Imagery-Based Risk Assessment Using Crowdsourcing Technology in Complex Workspaces

Cheng Zhang[1], Pingbo Tang[2], Pin-Chao Liao[3], and Yi Ren[4]

[1]Del E. Webb School of Construction, Arizona State University, 660 South College Avenue, Tempe, AZ, 85281; email: cheng.zhang.7@asu.edu
[2]Del E. Webb School of Construction, Arizona State University, 660 South College Avenue, Tempe, AZ, 85281; email: tangpingbo@asu.edu
[3]Department of Construction Management, Tsinghua University, Beijing 100084, P. R. China; email: pinchao@tsinghua.edu.cn
[4]School for Engineering of Matter, Transport & Energy, Arizona State University, Tempe, AZ, 85281; email: yiren@asu.edu

ABSTRACT

Risk assessment based on imagery data is becoming popular in construction project management because cheap imaging devices can capture reality in real time. One challenge is that image-based safety risk identification heavily relies on the subjective image interpretation. Well-trained inspectors could be a limited resource that might not be always available to meet the safety inspection requirements on large and busy jobsites. One approach to address this challenge is "crowdsourcing" – collecting risk recognition results through online image interpretation games, and aggregate answers from game players into comprehensive and reliable risk recognition results. Unfortunately, unreliable answers from some players who lack professional inspection training can distort the aggregated results, thus make the answer of the majority not trustable. This paper presents a study showing that without knowing the actual risks on construction sites, human-image interaction behavior analyses of anonymous online image interpreters can overcome the influences of mixed reliable and unreliable answers from anonymous online players and acquire reliable risk evaluations of jobsite imageries. We conducted this study with the following steps: 1) train anonymous assessors in a few minutes on limited safety rules and have them assess some images for collecting crowd-based risk detection data, 2) collect assessment results using an online crowdsourcing platform, and 3) automatically aggregate risk assessments using a Bayesian network model into the risk detection results. Results from an online image-based elevator installation risk assessment experiment shows that the proposed method can overcome the limitation of the majority-based voting and achieve comparable results as experienced safety inspection professionals.

INTRODUCTION

Imageries have shown potentials for supporting risk management in construction and civil infrastructure management. In both China and the United States, civil infrastructure management agencies use imaging sensors for collecting detailed spatiotemporal information of bridges, dams, and other large structures for detailed condition assessment and risk analysis (Zhu and Brilakis 2010). Efficient and effective uses of imagery data for risk recognition is thus becoming increasingly important for establishing a data-driven risk management framework for civil engineering projects (Zhang and Tang 2015). Unfortunately, subjective image interpretation manually conducted by inspectors brings uncertainties and biases in risk recognition results based on images. Even well-trained inspectors spend a lot of time for achieving comprehensive and reliable risk recognition from images (Lagasse et al. 2009). In some cases, the uncertainties and biases within manual image interpretation processes can mislead the decisions about construction safety management and civil infrastructure maintenance (Moore et al. 2001). Civil engineers have been developing methods to increase the reliability of manual image interpretation in construction safety (Chang and Liao 2012; Lattanzi and Miller 2014; Papaelias et al. 2016). Some researchers examined image processing algorithms that can automatically extract certain features from images to assist engineers in identifying risks of construction (Chang and Liao 2012) . However, engineers still need to decide how to setup and use such image processing algorithms so that the subjective factors still exist (Moore et al. 2001). At present, interpretation of the images based on human intuition and experiences seems still unavoidable.

Recently, crowdsourcing is becoming a promising tool for reducing the training costs of professional inspectors while maintaining or even improving the reliability of inspection (Brabham 2008; Kittur et al. 2008; Liu et al. 2015). A "crowdsourcing" approach collects risk recognition results through online image interpretation games, and aggregates answers from game players into hopefully more comprehensive and reliable risk recognition results (Brabham 2008; Poetz and Schreier 2012). Crowdsourcing integrates the recognition power of human individuals into formal reasoning and pattern classification algorithms for solving image analysis tasks that are challenging state-of-the-art computer vision methods. Such approach is promising in construction engineering/management and civil engineering domains with unpredictable human behavior and dynamic working environment. Examples of such tasks include image segmentation, object recognition, worker activity analysis, and scene understanding (Brabham 2008; Liu et al. 2015; Poetz and Schreier 2012; Ranard et al. 2014).

