Towards a robust, automated hurricane damage assessment from high-resolution images

Jim Thomas [a], Ahsan Kareem [b], Kevin Bowyer [c]

[a]*University of Notre Dame, Notre Dame, IN, USA, jthoma14@nd.edu*
[b]*University of Notre Dame, Notre Dame, IN, USA, kareem@nd.edu*
[c]*University of Notre Dame, Notre Dame, IN, USA, kwb@nd.edu*

1 INTRODUCTION

In the event of a natural disaster such as hurricane or earthquake, estimating the extent of damage is necessary for implementing fast and effective recovery measures. Images of affected areas are easily obtained through satellite or aerial sensors. The key objects of interest in such images are buildings, as damage directly impacts lives. This work aims to build a system capable of fine-grained damage analysis by comparing before and after storm images. While much of this type of assessment is currently still done manually, we present an automated system for hurricane damage analysis. Previous approaches were limited by the number of control points in manual registration (Bitelli et al, 2004). They (Chen and Hutchinson, 2009; Matsuoka and Yamazaki, 2004; Womble et al 2007) either do not extract buildings or only provide approximate estimates. Previous methods in damage classification do not combine color-based and edge-based measures, while a fusion of the two is explored this study. Thus the work presented in this study aims at minimizing the manual interaction required in previous works and introduces new techniques for damage classification. We also deal with the challenge of creating an automatic system that can work with aerial or satellite images. The main contributions of this work can be summarized as: 1) application of a pre-processing stage where images are registered and roof-tops are extracted by novel algorithms, and 2) proposal and evaluation of intuitive, high-level features that can be used to distinguish damage states of a roof-top.

In section 2, we describe the first step in pre-processing the images, image registration. Accurate image registration is required for damage analysis from before and after event images. We discuss a reliable and fast automatic technique that computes a large number of landmark points. Additionally, an efficient and robust way to match corresponding landmark points is described. In section 3, we introduce the second step in pre-processing; a novel algorithm that detects buildings of a broad range of variations in shape, texture or color. Introducing this robust object extraction technique into the framework of damage detection has opened up new possibilities in analysis from high-resolution imagery. Finer details of damage at a per-building basis can now be captured. In section 4, we identify relevant features that are intuitive and useful for classification into qualitative damage states. We evaluate and discuss the results from each of these individual steps in section 5. Finally, section 6 concludes the work presented in this paper.

2 IMAGE REGISTRATION

The satellite or aerial images from before and after a disaster may have been taken by different cameras at different altitudes, angles and positions. While image providers may provide registered or georeferenced images, often this is not the case. As a result, images of the same location may vary in overlap and viewpoint. By registering two images, such differences can

be corrected. For registering before and after damage images, we extract low-level features from both images and find a one-to-one matching.

2.1 *Feature extraction*

For feature extraction we adopt a scale- and rotation-invariant interest point detector and descriptor, coined SURF (Speeded Up Robust Feature) first introduced in (Bay et al, 2006). The choice of SURF features for this application is inspired by the fact that they are more robust and faster than other state-of-the-art detectors and descriptors while maintaining high accuracy in matching. The interest points thus detected are described by the distributions of the intensity content within their neighborhood using Haar wavelet responses in the vertical and horizontal directions. This is represented as a descriptor vector. For more details the reader is advised to refer to (Bay et al, 2008). See Figure 1 for an example of interest point detection.

Figure 1. Before (left) and after storm (right) images are shown with detected interest points marked in red circles.

2.2 *Matching and Transformation*

Once the interest points are detected, we need to find a one-to-one correspondence between the detected points. This task is complicated by the fact that there may be outliers. First, a k-Nearest Neighbor algorithm (Muja et al, 2009) is used to find the first and second nearest neighbors in terms of descriptor vectors. A Euclidean distance between the descriptor vectors is used. A matching pair is accepted if its distance is closer than the threshold (t = 0.7) times the distance of the second nearest neighbor. Once the matching is done, we find the transformation matrix. Satellite images can be treated as 2D projections being viewed through a camera viewfinder. Because the camera's position, orientation, and field of view may change for before-and-after disaster images, we assume that a homographic transformation on the matched points is sufficient to correct the changes. Given a homographic transformation matrix H with elements h_{ij} we determine the value of the matrix by minimizing the back projection error:

$$min \sum_i (x'_i - (h_{11} * x_i + h_{12} * y_i + h_{13}) \div (h_{31} * x_i + h_{32} * y_i + h_{33}))^2 + (y'_i - (h_{21} * x_i + h_{22} * y_i + h_{23}) \div (h_{31} * x_i + h_{32} * y_i + h_{33}))^2 .$$

