Big visual data analytics for damage classification in civil engineering

C.M. Yeum, S.J. Dyke*, J. Ramirez, and B. Benes

Purdue University, Indiana, USA
** Corresponding Author*

ABSTRACT Visual data provide a wealth of information to better understand the world around us. A tremendous amount of visual data is collected in civil engineering applications through efforts such as scientific experiments, field surveys, resource management, and reconnaissance missions. Among these efforts, visual data generate crucial and abundant information evaluating the condition of a civil structure. As a typical example, during a disaster such as a natural catastrophe or industrial explosion, vast amounts of perishable image data are collected that may be used to generate new knowledge from the consequences of that event. However, not only does this process require time-consuming data collection by human engineers, it is also tedious and expensive to manually search through these data sets to find the most informative images. Autonomous collection, processing and analysis offer great potential to support structural evaluation. In this study, we propose a novel autonomous evaluation method to examine large volumes of images. Recent deep convolutional neural network (CNN) algorithms are applied to extract visual information from the collected images. Task-oriented engineering knowledge and experience are incorporated into the procedures to increase accuracy. The target application addressed in this study is post-disaster building damage evaluation. Illustration of the technique and capabilities for collapse classification is demonstrated using large-scale images gathered from past earthquake events.

1 INTRODUCTION

An astonishing amount of visual data is being collected worldwide through scientific experiments and field surveys in civil engineering. For example, during each natural disaster, vast amounts of perishable visual data are collected formally by teams of experts. That data is collected in order to generate new knowledge by learning from that event. With the availability of ubiquitous visual data sources, such as social media, news media and unmanned aerial vehicles for hire, large volumes of useful visual data are available for various purposes (Voigt et al. 2007; Yates & Paquette 2011; Computing Community Consortium 2013; Measure & American Red Cross 2015; Wang et al. 2015). Currently the major approach available to responders and researchers for the analysis of such data is tedious manual sorting and analysis of these photographs or videos. Only a small portion of the growing volumes of visual data collected are actually being used for research and extracting information for onsite decisions. Engineers must exploit the power of information technology to extract such information in an efficient manner.

Autonomous collection, processing and analysis of data offer great potential to aid the human decision maker. Whether he/she is conducting scientific research, or making a decision, the human operator and decision maker needs to quickly sift and sort visual data to identify and evaluate the important scenes. There is a compelling need to provide support to these human decision makers by giving them the analytic power to search, filter, prioritize, classify and annotate hundreds or thousands of images.

In recent years, powerful computer vision methods and machine learning algorithms have been established within computer science and engineering, and related disciplines. In some applications these have nearly reached human-level performance (Taigman et al. 2014; He et al. 2015). These methods have been considered for a broad range of applications, ranging from speech or text recognition to autonomous driving (Ciresan et al. 2012; Hannun et al. 2014; Chen et al. 2015). In civil engineering, vision-based remote or crowdsourcing structural inspection and construction management techniques have been researched and have achieved improvements in accuracy and efficiency (Golparvar-Fard et al. 2009; Jahanshahi et al. 2009; Ghosh et al. 2011; Zhu et al. 2011; German et al. 2012, 2013; Torok et al. 2013).

Previously developed techniques have been validated for specific damage types using a small quantity of images that were collected with the intension of using them for a specific purpose or application. However, in real circumstances during a disaster with realistic time and resource constraints, there is no guarantee that one may be able to collect favorable images for such a specific purpose due to the large uncertainty of locations, viewpoints, or contents. Thus, classification and filtering of the images will be able to support the decision-maker, particularly when time is limited. Furthermore, there is no assurance that these methods will be able to handle large-scale, complex, and unstructured images in such a way as to be tractable.

At this time such methods, when used in isolation, are still severely limited in their ability to extract useful information. Real-world visual data are quite diverse in nature (e.g. quality, resolution, subject, composition, illumination). And although these traditional computer vision methods are quite powerful, we still lack a good understanding of how to implement them in a domain-centric and task-oriented manner. Objects of interest to a civil engineer often need to be understood and analyzed in their spatial configuration as well as in their background context, and to extract meaningful information such analysis should be firmly based on engineering experience and knowledge.

In this paper, we develop and demonstrate a novel method for autonomous big data analytics that is intended to support decision-making in the field. The target application we addressed here is post-event building evaluation during a disaster. We implement task-oriented computer vision methods capable of detection, classification and evaluation of large volumes of visual data. A key factor in the method is that we incorporate prior knowledge from our target application, while we also investigate and utilize required optimal resolution of images for the scene classification and object detection actions. Recently developed deep convolutional neural networks (CNN) algorithm are applied for image classification and object detection, while ensuring success by integrating engineering domain knowledge into the procedure (Krizhevsky et al. 2012; Girshick et al. 2013; Russakovsky et al. 2014; Zhou et al. 2014; Simonyan & Zisserman 2014). Here we provide a feasible solution for analyzing a large-scale collection of real images from disasters. The proposed method can be expanded to incorporate new or existing damage detection methods for broad application in a range of disasters.

