

Damage Pattern Mining in Hurricane Image Databases*

Shu-Ching Chen¹, Mei-Ling Shyu², Chengcui Zhang¹, Walter Z. Tang³, Keqi Zhang⁴
¹School of Computer Science, Florida International University, Miami, FL 33199, USA

²Department of Electrical & Computer Engineering,
University of Miami, Coral Gables, FL 33124, USA

³Department of Civil and Environment Engineering,
Florida International University, Miami, FL 33199, USA

⁴International Hurricane Center,
Florida International University, Miami, FL 33199, USA

{chens@cs.fiu.edu, shyu@miami.edu, czhang02@cs.fiu.edu, tang@eng.fiu.edu, zhangk@fiu.edu}

Abstract – We present a damage pattern mining framework for hurricane data on residential houses using aerial photographs with 1:3000 based scale. The vertical photographs are normally collected for an overview of a disaster area or for more detailed assessment. Damage on roof, especially for shingles torn off, is expected to be discovered in a more efficient and automatic way instead of going through the high-resolution aerial photograph for numerous details. The discovered damage patterns can then be used for hurricane damage assessment and impacts on different geographical areas. Our methodology is: (1) applying a novel and effective segmentation method on each residential house on the aerial photograph of one community, (2) using the segmentation results to obtain a set of indexing parameters for each house representing the damage level of roof cladding as well as the patterns of damage, (3) using these parameters to select several templates representing the damage patterns so that users can issue query-by-example (QBE) queries. The proposed segmentation method is an unsupervised simultaneous partition and class parameter estimation algorithm that considers the problem of segmentation as a joint estimation of the partition and class parameter variables. By utilizing this segmentation method, the indexing parameters can be obtained automatically. The QBE capability can assist in finding similar damage patterns on the roof of the residential houses in different locations in the image databases. Experiments based on the aerial photographs of Hurricane Andrew in 1992 are conducted and analyzed to show the effectiveness of the proposed Hurricane damage pattern mining framework.

Keywords: Multimedia data mining, Hurricane Andrew, damage assessment, segmentation, QBE (Query

By Image).

1 Introduction

After Hurricane Andrew in 1992, many areas of south Florida experienced damages with different degrees. Damage assessment and evaluation become very imperative after the disaster. Although there are a lot of innovations in the field of remote sensing, aerial photography is still the most widely used means for studying surface features. Vertical photographs are normally collected with enough overlap for stereo viewing and are suitable for an overview of a disaster area or for more detailed assessment. In this paper, the aerial photographs used are 1:3000 based scales.

There are some methods for hurricane damage assessment such as the quantitative damage assessment of homes and the study by FEMA [7]. One way is to gather quantitative damage assessment data by interpretation of aerial photographs and provide quicker assessment of damage. The methodology of interpreting aerial photographs can be based on the use of magnifying glasses or computer image scanning for greater magnification so that the photo interpreter could accurately assign a value of damage to each roof. In other words, this step is totally done by hand. Also, the construction of its damage database of residential houses is all completed by hand. It is not an efficient way especially for such a large amount of image data. The damaged area on the roof, especially for shingles torn off, is a very important criteria for evaluating overall damage, and it is expected to be estimated in a more efficient and automatic way instead of going through the high-resolution aerial photograph for numerous details. For this purpose, knowledge discovery in databases (KDD) or data mining techniques can be used. As pointed out by [10], there is a need and an opportunity for at least a partially-automated

*© 2003 IEEE.

form of KDD or data mining to handle the huge size of real-world database systems. Data mining or knowledge discovery is defined as a process of extracting implicit, previously unknown, and potentially useful information from data [1], [5], [6], [9], [12]-[15]. The objective of this process is to sort through large quantities of data and discover new information [8], [11].