This is done by using a RANSAC algorithm (Zhang Z et al, 1999) to try many different random subsets of 4 matched pairs each. To further ensure than the best subset of 4 matched pairs is used, we reject any subset where collinearity and cyclic order of the points are violated. For example, if 4 points in the before image are collinear, we expect them to be collin-

ear in the after image as well. It then estimates the homography matrix using each subset. The quality/goodness of each subset is computed based on the total number of inliers. The best subset is then used to produce the initial estimate of the homography matrix. The computed homography matrix is refined further (using inliers only) with the Levenberg-Marquardt method in order to reduce the re-projection error further. This homographic transformation can then be applied to one of the images to correct it with respect to the other (See Figure 2).

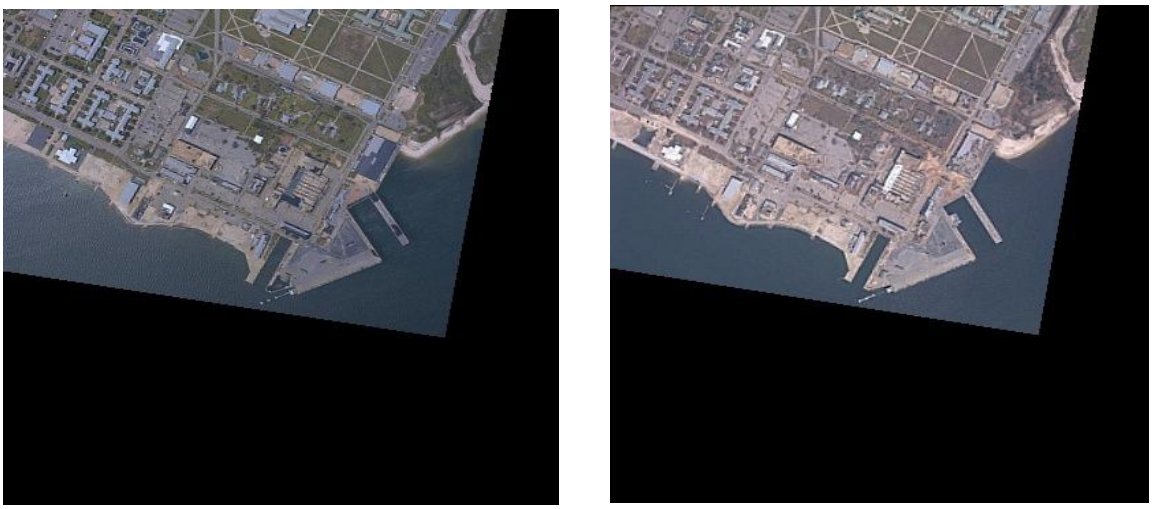

Figure 2. Before (left) and after storm (right) images are shown after homographic transformation.

3 BUILDING DETECTION

Many image repositories do not contain metadata, therefore we require automatic building detection from post-disaster images. However most repositories also lack stereo or multi-sensor information. Hence, the case addressed in this section is a challenging monocular object recognition task based on purely optical data. The common property of the previous techniques (Noronha and Nevatia, 2001; Sirmacek and Unsalan, 2009; Song et al, 2006) is that they are based on one or more specific assumptions (like presence of unique roof colors, shadows and shadow filters, strong edges, homogeneous roofs, only a few typical building structures, or simple 3-D models can be fit), but they fail if the features used are missing or less discriminative for the input data. Due to the several restrictions of edge-based techniques, we adopt a region-based scheme.

