Smart Visualization of Urban Area Earthquake Simulation

Hideyuki O-TANI¹, Jian CHEN², Muneo HORI³

¹RIKEN Advanced Institute for Computational Science
(7-1-26, Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo, 650-0047, Japan)
E-mail:h.o-tani@riken.jp

²RIKEN Advanced Institute for Computational Science
(7-1-26, Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo, 650-0047, Japan)
E-mail:jchen@riken.jp

³Earthquake Research Institute, the University of Tokyo
(1-1-1, Yayoi, Bunkyo-ku, Tokyo, 113-0032, Japan)
E-mail:hori@eri.u-tokyo.ac.jp

Smart visualization of output from Integrated Earthquake Simulation (IES) performed for various earthquake scenarios is studied by taking advantage of existing methodology of big data visualization and by developing our own methodology which is suitable for earthquake disaster assessment. A key issue of being smart is to deliver information generated by IES to not only professional earthquake engineers, but also non-professional government officials and city residents. Simple visualization for city residents, wide area visualization for government officials, and dual visualization for professional engineers are proposed to bring common recognition to IES users. As a trial of *smart* visualization, both structure-wise and scenario-wise damage condition statistics are graphed by means of histograms and mosaic plots; spatial clustering analysis for damaged structures is introduced and the distance between the damage conditions of earthquake scenarios is defined and visualized, accompanied by scenario cluster analysis.

Key Words: *integrated earthquake simulation, earthquake disaster assessment, output visualization*

1. INTRODUCTION

2011 Tohoku Earthquake reveals the limitation of earthquake disaster assessment that is based on a unique scenario of earthquake. If an actual earthquake exceeds one which is used for the assessment, the resulting disaster naturally exceeds the assessed one. In seismic design, two scenarios are considered for ground motion, by studying numerous examples of past structure damage. Although it is a challenging task, it is worth being studied to develop earthquake disaster assessment based on various earthquake scenarios, so that two or a few suitable scenarios will be selected.

Integrated Earthquake Simulation (IES)^{(1),(2)} is a seamless simulation of the three phases of earthquake in an urban area, namely, strong ground motion generation, seismic structure response, and human/community disaster actions. In IES, an earthquake scenario is used as input, and the resulting disaster is computed objectively. Therefore, earthquake disaster assessment based on various earthquake sce-

narios can be made if all of these scenarios are used as input of IES.

Output of IES is huge, since it computes a whole target area; the data size reaches tera byte order for a case of an urban area in which 1,000,000 of buildings are located. Visualization of this output is not a simple task. Moreover, if 1,000 cases of different earthquake scenarios are simulated, visualization of such big data would be most challenging; showing the mean and variation of the disasters is not suitable even though straightforward.

In this paper, we study *smart* visualization of IES output for various cases of earthquake scenarios. The smart visualization is the first step that is critical to realize earthquake assessment based on various scenarios. We take advantage of existing methodology of big data visualization, as well as develop our own methodology which is suitable for earthquake disaster assessment. A key issue of being smart is to deliver information that is generated by IES to professional earthquake engineers and non-professional government officials.

2. SUMMARY OF IES

The program architecture of IES is a layer structure, consisting of three layers, namely, data layer, simulation layer and visualization layer. The data layer is data stored in available Geographical Information System (GIS) and data for earthquake scenario. The simulation layer executes programs for strong ground motion generation, seismic structure response, and human/community disaster actions. The visualization layer converts the simulation output to suitable files for visualization. For the inter-layer communication, used are data of pre-defined structure, called Common Modeling Data (CMD). This is a key point of IES visualization; output of the simulation layer is converted to CMD, so that the visualization layer works only for CMD.

Enhancement of high performance computing (HPC) is essential to IES, since it analyzes the entire urban area. For instance, the seismic structure response is computed for an analysis model of each building. Embarrassing parallel computation is readily applicable to this simulation if interaction among neighboring buildings is ignored, as in the case of seismic design. However, the HPC enhancement results in large amount of data; if one building model has 10^2 degree-of-freedom and there are 10^6 buildings, the output of dynamic analysis for 10^4 time steps (0.01 sec time increment and 100 sec duration) surely exceeds 10^{12} bytes, assuming that one degree-of-freedom at one step generates 8 byte data.