In this paper, we propose a damage pattern mining framework for hurricane data on residential houses. The main idea of our proposed framework is to apply an efficient image segmentation method to distinguish the damaged area and the undamaged area within the roof. By doing so, we can easily and automatically discover the relative amount of roof cover missing instead of obtaining assessment by photo interpreters. Furthermore, since many other useful parameters describing the damaged/undamaged area can also be obtained during the procedure of segmentation, we can use these parameters to represent the damage patterns of residential houses. The steps of our methodology are:

1. Applying a novel and effective segmentation method on each residential house on the aerial photograph of one community.
2. Using the segmentation results to obtain a set of indexing parameters for each house representing the damage levels of roof cladding as well as the patterns of damage.
3. Using these parameters to select several templates representing the damage patterns so that users can issue query-by-example (QBE) queries.

The essential part of our proposed framework is a segmentation method called an unsupervised simultaneous partition and class parameter estimation algorithm, that considers the problem of segmentation as a joint estimation of the partition and class parameter variables. In this paper, we improve its performance and apply the enhanced segmentation method to the residential houses on the aerial photographs. By utilizing this segmentation method, the indexing parameters can be discovered automatically. The house images as well as their segmentation parameters constitute the damage database of residential houses. Based on the construction of the damage database, our framework can answer quick roof damage level queries and effective roof damage pattern queries. Also, we use a multi-filter strategy to filter most of the bias house images that are far different from the example query image at the very beginning of query, which can greatly reduce the overhead. The test results based on the aerial photographs of Hurricane Andrew in 1992 are conducted and analyzed to show the effectiveness of the proposed damage pattern mining framework. Figure 1 gives the workflow of the proposed framework.

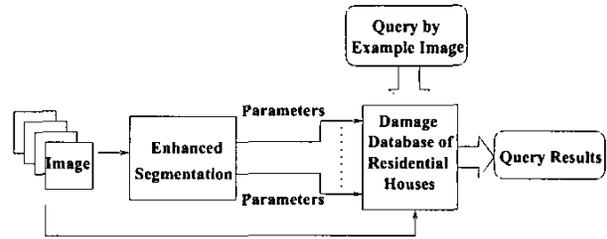


Figure 1: The workflow of the proposed damage pattern mining framework and query-by-example of hurricane data on residential houses.

The paper is organized as follows. In Section 2, the proposed damage pattern mining framework is presented. The enhanced segmentation method, the parameters obtained by segmentation, and the multi-filter strategy for image queries are included in this section. Section 3 gives query results and discussions. Conclusions are presented in Section 4.

2 Hurricane Damage Pattern Mining Framework

The proposed framework consists of two main parts: the knowledge discovery part and query mechanism part. An enhanced unsupervised segmentation method is proposed for the knowledge discovery part. The following two subsections discuss these two parts respectively.

2.1 Segmentation Method

In this subsection, an overview of the original unsupervised segmentation method is given. Then, an enhanced segmentation method used in our framework is introduced.

2.1.1 Unsupervised Segmentation Method

As mentioned previously, the segmentation method is an unsupervised image segmentation method to partition images. In this algorithm, the partition and the class parameters are treated as random variables. The method for partitioning a still image starts with a random partition and employs an iterative algorithm to estimate the partition and the class parameters jointly [2, 3, 4]. A randomly generated initial partition is used for each still image.

Suppose we have two classes *class1* and *class2*. Let the partition variable be $c = \{c_1, c_2\}$, and the classes be parameterized by $\theta = \{\theta_1, \theta_2\}$. Also, suppose all the pixel values y_{ij} (in the image data Y) belonging to class k ($k = 1, 2$) are put into a vector Y_k . Each row of the matrix Φ is given by $(1, i, j, ij)$ and a_k is the vector of parameters $(a_{k0}, \dots, a_{k3})^T$.

$$y_{ij} = a_{k0} + a_{k1}i + a_{k2}j + a_{k3}ij, y_{ij} \in c_k \quad (1)$$

$$Y_k = \Phi a_k \quad (2)$$

$$\hat{a}_k = (\Phi^T \Phi)^{-1} \Phi^T Y_k a_k \quad (3)$$

We estimate the best partition as that which maximizes the a posteriori probability (MAP) of the partition variable given the image data Y . Now, the MAP estimates of $c = \{c_1, c_2\}$ and $\theta = \{\theta_1, \theta_2\}$ are given by

$$\begin{aligned} (\hat{c}, \hat{\theta}) &= \underset{(c, \theta)}{\text{Arg max}} P(c, \theta | Y) \\ &= \underset{(c, \theta)}{\text{Arg max}} P(Y | c, \theta) P(c, \theta) \end{aligned} \quad (4)$$