In the proposed region-based scheme, the first stage is segmentation where we group together pixels of similar spectral content. We assume an N dimensional multispectral space. While the images used in this evaluation have N = 3 (R, G and B channels), the use of additional bands can improve the segmentation performance. We use maximum likelihood classification to predict the spectral class of each pixel. Sufficient training pixels for each spectral class must be available to allow reasonable estimates to be obtained. For each image, regions belonging to 11 classes were manually labeled. For each class, 4 regions with at least 100 pixels each were used for training. Classes included ocean, grass, vegetation, road, white roof, blue roof, dark roof, gray roof, pavement, land and shadows. The maximum-likelihood classification results in salt and pepper noise as the class conditional mean vector and covariance matrix is calculated only from a fraction of the total pixels. To reduce the noise, classification, over-segmentation and under-segmentation, we require more training pixels. However, to minimize supervision in the detection process, we instead use a k-means clustering algorithm to improve the results of the maximum likelihood classifier. Using the means of the labeled training classes as seeds, a clustering is performed over pixels of the entire image. The means and covariance of the resultant clusters are then used with the maximum-likeli-

hood classifier and shadows. The resultant segmentation can now be used to identify buildings.

Buildings are considered to be objects composed of one or more building segments, and which fulfill certain properties such as presence of shadows in a certain direction, compactness, meeting area threshold etc. However, in practice due to variations in rooftop texture buildings are usually over-segmented. A highly robust way to overcome this problem is to consider buildings as objects consisting of at most 'r' segment classes ,which may include non-building classes as well (r = 4 in our experiments). Each segment formed from each combination of r classes can then be tested with threshold values for shape and shadow based measures to make a building/non-building decision. The shape based measures include area (defined as number of pixels within a segment) and compactness. Compactness is defined as:

$$Compactness = 4\pi \frac{area}{perimeter^2}$$

and has a maximum value of 1 for a circle,. We propose a new measure, the measure of estimated and predicted shadow (MEPS), to describe the 'shadow strength' of an object. The first step in computing the MEPS feature is finding an approximate estimation of the direction in which we expect the shadow to be present. This shadow direction can be denoted by a direction vector $\vec{v_{sh}}$. If the time at which the image was acquired is available then computing $\vec{v_{sh}}$ is trivial. Otherwise segments that were classified as buildings can be used to estimate $\vec{v_{sh}}$. Due to noise and misclassifications there maybe disagreements between building segments on the direction of $\vec{v_{sh}}$. This can be overcome by a voting process. In the second step, based on $\vec{v_{sh}}$ we predict the pixels that could be marked as shadow along the perimeter of any object. The MEPS feature is then computed based on the number of predicted pixels. We can define MEPS feature $MEPS_x$ for a contour *x* as $MEPS_x = N_{sh} \div N_x$

where N_{sh} is the number of points on the contour where at least a shadow pixel was found along the profile and N_x is the number of points where we predicted shadow to be present. $MEPS_x$ takes up a value between 0 and 1 with higher values for for elevated surfaces like rooftops and lower values for flat areas like roads, pavements and parking lots. Finally, as described earlier, each segment formed from each combination of 'r' classes can then be tested with threshold values for compactness C_{th}, area A_{th} and MEPS value $MEPS_{th}$. In our experiments we ignored shadow and grass classes as their omission did not affect performance. See Figure 3 for an example of building detection.

4 DAMAGE ESTIMATION

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 (Womble, 2005). Our approach assumes that there can be three basic types of damage. Minor damage includes removal of tiles, slight irregularities in edges etc. Moderate damage includes holes in the roof, dislodged decking and partial change in elevation of the roof. The third kind is when the rooftop collapses or is missing. The change measures extracted from the images should be able to discriminate well between these three general categories. Change measures are not calculated per building. Instead, we divide each building into a grid of cells, extract the features and calculate the damage measure for each cell. This allows for determining different levels of damage for different parts of a building.

4.1 *Damage-based features*

To characterize the 3 types of damage (collapsed roof, cavities and removed tiles) we use 3 different features. When a roof collapses, the image shows an increase in the number of new edges and non-linearity of existing edges. Canny edge detection performed on the difference

of before-and-after images can capture this change in appearance of edges. We propose edge density as a measure of number of pixels that appear as new or changed edges per unit area.

Figure 3. An image from coastal florida (left). Extracted buildings are shown in white (right).