IES is being implemented in the K computer, with an ambition that massive parallel computation is made for assessment of earthquake and tsunami disaster in urban areas. Most advanced programming is applied to the implementation, with the help of expert computer engineers. At this moment, it is capable to analyze the whole Tokyo area, and the computation time for seismic structure analysis in the area is reduced to less than one day even if most complicated models are used.

A prototype of IES has been originally developed for a certain area of Tokyo Metropolis. Since a module for automated model construction is developed, it is readily extended to other areas. Kobe City is an appropriate target city since it suffered 1995 Kobe Earthquake. Numerous reports have been made for this urban area earthquake disaster, and data stored in them are available to validate IES simulation.

3. ANIMATION FOR ONE SCENARIO

Animation can help IES users explorer and un-

Fig.1 Visualization of a building surface model (a), and a computer model and edge of a building (b).

derstand computer simulations. It is expected to bring them common recognition of possible earthquake hazards and disasters. The common recognition is vital for rational design of disaster countermeasures.

Results of IES are expected to be understood by local government officials and city residents as well as by professional engineers. The concerns and knowledge of officials and residents are different from those of engineers, and the animations which are made for officials and residents should be adjusted by alternating the type and the amount of information or visualization area size.

In the preceding work, output of IES is animated only to exhibit the displacement time histories of buildings. The time histories obtained from an urban earthquake disaster simulation are expressed by time-varying colors of building surface models^{3),4)} as shown in **Fig.1(a)**, or animated with the movement of computer models with building edges⁵⁾ in **Fig.1(b)**. The former visualization is suited for wider area visualization where individual movements of buildings are inconspicuous due to its small scale, and the latter visualization is appropriate for observing the detailed response of buildings in a local area. However, both visualizations are intended for specific users, such as developers of IES and professional engineers, and not for non-professional government officials or city residents.

In this work, we propose several building damage visualizations of urban area simulations expected to bring common recognition of possible earthquake hazards and disasters among IES users. Since their backgrounds and knowledge are various, we discuss appropriate visualizations for respective users. These visualizations will be incorporated to a system which can smoothly convert the target earthquake scenarios to be visualized in our future work. Such system will help us to understand the differences between the earthquake scenarios.

Fig.2 A snapshot of an animation aimed at city residents.

(1) Simple visualization for city residents

There are two types of considerable visualizations for city residents with no expert knowledge: one is a simple visualization and the other is a realistic one. Both visualizations can help city residents imagine the earthquake hazards and disasters. The realistic one is effective only when the computed disaster is large and the changes or movements of buildings are prominent, while the simple one can effectively express disasters by magnifying its scale.

Here, we show an example of the simple visualization in **Fig.2**. The color of each building model represents the maximum of the damage degree determined by the story drift angles. Displacements of buildings are magnified and displayed as the movement of buildings. This visualization is designed for city residents to understand that the IES outputs are physics-based, which implicitly enhances the reliability of the output.

(2) Wide area visualization for officials

Local government officials would be interested in how many and which buildings are damaged. As an example of visualization for the officials, high angle view of buildings is shown in **Fig.3**. Each building model has a color determined according to the maximum of its damage degree; blue for no damage, white moderate damage, and red severe damage. Displacement time histories of buildings are exhibited by magnified movement of building models when visualization area is small enough to show the movement.

(3) Visualization for professional engineers

Various computer models for seismic response analysis of buildings are incorporated to IES. To visualize the results of different computer models in the same manner, IES can convert the results of different computer models to those of MDOF systems as have been shown in **Fig.1(b)**. The visualization of these converted models can provide professional engineers

Fig.3 Wide area animation aimed at government officials.

Fig.4 Visualization of computer models with spheres of radii cubic roots of floor weights.