Let $J(c, \theta)$ be the functional to be minimized. With appropriate assumptions, this joint estimation can be simplified to the following form:

$$(\hat{c}, \hat{\theta}) = \underset{(c, \theta)}{\text{Arg min}} J(c_1, c_2, \theta_1, \theta_2) \quad (5)$$

$$\begin{aligned} J(c_1, c_2, \theta_1, \theta_2) &= \sum_{y_{ij} \in c_1} -\ln p_1(y_{ij}; \theta_1) \\ &+ \sum_{y_{ij} \in c_2} -\ln p_2(y_{ij}; \theta_2) \end{aligned} \quad (6)$$

The algorithm starts with an arbitrary partition of the data and computes the corresponding class parameters. Using these class parameters and the data, a new partition is estimated. Both the partition and the class parameters are iteratively refined until there is no further change in them. After the segmentation, a set of parameters describing both of the two classes is obtained automatically, and some of these parameters are selected for future query use. The details about the parameters are put into Section 2.2. Figure 2 gives one example of segmentation result for residential house image house15. The left image is the original image while the right one is the segmentation result of the original one. Corresponding to our application domain of roof damage assessment, the gray area in the segmentation result represents the damaged area (such as shingles torn off) within the roof while the black area represents the undamaged area. It is very natural to think about using the segmentation results to query the relative amount of roof damage area (i.e. the area of cladding missing divided by the area of the whole roof resulted in a percentage value). In fact, our framework does support this kind of query, and the results are shown in Section 3.

2.1.2 An Enhancement to the Segmentation Method

As mentioned above, the original segmentation method begins with randomly generated initial partitions. By beginning with different initial partitions, we

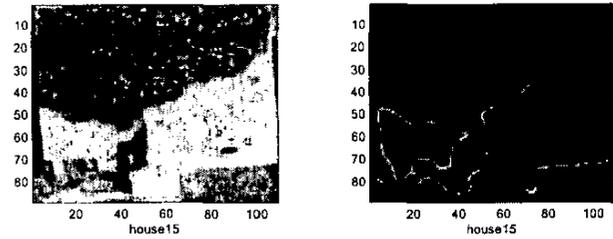


Figure 2: An example segmentation for residential house image house15.

get different local minima. The best among them, i.e. one with the least value gives the desired solution. Of course, there is no guarantee that it is the global minimum. In our method, we compute a number of (usually < 20) local minima in order to give the desired solution. Since the computational requirement for each local minimum is very little, the overall computation needed for the best local minimum is not much by doing so.

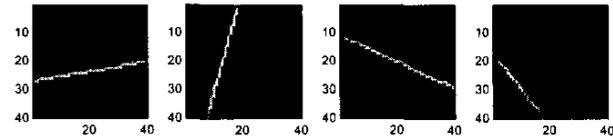


Figure 3: Fours examples of randomly generated initial partitions for the proposed segmentation method.

When generating the twenty initial partition candidates, we combine two methods: the randomly generated straight-line partitions and the predefined templates. Figure 3 gives four examples of the randomly generated straight-line partitions (the number of classes is two). The area of the original house images is partitioned by an arbitrarily generated straight-line across the whole image area. Different areas separated by the straight-line represent different classes. In many cases, the randomly generated straight-line partitions are good enough to get the desired initial partition, but in many other cases it cannot work well because of the limitations of the straight-line partition. In order to get a good initial partition as quick as possible, we combine the technique of predefined templates into the generation of initial partitions. As shown on Figure 4, eight predefined templates are selected as candidates in the selection of the desired initial partition. Another important issue about the initial partition is how to select the “best” one among those candidates. The criteria for evaluating the candidates involve two aspects: one is the local minimum, and the other is the standard deviation of each class within a house image. We choose two candidates when each of them has either the lowest local minimum or the lowest standard deviation. Then compute the two candidates’ global minimum. The one

with the lower global minimum should be chosen as the final partition. Our experiments on hurricane house images show that, by doing so, greater than ninety percent of the residential house images can be partitioned very well, which means the accuracy of our enhanced segmentation method is greater than ninety percent.