2 PROPOSED DAMAGE EVALUATION METHOD

An overview of the procedure developed here is shown in Figure 1. In step 1, images collected during a disaster are automatically filtered and prioritized based on available metadata from the images. Metadata acquired at the time of image acquisition includes geospatial and temporal (e.g. time/date, GPS) data, as well as information relevant to the event itself (e.g. previous building images, event intensity map). When available, these items can be incorporated into the overall process. The use of such metadata is beneficial for rapid access and flexible mining of a set of valuable images needed to explore their visual contents. In step 2, images are classified according to their content, particularly in terms of containing scenes associated with the target application. Scenes may be defined in terms of single or multiple objects and their spatial configuration. For instance, a scene of a building façade is composed of one or more objects and their spatial arrangement including an entrance at the bottom, and an array of windows or floor borders. Because scenes are typically recognized by low-dimensional features (e.g. general shape, colors, or compositions) and need not be interpreted using a detailed appearance of objects, they can readily be recognized in low resolution images. Thus, efficient and rapid computation is possible in this step. In step 3, specified target objects are identified and localized within the scenes classified in step 2. Here, the target objects encompass damage (e.g. spalling or cracking), and also the objects in which such damage is present (e.g. beam, column, wall). They may also include geometric or pattern features (e.g. window opening, stair-step cracking), which may be used subsequently for damage evaluation. Such object and scene categories must be designed and defined based on the applications for which they will be used. Lastly, in step 4, damage contained in those images is evaluated based on prior knowledge of the target applications. This step is performed to understand damage using its presence as well as its location on the object, appearance or surrounding objects. The steps proposed here are flexible depending on level of information provided with images.

A good example to illustrate the proposed concept is buckling detection after an earthquake. Buckling is detected by observing whether vertical rebar on a structural column is exposed and yielded. Such a domain-based definition of the problem provides good prior information to design the proposed technique. Instead of direct detection of rebar across a large collection of high-resolution images, the proposed method includes several steps: metadata filtering (with respect to location, date, or time), scene classification of indoor or outdoor building (or building façade), column/spalling object detection, and rebar detection. These steps are performed sequentially by gradually increasing the resolution of the images used for the relevant steps. Finally, we are able to evaluate the condition (e.g. bending or break) from the high resolution image containing the detected rebar. At each step, the number of images that are needed for processing with higher resolution decreases at each step. Thus, this approach is especially appropriate when time and resource constraints exist.

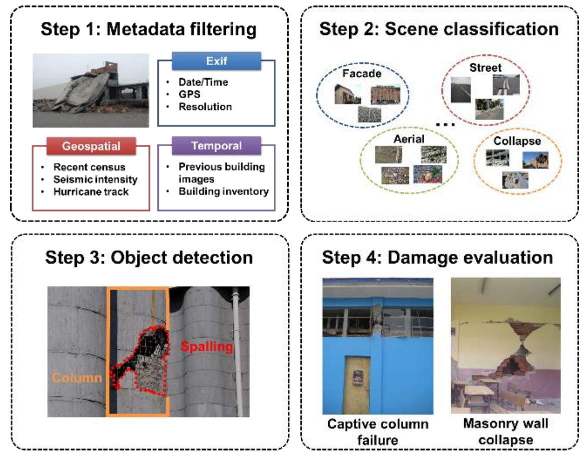

Figure 1. Overview of the proposed damage evaluation method

3 COLLAPSE CLASSIFICATION

As a pilot study, we use the case of collapse to demonstrate scene classification in the proposed method. The term collapse here refers to both significant damage/collapse of the building structure, as well as major damage/collapse of a single structural component. The reasons for selecting this damage case are that (1) collapse of buildings or their components represents a major mode of damage that is of interest in earthquake reconnaissance; (2) a large number of appropriate images are available for training and validation; and (3) there is no existing annotation image database for this situation. To perform this case study, a large annotation image collection is established and used for validation of the method.

We first introduce our post-event reconnaissance image data collection. We have gathered a collection of 67,000 color images acquired by various researchers and practitioners after past natural disasters including hurricane, tornado and seismic events (e.g., from datacenterhub.org at Purdue University, disaster responders, or Earthquake Engineering Research Institute (eeri.org)). Nearly all of these images preserve the original quality (resolution) as well as the basic information (e.g. date, time, and event), and a small portion of images have GPS information or a picture of a GPS navigator. However, no annotation was available for the visual contents of the images. At this time, the distribution across the types of events is earthquake (90%), hurricane (6%), tornado (3%), and others (1%). These images are collected from several different events such as earthquakes (e.g. Haiti in 2010, L'Aquila in 2009, Nepal in 2015), hurricanes (e.g. Florida in 2004, Texas in 2008), tornadoes (Florida in 2007; Greensburg in 2007). We will continue to collect images from such events to further integrate into the collection.