Conversion into HSV color model helps identify features that intuitively have more discriminating capacity. The V (value) component can be used to discriminate roofs with cracks, openings or holes. This happens when the roof structure is partially or completely destroyed, leaving a huge cavity. A Bhattacharya distance of V histograms of before and after cells respectively is computed for each cell.

Lastly, the variation in color that occurs in the case of milder damage like removed tiles, exposed decking etc is noticeable in the hue spectrum. Hence we compute another measure, the hue means, to capture the change in hue spectrum. We define Hue means as the distance between means of hue components for each cell in before and after images.

4.2 *Damage classification*

Now we have three features for each cell of a grid; the edge density, V histogram and H means (see Figure 4). A grid is placed over each building such that it is a rectangle of minimum area that entirely circumscribes the building. The grid is used with the intention of finding damages that are partial and localized, such as a partial collapse, removed tiles etc. For each feature, the values are discretized into the k bins. For each building we count the number of grid cells that belong to each of the k bins. These numbers are concatenated for the 3 features, to get a $k \times 3$ feature vector.

This process captures localization; i.e., small areas of damage, but not necessarily the spatial proximity of damage. This property can be captured by computing the proximity measures for each cell. Proximity of a cell is defined as the number of neighboring cells in the 8-connected neighbors that lie in the same bin/interval as the cell. For each of the k bins, find the totals of proximity measures for cells that belong to each bin forming another $k \times 3$ vector. Thus concatenating the proximity values with the cell counts described earlier, we have a total of $k \times 6$ features for each building.

The proposed feature vector can be used to classify buildings in an image into categories ranging from no damage to severely damaged. In our evaluation we used $k=5$ and experimented with many supervised learning algorithms for predicting the damage state of buildings as described in section 5.

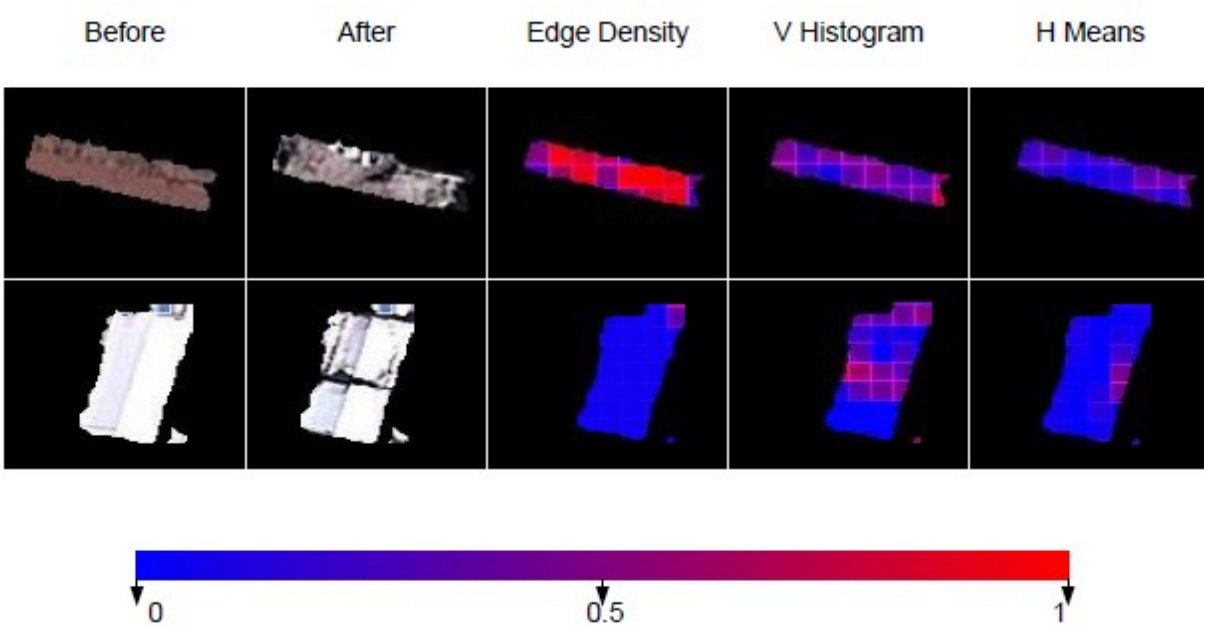

Figure 4. The first row shows a collapsed building and the corresponding false color images for edge density, V Histogram and H means. Notice that the edge density values are significantly higher in this case. The second row corresponds to a partially damaged building with a cavity in the rooftop. V histogram indicates a significant change in the cavity area and H means shows a minor damage on the roof.