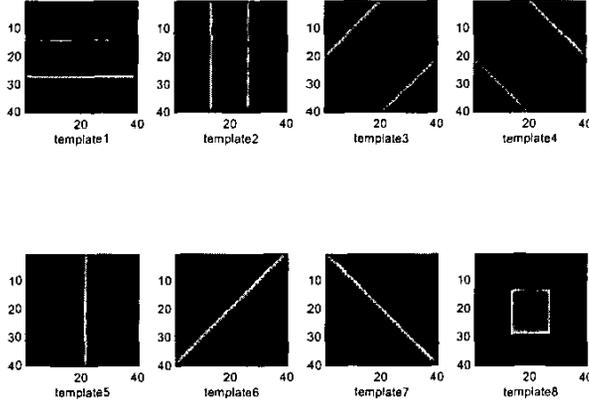


Figure 4: The eight predefined initial partition templates.

2.2 Query Strategy

In order to explain the query strategy in our proposed framework, we should first look through the parameters obtained through segmentation procedure. After the segmentation on each residential house image, a set of parameters for each image is obtained automatically. The parameters can be used in discovering the damage patterns in the hurricane image database by the proposed query strategy. Compared with the high-dimension of the original image, the number of parameters for each image is much smaller. Some of these parameters are selected for query use. Since our segmentation method uses the functions of the spatial coordinates of the pixel as the mathematical description of a class, those parameters relative to spatial information should be able to represent the spatial damage pattern of the residential houses.

Parameter MEAN: After the segmentation, each pixel within a house image has its class identification. That is, if there are two classes, then the class identification for each pixel is either 1 or 2. As mentioned above, the relative position of a pixel within a house image is denoted by its spatial coordinates. Suppose the number of pixels belonging to *class1* is $Nk1$ while $Nk2$ denotes the number of pixels in *class2*. Parameter *MEAN* is a 2×2 matrix and is given by $(Mean_c1_cord, Mean_c2_cord)$, where *Mean.c1.cord* and *Mean.c2.cord* are vectors denoted by:

$$Mean_c1_cord = (Mean_c1_x, Mean_c1_y)^T \quad (7)$$

$$Mean_c2_cord = (Mean_c2_x, Mean_c2_y)^T \quad (8)$$

$$Mean_c1_x = \left(\sum_{y_{ij} \in class1} i/Nk1 \right) \quad (9)$$

$$Mean_c1_y = \left(\sum_{y_{ij} \in class1} j/Nk1 \right) \quad (10)$$

$$Mean_c2_x = \left(\sum_{y_{ij} \in class2} i/Nk2 \right) \quad (11)$$

$$Mean_c2_y = \left(\sum_{y_{ij} \in class2} j/Nk2 \right) \quad (12)$$

Parameter CV: It is the covariance matrix of matrix *cvecs.n* ($n = 1$ or 2).

$$cvecs_n = (stkr_n, stkc_n)^T - mn_n \times ones_Nkn \quad (13)$$

where *stkr.n* is a column vector with each row being the *i.coordinate* of $y_{ij} \in class_n$, and *stkc.n* denotes the column vector of *j.coordinates* of $y_{ij} \in class_n$. Here, *mn.n* is a column vector with 2 elements, each of them representing the mean of the *i.coordinates* or *j.coordinates* $y_{ij} \in class_n$. *ones.Nkn* is a unity column vector of Nkn elements (i.e., all of them have value 1).

$$CV = (CV_1, CV_2) \quad (14)$$

$$CV_n = (cvecs_n \times cvecs_n^T)/Nkn \quad (15)$$

The parameter *CV* represents the spatial distribution pattern of each class, which is very essential for damage pattern query.

Parameter DET: It contains the determinants of matrix *CV.n*, where n is 1 or 2.

$$DET = Determinant(CV_1, CV_2) \quad (16)$$

The reason why we need the parameter *DET* is that *CV* has eight (2×4) elements but *DET* has only two elements. Remember that even the parameter *MEAN* has four elements. Since we use Euclidean distance for comparing two matrices, the less the dimension of matrix is, the better the performance is. Of course the information included in *DET* may be not enough to achieve good query results. However, if we use it as the first level filter in the query framework, the overall computation can be reduced such that the framework can answer queries quicker and more efficient. In fact, we apply a multi-filter architecture in our framework. Figure 5 shows how the proposed framework works. The performance of this architecture will be discussed in Section 3.