Fig.5 Visualization of computer models with disks of radii square roots of weights.

the detailed information about seismic response of buildings.

Fig.4 and **Fig.5** show the building edges and the computer models of MDOF systems. Story weights of buildings are expressed by spheres in **Fig.4** and disks in **Fig.5**. Compared to the spheres of radii cubic roots of floor weights, the disks of radii square roots

Fig.6 Dual visualization for professional engineers.
(computer model: MDOF)

Fig.7 Dual visualization for professional engineers.
(computer model: OCM)

of weights make clear the differences among the story weights of buildings.

In addition to the computer models of MDOF systems, translucent surface models of buildings are visualized in **Fig.6**. In this dual visualization, the colors of the surface models represent the maxima of story drift angles, and those of computer models represent the time histories of story drift angles. Displacement time histories are displayed by motion of computer models. The floor weights are indicated by spheres because large disks are irritating to eyes in the dual visualization. The advantage of the dual visualization with converted computer models is that it can express the maximum damage degree and the time history of the story drift angle simultaneously together with the displacement time history.

It is also important to explicitly depict what computer models are adopted to the seismic response analysis of buildings. **Fig.7** shows an example of the dual visualization with computer models of OCM systems. We will further develop the dual visualization with different computer models in the future work.

(a) scenario A

(b) scenario B

Fig.8 Wide area visualization for two different earthquake scenarios.

4. TRIAL OF SMART VISUALIZATION

(1) Visualization of standard statistical analysis

As we mentioned, earthquake disasters computed by IES depend on an input earthquake scenario. As a typical example, **Fig.8** presents two cases of the disasters which are computed for two scenarios. The difference is large, and it may not be appropriate to use ordinary statistical analysis which computes the average and standard deviation.

To demonstrate the variability of structure damage conditions, we carried out IES simulations with the following settings: 1) a target region in Kobe city of 500 m $\times$ 500 m with 710 structures is studied; 2) one component model (OCM) is applied for structure re-

Fig.9 Histograms for the damage conditions: the mean of the 710 structures and 3 structures with different distance to the mean.

sponse analysis; 3) the JR-Takatori wave is used and its relative orientations to the structures are shifted at every 3 degrees to obtain 120 different scenarios. The damage conditions of structures are determined by the maximum drift angles of their floors and classified into three catalogues, according to assumed thresholds listed in **Table 1**.

Table 1 Classification of damage conditions

catalogue	damage condition	maximum drift angle θ_d
severe damage (S)	1.0	$\theta_d > 0.01$
moderate damage (M)	0.5	$0.005 < \theta_d \leq 0.01$
no damage (N)	0.0	$\theta_d < 0.005$

The statistics of the damage conditions of a single structure in multiple earthquake scenarios can be represented by a histogram, h_j ($j = 1, 2, 3$), which is the number of scenarios that fall into each of the damage conditions (0, 0.5, 1). The mean ($\bar{h}_j$) and standard deviation of the histograms of the 710 structures are shown in **Fig.9(a)**. The distance between the histogram of a structure to the mean histogram can be defined as $\sqrt{\sum_{j=1}^3 (h_j - \bar{h}_j)^2}$. The histograms of the structures corresponding to the 5th, 50th and 95th percentiles of the distances are shown in **Fig.9(b), (c)** and **(d)**. As is seen, the distance from the mean is quite different for each percentile. It does not seem appropriate to summarize such varying data assuming a normal distribution or another simple probabilistic distribution.

The large variation of the damage condition distribution can be more clear if a mosaic plot is used; see **Fig.10**, which corresponds to the two cases shown in

Fig.10 Mosaic plots of the statistics of structure type versus damage condition of the two scenarios in **Fig.8**: (a) scenario A and (b) scenario B. (N, M, S stand for no, moderate and serve damage.)