For assessment of the proposed technique, and algorithmic training, all images are first manually annotated using in-house annotation software. A single image is shown centered in the screen and annotators are asked to answer a yes or no question of "Does the image contain a collapse scene?". Based on our experience, such a binary classification yields better results than multiple choice questions. Manual annotation can be quite taxing for the human if there are several buttons/options and there is high potential for error.

Figure 2 shows several sample images used for collapse scene classification (from datacenterhub.org at Purdue University). For collapse classification, the dataset is composed of 1918 collapsed building and building components (b&bc) data as positive and 3427 other data as negative, which are composed of minor damage b&bc, irrelevant images, and undamaged b&bc in Figure 2. Such sampling of the negative dataset is designed to represent non-collapse image collection from a real earthquake reconnaissance scenario.

As mentioned in the introduction, the CNN algorithm will be implemented for collapse classification. In the last few years, CNNs have led to major breakthroughs in computer vision areas and have enabled the development of high-level abstractions using large-scale databases of general everyday objects. CNNs typically have one or more convolutional layers tied with weights and pooling layers to extract scale, translation and rotation tolerant features, and fully connected layers connected with these features classify image category or object(s). The parameters of CNNs are trained in advance of their implementation using large-scale training image data (Krizhevsky et al. 2012; Girshick et al. 2013; Russakovsky et al. 2014; Zhou et al. 2014; Simonyan & Zisserman 2014).

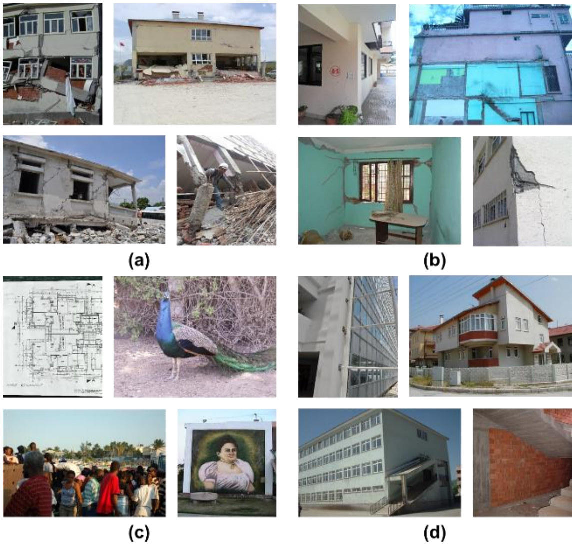

Figure 2. Sample images used for collapse scene classification: (a) collapse buildings and building components (b&bc), (b) minor b&bc, (c) irrelevant images, and (d) undamaged b&bc. (a) is assigned in a positive class and the others are in a negative class.

In this study, we used Vgg-f CNN architecture implemented in MatConvNet (Vedaldi & Lenc 2014), which comprises 8 learnable layers, 5 of which are convolutional, and the last 3 are fully-connected. Fast processing is ensured by the 4 pixel stride in the first convolutional layer. In the data augmentation process which produces a suitable set of input images for the CNN, a square region of input images are randomly cropped in original images followed by resizing them as 224 x 224 pixels. In each epoch, a batch at each iteration is assigned using randomly ordered pictures, and these data are augmented. Stochastic gradient descent with a batch of images are learned to optimize the parameters of the network.

Figure 3. Examples of the collapse classification result: (a) classification of positive images and (b) classification of negative images. Note that the text labels indicate classification results, and blue and red colors are true and false classification, respectively (e.g. an image in (a) having a label of "collapse" in blue is true classification collapse building image)

A workstation having a Xeon E5-2609 CPU, 12 GB memory and two GPU, NVidia Titan X and Telsa k40, a total of 24 GB video memory is used for training and testing the algorithm. The MatConvNet library installed on Matlab 2014b is used for this demonstration (Vedaldi & Lenc 2014). To obtain these results, 80% of annotated images are used for training and the remainder are used for testing and assessment of the classifiers. Less than 20 epochs are required to reach convergence and this training processing required approximately 10 hours.