Another issue about ranking the retrieved images is relatively simple. It just uses the sum of the weighted Euclidean distances on parameters *CV* and *MEAN* between the query image and the retrieved image to determine the ranking. The values of the weights are derived from the experiential values.

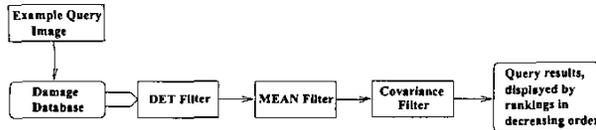


Figure 5: The multi-filter query architecture.



Figure 6: An example piece of aerial photographs.

3 Query Results and Discussion

The example residential house images are collected from the grayscale aerial photographs of southwest Miami area after Hurricane Andrew. Figure 6 shows one piece of such aerial photographs. Total 550 house images are cropped from the original aerial photographs by hand and saved as individual images for query use. The size of each image is no less than 200 rows and 200 columns. Most of the houses experienced different levels of damages. Two typical examples of the assessed houses are shown below to demonstrate how the damage pattern queries are executed. As we mentioned before, the white blemishes or spots evident in the house images represent the torn off shingles or damaged sheathing, and the black areas are those that are almost undamaged. We apply our segmentation method to all of these images, using two classes to represent the damaged area and undamaged area respectively within a house roof area (the damaged area is represented by gray areas in the segmentation result, while the undamaged area is represented by black areas). The house images as well as their segmentation parameters constitute the damage database of the residential houses. We will use two kinds of queries to illustrate the performance of our proposed framework.

3.1 Query for Roof Cover Missing Level

An example roof cover missing level query is “Show me more images which has similar roof cover missing level like this.” The parameter we use for query was determined by the relative amount of cover missing (i.e., the area of cladding missing divided by the area of the whole roof results in a percentage value). This parameter is automatically obtained after the procedure of segmentation.

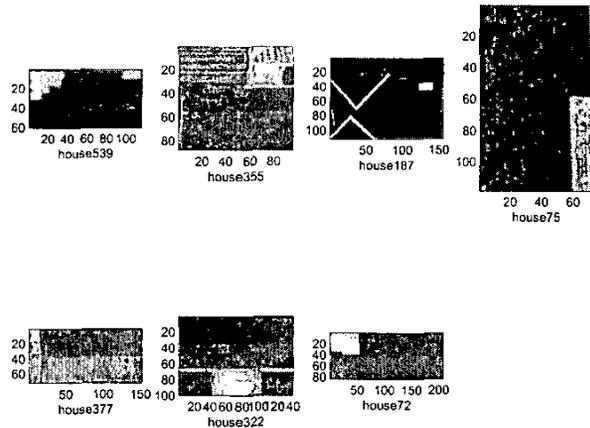


Figure 7: Roof cover missing query results on the hurricane damage database of the residential houses. Query image house539 is on the top left and the matches of the image are from top left to bottom right in decreasing order of their similarity.

Figures 7 and 8 are given to show the query results for query image house539. Figure 7 shows the first 7 original house images being retrieved. The top left image in Figure 7 is the query image (house539) and the matches of the images are from top left to bottom right in decreasing order of similarity. Figure 8 shows the segmentation results of those house images. In this case, the roof cover missing level of image house539 is about 15 percent. Since roof cover missing level is an important criteria for evaluating the overall damage, this kind of query is very useful and convenient for quantitative damage assessment of the residential houses.

3.2 Damage Pattern Query

In our framework, the damage pattern query that we use is a similarity query, for example, “Show me more house images which have similar damage patterns like this.”

