Future work needs to be done in combining measures obtained from change detection with other measures that describe feature organization, debris scattering etc. A more accurate automated assessment of windstorm damage might be achieved by combining these measures.

## ACKNOWLEDGMENTS

The support for the study was provided by the Collaborative Research Program at the NatHaz Laboratory as a part of the Global Center of Excellence at Tokyo Polytechnic University: *New Frontier of Education and Research in Wind Engineering*. The center is funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. The opinions presented in this paper do not necessarily reflect the views of the collaborators or the funding agency. We would also like to thank Dr. A. Womble for providing us with the sample images used in his work and for sharing hurricane images.

## REFERENCES

- [1] Y. Z. Hsu, H.-H. Nagel, and G. Reckers, "New likelihood test methods for change detection in image sequences," *Computer Vision, Graphics, and Image Processing*, vol. 26, 1984, pp. 73–106.
- [2] K. Skifstad and R. Jain, "Illumination independent change detection for real world image sequences," *Computer Vision, Graphics, and Image Processing*, vol. 46, no. 3, 1989, pp. 387–399.
- [3] Dillencourt, M. B., Samet, H., and Tamminen, M, "A general approach to connected-component labeling for arbitrary image representations", *J. ACM* 39, 2 (Apr. 1992), 253-280.
- [4] Adam W. Hoover, Gillian Jean-Baptiste, Xiaoyi Jiang, Patrick Flynn, Horst Bunke, Dmitry Goldgof, Kevin W. Bowyer, David Eggert, Andrew Fitzgibbon, and Robert Fisher. "An Experimental Comparison of Range Image Segmentation Algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18, (7), 1996, pp. 673-689.
- [5] R..J Radke , S. Andra , O Al-Kofahi , B Roysam, "Image Change Detection Algorithms: A Systematic Survey", *IEEE Transactions on Image Processing*, 14(3):294-307 March 2005.
- [6] J. A. Womble, "Remote-Sensing Applications to Windstorm Damage Assessment", *Doctoral Dissertation in Civil Engineering, Texas Tech University, December 2005*.
- [7] J. A. Womble, K. C Mehta, B. J Adams, "Remote-Sensing Assessment of Wind Damage", *5th International Workshop on Remote Sensing Applications to Natural Hazards, September 2007*.