Fig.8. Two types of structure model are used by OCM analysis and their numbers determine the split ratio of the horizontal edges of a mosaic block. As is seen in **Fig.10**, there are more SRC (type II) structures than RC (type I) structures in the target region. The vertical edges of a mosaic block are further split according to the ratio of the damage conditions of the corresponding structure types. The two mosaic plots in **Fig.10** show that the type I structures are safe in both scenarios while the damage conditions for the type II structures vary considerably, as in scenario A almost all structures of type II are damaged to some extent and in scenario B nearly half of them are of no damage.

The mosaic plot supports a usefulness of smart visualization, as it vividly delivers the large variability of the earthquake disaster, or, more specifically, the dependence of the disaster on the input scenario. However, it does not help us consider effective countermeasures for the disasters. We thus seek to extract information which is useful for this purpose from such varying output of the earthquake disasters. To this end, again, we take advantage of smart visualization, in order to make a spatial cluster or a scenario cluster. The spatial cluster is a small area in which buildings are damaged for various scenarios, and the scenario cluster is a set of scenarios which produces more or less similar distribution of damages.

(2) Visualization of damage correlation

To visualize the correlation of two variables, we usually use scatter plot. For multiple variables, we may use trellis plot⁶⁾ and mosaic plot⁷⁾ as shown in **Fig.10**. However, the correlation obtained by these plots is not related to a spatial distribution, since these plots are made for variables which are not related to position.

For spatial clustering, we need to visualize the variables which are related to position or to visualize a

(a) Clustering result for DBSCAN input parameter set 1

(b) Clustering result for DBSCAN input parameter set 2

Fig.11 Clustering results with two different sets of DBSCAN input parameters for the same damaged structures in one scenario: $r_n = 40$ m and $M = 12$ (a); $r_n = 40$ m and $M = 15$ (b). Core points are represented by triangles and border points as circles within the dashed lines.

spatial point which has certain variables. Clustering algorithms which are developed for data mining can be adopted for the purpose of making spatial clustering. We apply the Density-Based Spatial Clustering Analysis with Noise (DBSCAN) algorithm, introduced by Ester et al.⁸⁾ DBSCAN is for a set of points with variables, with a distance function for any two of the points which is computed by using their variables, and is applicable to the damage distribution of one scenario or others. DBSCAN needs two pre-set parameters, the neighborhood radius (r_n) and the minimum number of points (M), and classifies each point into the following three types: 1) a core point, around which enough points ($\geq M$) are located within the radius r_n ; 2) a border point, which is a neighbor of any

core point (distance to the core point $\leq r_n$) but does not have enough neighbors ($< M - 1$); and 3) a noise point, which is neither a core nor a border point. A cluster is formed by a core point and its border points. If the core points of two clusters are neighbors, the two clusters are merged into a larger one. As is seen, the number and size of the clusters change depending on r_n and M , although DBSCAN makes clustering objectively for a given set of r_n and M .

As for the present case, a point is the location of the structure and the variable is the degree of the damage. We define the distance function for two structures as the Euclidian distance between the points when they have been damaged. As a trial of DBSCAN, we set $r_n = 40$ m and use $M = 12$ and 15. The resulting clusters are shown in **Fig.11**. It is hard to tell which case is more appropriate. However, we can see two common clusters located in the center for both cases. With the assistance of DBSCAN, we may choose suitable spatial clustering which is less sensitive to the pre-set parameters even for many scenarios, by filtering similar clusters which appear in various scenarios. The optimal choice of pre-set parameters for DBSCAN remains a challenging task.

(3) Smart visualization using big data tools

The data of structure damage conditions can be organized as a matrix and be visualized by treating each element of the matrix as a pixel; see **Fig.12**, where a horizontal row is structure-wise and a vertical column is scenario-wise. This visualization is useful for exploring patterns in the data; for instance, two horizontal rows which are similar to each other correspond to two structures which suffer damage in a similar manner to various scenarios. In particular, vertical columns which have similar patterns correspond to scenarios which produce similar disasters. Indeed, we can identify four groups of such columns. Such regular pattern is due to the assigned scenarios generated by sequentially changing the direction of input ground motion. Two scenarios with a difference of 180 degrees are very similar, e.g. the damages in scenarios 1 to 20 resemble those in scenarios 61 to 80.