Figures 9 and 10 are given to illustrate the query results for query image house433. Figure 9 shows the first 5 original house images being retrieved. Query image house433 is on the top left, and the matches are from top left to bottom right in decreasing order of their similarity. Figure 10 shows the segmentation results of those

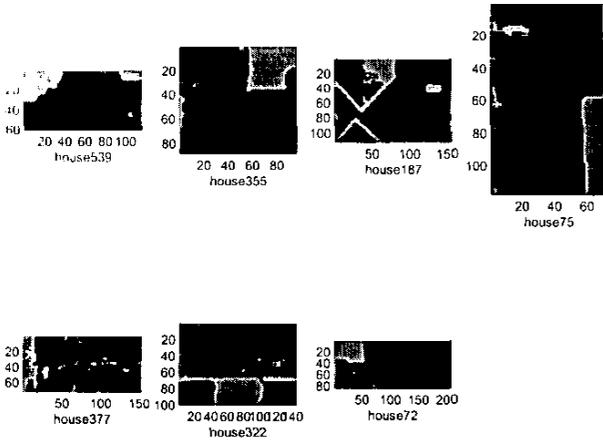


Figure 8: Roof cover missing query results on the hurricane damage database of the residential houses after the segmentation. Again, query image house539 is on the top left and the matches of the image are from top left to bottom right in decreasing order of their similarity.

house images. From the segmentation results, we can see that the damage pattern of image house433 is of relatively regular shape (rectangular), and the damaged area is distributed almost evenly on the left and right sides of the roof. The query results looks pretty good from the observation of human eyes. The damage patterns of house431, house430 and house432 are very similar to query image house433. Even the image house246, which looks much more different with the query image than the others, has the similar damage pattern to the query image by the fact that the shape of its undamaged area is near to rectangular, and it has a large part of damaged area distributed on the left and right sides of its roof.

Figures 11 and 12 show the query results for query image house6. Figure 11 gives the first five original house images being retrieved, and Figure 12 gives the segmentation results of those images. The reason to choose image house6 as an example query image is that it has significantly different damage pattern with query image house433. Compared with house 433, the undamaged area of house6 is of irregular shape and with many indents on the top and bottom edges of it, and the damaged area is mainly distributed on the top and bottom parts of the image with the top part being a little smaller than the bottom part. From the observation of the segmentation results, we can see that the damage patterns of house20 and house545 look very similar to the query image. When we take a look at the segmentation result of house509, we can find that it has many indents on the top and bottom edges of its undamaged area, and its damaged area is mainly located on the top right part and bottom part where the former (top right part) is a little bit smaller than the latter (bottom part).

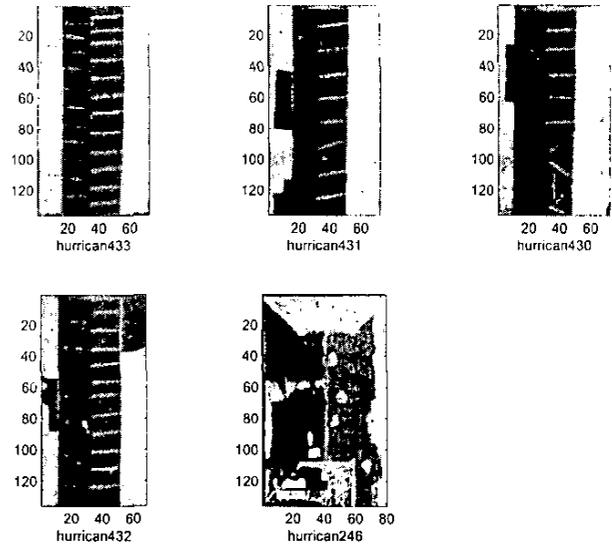


Figure 9: Damage pattern query results on the hurricane damage database of the residential houses. Query image house433 is on the top left. Matches of the images are from top left to bottom right in decreasing order of their similarity.