5 EVALUATION

The aerial imagery used for this research was acquired by the NOAA Remote Sensing Division. The images were uncorrected and approximately 50 cm resolution. Each image is 4077x4092 pixels. The NOAA images used were Hurricane Dennis (2005) and Hurricane Ivan (2004) images of Penascola, Florida. All images were in 3-channel 24-bit JPEG format. In addition, Google Earth images of the 2010 Haiti Earthquake and QuickBird satellite images of Punta Gorda, Florida from Digital Globe were also used. The QuickBird images contained 4 spectral bands. QuickBird collects panchromatic (black and white) imagery at 60-70 cm resolution and multispectral imagery at 2.4- and 2.8-meter resolutions.

5.1 *Image Registration performance*

To study the effect of image resolution on the SURF-based registration process, we used images from 15 locations in coastal Florida before and after Hurricane Dennis. These images were 4077x4092 and 50cm resolution originally. The image pairs vary in overlap, lighting conditions and viewpoint. They were downsampled to smaller versions with 1m, 2m, 4m and 5m resolutions. For best performance, the color to grayscale conversion was done by taking the maximum of the R, G and B values for each pixel. While 14 of the 15 images registered successfully at 50cm, 1m and 2m, the performance dropped to 12 out of 15 at coarser resolutions. The registration error for all registered images was found to be 1 pixel. The registration performed well even with low overlap and varying lighting conditions. While experiments show that repeatability of the SURF detector was adversely affected at lower resolutions, the proposed scheme is well suited for high-resolution satellite imagery.

5.2 *Building Detection*

Building detection experiments were performed on 3 images of coastal Florida before Hurricane Dennis. Let TP be the true positives found by the detection and FP be the false posit-

ives. Let P or actual positives be the number of buildings in the Ground Truth. We define true positive rate (TPR) as TP/P and positive predictive value (PPV, analogous to precision) as TP/(TP+FP). Table 1 show the results of the detection after comparison with the ground truth. A high TPR ranging between 85% to 91% was observed for all images. However the PPV values were lower, with image 3 having the lowest value of 76%. This is due to higher FP with lower threshold values of Cth = 0.3 and Mth = 0.3. The longest running time observed was 25 minutes (for image 2).

Table 1. Performance of Building Detection

Image	Size	Buildings	DP	Accuracy	Time (mins)
1	785 × 719	50	86%	88%	5
2	2056 × 1320	250	88%	79%	25
3	659 × 496	35	91%	76%	2

5.3 *Damage Classification performance*

To do a thorough assessment of the classification process, we first prepared the ground truth by visually inspecting the damage states and making a manual judgment of the appropriate damage state. The labeling is done into one of the various classes based on the damage metric scale used. A fine-grained 4-scale damage metric ranging from RS-A (no damage) to RS-D (total collapse) similar to the one used in (Womble et al, 2005) is used. A simpler 3-scale metric formed by combining classes B and C was also used. A total of 150 buildings have been labeled into both the scales. We then used supervised classification techniques to classify each building into a damage state by using the damage features mentioned in section 4.

For the experiments, 75 buildings were used in the training set and another 75 in the testing set. We experimented with decision trees, rule learners, perceptron based techniques, Bayesian Networks, and ensembles of classifiers and report only the top two performers here. An ensemble of J48 trees with AdaBoosting (Freund et al, 1996) performed the best for the 4-scale damage metric while RandomForest (Breiman et al, 1996) performed better for the 3-scale damage metric. The best corresponding accuracies were 72% and 80% respectively. While studying the performance of RandomForests with varying number of trees, it was observed that for 3-scale metric we find a consistent 80% accuracy when the number of trees is between 6 and 10. Similarly, while studying the performance of AdaBoosting with varying number of iterations, it was observed that the peak of accuracy for 4-scale is achieved at 3 iterations and for 3-scale at 7 iterations. The confusion matrices in Table 2 show the classifier performance for individual classes. In spite of the effects of lighting, shadows etc., both classifiers perform well with no-damage and total-damage classes (A and C for 3-scale, A