More quantitatively, we define a distance between two scenarios in terms of structure damage conditions, as follows:

$$D(i, j) = \sqrt{(1/K) \sum_{k=1}^K (d_i(k) - d_j(k))^2}, \quad (1)$$

where $d_i(k)$ is the damage condition of the k -th structure in the i -th scenario and K is the total number of structures. Since the main interest is scenario clustering, we filter D with an upper limit of $D(i, j) = 0.2$. Note that $D(i, j)$ becomes 0.5 if all the structure have

damage 0 in the i -th scenario and 0.5 in the j -th scenario. The threshold value of 0.2 corresponds, for example, to the case when 4% of the structures change their damage conditions from 0 to 1 or the case when 16% from 0 to 0.5 or 0.5 to 1. The filtered distance matrix is visualized in **Fig.13**. We can clearly identify four blue areas along the diagonal line of **Fig.13**. These areas indicate four scenario groups, within which the distances between scenarios are relatively small (less than 0.1 with blue color). Further analysis of **Fig.13** horizontally (or vertically) shows that the four groups can be classified as two clusters, each of which contains two close groups.

5. CONCLUSIONS

Smart visualization of the IES outputs generated for various earthquake scenarios is studied by taking advantage of existing methodology of big data visualization, as well as by developing our own methodology which is suitable for earthquake disaster assessment. A key finding is that such smart visualization techniques will help us find spatial and scenario clusters of the synthesized earthquake damages, so that officials and residents as well as professional engineers can have common recognition on earthquake damage patterns. Even though there could be a millions of possibilities of earthquake, the disaster which we have to prepare might be extracted.

Further investigation is surely needed, to develop a smarter visualization technique of the big data produced by IES. It must be a separate issue from improving the reliability of IES and its results. This is because increasing the number of earthquake scenarios linearly increases the amount of the data, and it will eventually exceed the human capacity of understanding them. Spatial and scenario clustering discussed in this paper is a way to extract more essential information from the synthesized data.

REFERENCES

- 1) Ichimura, T., Hori, M., Terada, K. and Yamakawa, T.: On integrated earthquake simulator prototype, *STRUCTURAL ENGINEERING EARTHQUAKE ENGINEERING*, Vol. 22, No. 2, p. 233, 2005.
- 2) Hori, M. and Ichimura, T.: Current state of integrated earthquake simulation for earthquake hazard and disaster, *Journal of Seismology*, Vol. 12, No. 2, pp. 307–321, 2008.
- 3) Lalith, M. and Hori, M.: Application of High Performance Computation for the Prediction of Urban Area Earthquake Disaster, *SSMS Internet Journal*, 2011.
- 4) Yamashita, T., Hori, M. and Kajiware, K.: Petascale Computation for Earthquake Engineering, *Computing in Science &*

Fig.12 Visualization of damage conditions of all structures from multiple scenarios: column indices for the scenario number and row indices for structure numbers.

Fig.13 Visualization of the filtered distance matrix of multiple scenarios.

Engineering, Vol. 13, No. 4, pp. 44–49, 2011.

- 5) Zhu, P., Hori, M., Kiyono, J. and Fujino, Y.: Towards Constructing Computer Model of Meteropolis for Consensus Recognition of Earthquake Hazard, *SOCIOTECHNICA*, Vol. 2, pp. 435–443, 2004.
- 6) Murrell, P.: *R Graphics*, Chapman & Hall/CRC, 2005.
- 7) Friendly, M.: Mosaic displays for multi-way contingency tables, *Journal of the American Statistical Association*, Vol. 89, pp. 190–200, 1994.
- 8) Ester, M., Kriegel, H.P., Sander, J. and Xu, X.W.: A density-based algorithm for discovering clusters in large spatial databases with noise, *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pp. 226–231, 1996.

(Received 5 Nov. 2012, Revised 5 Feb. 2013, Accepted 23 Feb. 2013)