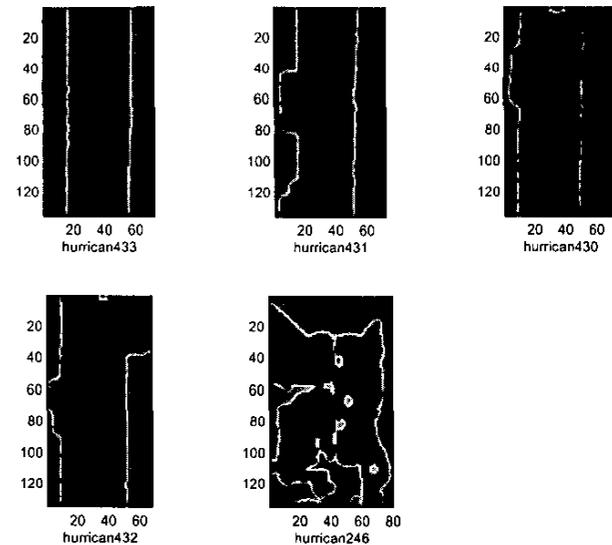


Figure 10: Damage pattern query results on the hurricane damage database of the residential houses after the segmentation. Query image house433 is on the top left. Matches of the image are from top left to bottom right in decreasing order of their similarity.

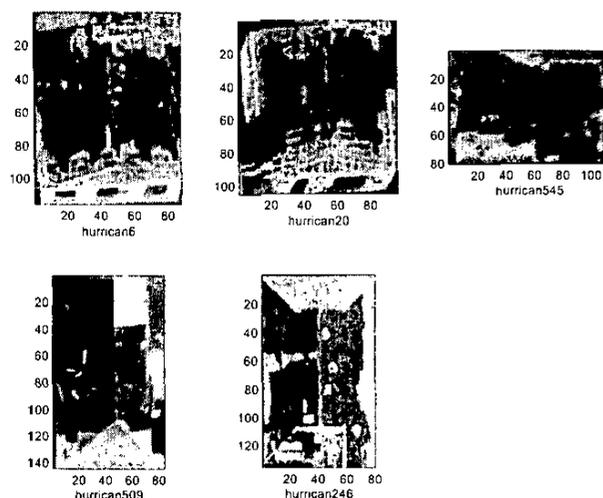


Figure 11: Damage pattern query results on the hurricane damage database of the residential houses. Query image house6 is on the top left. Matches of the images are from top left to bottom right in decreasing order of their similarity.

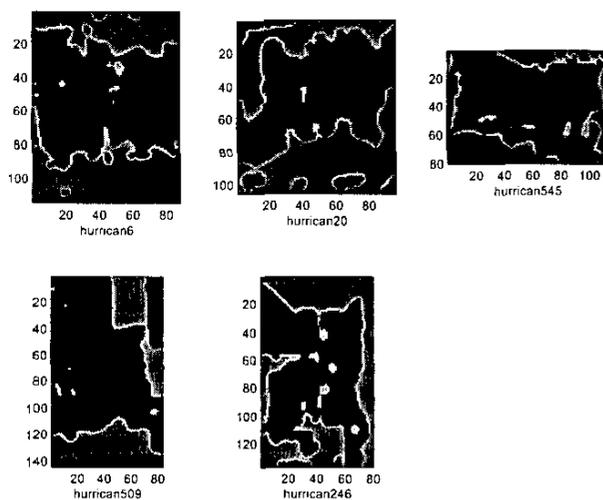


Figure 12: Damage pattern query results on the hurricane damage database of the residential houses after the segmentation. Query image house6 is on the top left. Matches of the images are from top left to bottom right in decreasing order of their similarity.

By analyzing the query results for different types of example query images, it is very promising to see that the proposed content-based retrieval framework for hurricane damage database of residential houses is able to retrieve those house images with the similar damage pattern, which is very useful for damage assessment of the residential houses. Moreover, since the proposed segmentation method is an unsupervised simultaneous partition and class parameter estimation algorithm, all the needed parameters are automatically obtained and indexed offline without any user interactions. Another good feature of this framework is, unlike many other QBE methods, it is insensitive to illumination changes since it is focused on the spatial-distribution of each area (damaged and undamaged). Also, the uses of multiple filters (*DET*, *CV* and *MEAN*) can greatly reduce the number of retrieved images at each step, which is essential for answering queries quickly. For example, when we use image house6 as a query image, the number of retrieved images sharply dropped over ninety percent after the *DET* filter and *MEAN* filter, and this is even more significant for query image house433. Since the dimension of the segmentation parameters is far more smaller than that of the images, We can believe this framework is a good foundation for further hurricane damage assessment.