Finally, in this demonstration we obtain rates of 86.6% true positive (true collapse detection), 13.3% of false-positive, and 93.7% of false negative, respectively. The proposed collapse classification successfully attains a relatively high rate for true-positives. A sample of images showing the classification results are shown in Figure 3. Note that these rates will vary slightly depending on CNN architectures and their parameters. Overall the performance of this approach is quite successful. The method shows great promise for supporting decisions in the field and for enabling research using large volumes of image data.

4 CONCLUSION

A novel method for automated post-disaster image classification is proposed to perform processing and analyzing big visual data. The method is demonstrated on a specific example classification focused on collapse classification. However, the general method can be applied to other civil applications that use large-scale visual data. In the future we plan to incorporate and validate a broader array of damage evaluation methods for broader application.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Santiago Pujol at Purdue University, Dr. Chungwook Sim at University of Nebraska, Dr. Alberto Pavese, and the EERI teams for providing access to large-volumes of post-earthquake reconnaissance images and Jongseong Choi at Purdue University for annotating large images.

REFERENCES

Computing Community Consortium (CCC). 2013. Computing for Disasters, *a workshop report funded by the National Science Foundation online at*: http://www.cra.org/ccc/files/docs/init/computingfordisasters.pdf

Ciresan, D.C., Meier U., & Schmidhuber, J. 2012. Transfer Learning For Latin And Chinese Characters With Deep Neural Networks, *The 2012 International Joint Conference on Neural Networks (IJCNN). IEEE.*

Chen, C., Seff, A., Kornhauser, A., & Xiao, J. 2015. DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving. *arXiv preprint* arXiv:1505.00256.

German, S., Brilakis, I., & DesRoches, R. 2012. Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments, *Advanced Engineering Informatics*, **26(4)**, 846-858.

German, S., Jeon, J. S., Zhu, Z., Bearman, C., Brilakis, I., DesRoches, R., & Lowes, L. 2013. Machine vision-enhanced postearthquake inspection, *Journal of Computing in Civil Engineering*, **27(6)**, 622-634.

Ghosh, S., Huyck, C. K., Greene, M., Gill, S. P., Bevington, J., Svekla, W... & Eguchi, R. T. 2011. Crowdsourcing for rapid damage assessment: the global earth observation catastrophe assessment network (GEO-CAN). *Earthquake Spectra*, **27(S1)**, S179-S198.

Girshick, R., Donahue, J., Darrell & T., Malik, J. 2013. Region-based Convolutional Networks for Accurate Object Detection and Segmentation, *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 99, 1-11.

Golparvar-Fard, M., Peña-Mora, F., & Savarese, S. 2009. D4AR–a 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication, *Journal of information technology in construction*, **14**, 129-153.

Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E.,Prenger, R., Satheesh, S., Sengupta, S., Coates, A. & Ng, A. Y. 2014. DeepSpeech: Scaling up end-to-end speech recognition, *arXiv preprint* arXiv:1412.5567.

He, K., Zhang, X., Ren, S., & Sun, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, *arXiv preprint* arXiv:1502.01852.

Jahanshahi, M. R., Kelly, J. S., Masri, S. F., & Sukhatme, G. S. 2009. A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures, *Structure and Infrastructure Engineering*, **5(6)**, 455-486.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, Rahul, Fei-Fei, L. 2014. Large-Scale Video Classification with Convolutional Neural Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Measure & American Red Cross. 2015. Drones for disaster response and relief operations.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. & Fei-Fei, L. (* = equal contribution) 2014. ImageNet Large Scale Visual Recognition Challenge, arXiv:1409.0575.

Simonyan, K., & Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556.

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. 2014. Deepface: Closing the gap to human-level performance in face verification. *In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on IEEE*, 1701-1708.

Torok, M. M., Golparvar-Fard, M., & Kochersberger, K. B. 2013. Image-based automated 3D crack detection for post-disaster building assessment, *Journal of Computing in Civil Engineering*, **28(5)**, A4014004.

Vedaldi, A., & Lenc, K. 2014. MatConvNet-convolutional neural networks for MATLAB. *arXiv preprint* arXiv:1412.4564.

Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., & Mehl, H. 2007. Satellite image analysis for disaster and crisis-management support, *Geoscience and Remote Sensing, IEEE Transactions on*, **45(6)**, 1520-1528.

Wang, S., Purnell, N. & Bhattachary, S. 2015. Nepal aid workers helped by drones, crowdsourcing, *Wall Street Journal-Asia Ed.*, May 1.

Yates, D., & Paquette, S. 2011. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International Journal of Information Management*, **31(1)**, 6-13.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. 2014. Learning deep features for scene recognition using places database, *In Advances in Neural Information Processing Systems*, 487-495.

Zhu, Z., German, S., & Brilakis, I. 2011. Visual retrieval of concrete crack properties for automated post-earthquake structural safety evaluation, *Automation in Construction*, **20(7)**, 874-883.