Table 2. A confusion matrix for 3-scale metric with RandomForest, number of trees = 7 (left) and A confusion Matrix for 4-scale metric with AdaBoost, number of iterations = 3 (right)

	Actual		
Predicted	A	B	C
A	35	3	3
B	2	13	2
C	2	3	12

	Actual			
Predicted	A	B	C	D
A	32	5	2	1
B	4	3	2	1
C	0	0	4	0
D	3	0	3	15

and D for 4-scale), however the predictions for intermediate classes (B and C) are more adversely affected in the 4-scale metric. This might either be due to lesser discriminant capability of our features for partial damages or ambiguity in the definition of these classes that may have affected the manual ground truth preparation.

6 CONCLUDING REMARKS

This paper presented an automated system for the assessment of damage from high resolution imagery. While the accuracy in previous approaches was limited by the number of control points in manual registration, our application of SURF-based feature detection was found to produce near perfect registration in 14 of the 15 image pairs at 50cm-2m resolution. We proposed a novel segmentation-based building detection algorithm. Our algorithm was able to accurately extract the boundary contours of buildings in a reasonable amount of time and gave a TPR (True Positive Rate) in the range of 85%-90%. We proposed change detection measures that reflect the kind of damage that occur after a windstorm. We used a combination of edge-based and color-based measures to classify damage into qualitative states. The final results in classification were promising; 80% accuracy for a 3-scale damage metric and 72% accuracy for a finer-grained 4-scale damage metric. While the proposed system attempts at automating every aspect of damage assessment, there are several areas that require manual supervision. Further, there are areas that require improvement in terms of computational time and robustness. All these provide new challenges to be addressed in future work.

7 ACKNOWLEDGEMENTS

The support for this study was provided in part by the Global Center of Excellence at TPU funded by the MEXT, Japan.

8 REFERENCES

Bay H., Tuytelaars T., Gool L. V., 2006. Surf: Speeded up robust features. In ECCV, pages 404–417.

Bay H., Tuytelaars T., Gool L. V., 2008. Speeded-up robust features (surf). Computer Vision and Image Understanding, pages 346–359.

Bitelli G., Camassi R., Gusella L., Mognol A, 2004. Image change detection on urban areas: The earthquake case. in: Proceedings of the ISPRS XXth Congress Volume XXXV, pages 6–14.

Breiman L.,1996. Bagging predictors. Machine Learning, pages 123–140.

Chen Z., Hutchinson T. C., 2005. Urban damage estimation using statistical processing of satellite images: 2003 bam, iran earthquake. Color Imaging X: Processing, Hardcopy, and Applications, pages 289–300.

Freund Y., Schapire R. E., 1996. Experiments with a new boosting algorithm. In Proceedings of the Thirteenth International Conference on Machine Learning, pages 148–156.

Matsuoka M., Vu T. T., Yamazaki F., 2004. Automated damage detection and visualization of the 2003 bam, iran earthquake using high-resolution satellite images. In 25th Asian Conference on Remote Sensing, Thailand, pages 841–845.

Muja M., Lowe D., 2009. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration, in International Conference on Computer Vision Theory and Applications.

Noronha S., Nevatia R., 2001. Detection and modeling of buildings from multiple aerial images. In IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 501–518.

Sirmacek B., Unsalan C., 2009. Urban-area and building detection using sift keypoints and graph theory. In IEEE Transactions on Geoscience and Remote Sensing, pages 1156–1167.

Song Z. Y., Pan C. H., Q. Yang., 2006. A region-based approach to building detection in densely build-up high resolution satellite image. In Intl Conference on Image Processing, pages 3225–3228.

Womble J. A., 2005. Remote-sensing applications to windstorm damage assessment. Doctoral Dissertation in Civil Engineering, Texas Tech University, pages 1–100.

Womble J. A., Mehta K. C., Adams B. J., 2007. Remote sensing assessment of wind damage. in: Proceedings of the 5th international Workshop on Remote Sensing Applications to Natural Hazards.

Zhang Z., 1999. Flexible camera calibration by viewing a plane from unknown orientations. In IEEE International Conference on Computer Vision, volume 1, pages 666–678.