4 Conclusions

In this paper, a damage pattern mining framework for hurricane image databases is presented to facilitate the establishment of predictive models for the real-time assessment of solid debris. The test result based on the aerial photographs of Hurricane Andrew in 1992 shows that the proposed framework is very effective and promising in roof cover damage evaluation and damage pattern classification. By using a novel and effective segmentation method, a set of indexing parameters for each house representing the damage levels and patterns can be discovered automatically; whereas the old damage assessment methods depend on human photo interpreters. The parameters can then be used in discovering the damage patterns in the hurricane image database by the proposed multi-filter query strategy. The multi-filter query strategy can greatly reduce the number of possible image candidates so that it can answer queries quicker. This is essential because using aerial photographs in damage assessment is valuable only if it is based on less precise and more quickly-analyzed photo interpretation. The discovered damage patterns can then be used for hurricane damage assessment and impacts on different geographical areas.

References

- [1] J. Chatratchat, J. Darlington, and M. Ghaem, "Large Scale Data Mining: Challenges and Responses", Proc. the third International Conference

- on Knowledge Discovery and Data Mining, pp. 143-146, 1997.
- [2] S.-C. Chen, S. Sista, M.-L. Shyu, and R. L. Kashyap, "Augmented Transition Networks as Video Browsing Models for Multimedia Databases and Multimedia Information Systems", Proc. 11th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'99), pp. 175-182, November 9-11, 1999.
- [3] S.-C. Chen, S. Sista, M.-L. Shyu, and R. L. Kashyap, "An Indexing and Searching Structure for Multimedia Database Systems", Proc. IS&T/SPIE conference on Storage and Retrieval for Media Databases 2000, pp. 262-270, January 23-28, 2000.
- [4] S.-C. Chen, M.-L. Shyu, and R. L. Kashyap, "Augmented Transition Network as a Semantic Model for Video Data", *International Journal of Networking and Information Systems*, Special Issue on Video Data, Vol. 3, No. 1, pp. 9-25, 2000.
- [5] U.M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery: An Overview", in U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, pp. 1-34, AAAI/MIT Press, 1996.
- [6] U.M. Fayyad, "Data Mining and Knowledge Discovery: Making Sense Out of Data", *IEEE Expert*, Vol. 11, pp. 20-25, 1996.
- [7] Federal Emergency Management Agency (FEMA). "Progress Report on a Study of Rapid Damage Assessment by Remote Sensing." 10 (December 1983).
- [8] R. Groth. *Data Mining: Building Competitive Advantage*, Prentice Hall, 2000.
- [9] R. Groth. *Data Mining: A Hands-on Approach for Business Professionals*, Prentice Hall, 1998.
- [10] G. Piatetsky-Shapiro, "Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop", *AI Magazine*, Vol. 11, No. 5, Special issue, pp. 69-70, Jan. 1991.
- [11] M.-L. Shyu, S.-C. Chen, and R. L. Kashyap, "Generalized Affinity-Based Association Rule Mining for Multimedia Database Queries", *Knowledge and Information Systems (KAIS): An International Journal*, Vol. 3, No. 3, pp. 319-337, August 2001.
- [12] E. Simoudis, "Reality Check for Data Mining", *IEEE Expert*, Vol. 11, pp. 26-33, 1996.
- [13] J. Oh and B. Bandi, "Multimedia Data Mining Framework for Raw Video Sequences", Proc. ACM Third International Workshop on Multimedia Data Mining (MDM/KDD2002), pp. 1-10, Edmonton, Alberta, Canada, July 23 -26, 2002.
- [14] M. Datcu and K. Seidel, "An Innovative Concept For Image Information Mining", Proc. ACM Third International Workshop on Multimedia Data Mining (MDM/KDD2002), pp. 11-18, Edmonton, Alberta, Canada, July 23 -26, 2002.
- [15] Z. Zhang, "Mining Surveillance Video for Independent Motion Detection", Proc. 2002 IEEE International Conference on Data Mining (ICDM'02), pp. 741-744, Maebashi City, Japan, December 09 - 12, 2002.