Unfortunately, while applying such methods to risk assessment based on images, unreliable answers from some game players who lack professional inspection training can cause biases in the aggregated results. As a result, taking the answer of the majority as the risk assessment result could be wrong (Burnap et al. 2015). Figure 1 visualizes such a bias of a groups of anonymous online image interpreters the majority of who provide a wrong answer about the risk captured in an image. Specifically, Figure 1(A) shows a crowdsourced risk assessment result of a jobsite picture which contains a violation of a construction safety rule. In the assessment result, the majority (65.5%) of the online anonymous image interpreters correctly identified the violated construction safety rule after observing the picture. Figure 1(B) shows the result of a jobsite picture which contains no violation, wherein the blue region shows the number of correct answers, which is "no violation." Unfortunately, in this case, the majority of the answers about type of safety rule violated in that picture mistakenly identified that the scene captured in the picture violated a particular safety rule due to the lack of knowledge about that particular rule (green region). Figure 1(B) also shows that the second largest portion of the game players chose "no violation," which is actually quite close to 23.8%, the majority, and other answers are not very far behind. Such diverse answers reflected confusions and disagreements among people about the image contents, and slight biases can easily cause switch of the "majority answer." New theories are thus necessary to overcome the limitation of majority-based approach to alleviate the distortions caused by biased or unprofessional answers.

This paper presents a study showing that without knowing the actual risks on construction sites, human-image interaction behavior analyses of anonymous online image interpreters can overcome the limitation of the majority-based approach and acquire reliable risk detection from jobsite imageries. We conducted this study with the following steps: 1) train anonymous interpreters in a few minutes on limited safety rules and then have them assess a few images to identify violated safety rules, and 2) automatically aggregate the risk assessments of trained interpreters using a Bayesian network model into final results. Results from an online image-based elevator installation risk assessment experiment shows that the proposed method can achieve reliable results comparing with the result following the majority's vote.

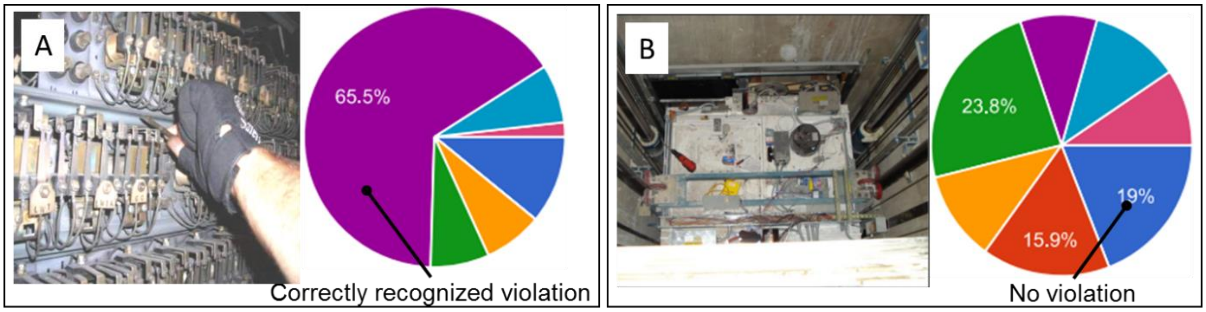

Figure 1 Crowdsourced risk assessment result of different jobsite pictures. A: picture with a violation of "wearing or holding metallic objects around live equipment"; B: picture with no violation.

CROWDSOURCING-BASED RISK ASSESSMENT METHOD

Overview

The overall goal of this research is to test the hypothesis that without knowing the actual risks on construction sites, the crowdsourcing-based risk assessment method can automatically aggregate risk assessments of anonymous people who just took simple training about construction safety rules into reliable risk evaluations. Figure 2 shows the IDEF0 model of the proposed crowd-based assessment method. Its inputs consist of three parts: 1) pre-defined safety rules of the construction activities; 2) jobsite pictures that may contain cases that violating safety rules; 3) the common sense about safety from anonymous participants on the internet. The output is the probability of the event that the assessed jobsite picture violates each safety rule in the predefined list of rules. The constraints are the Bayesian's rule and the training data set. The mechanisms of this crowdsourcing-based risk assessment involve the training process of anonymous image interpreters, the testing process for collecting risk assessments of these interpreters, and the decision-making process that determines the probabilities of the image certain rule violations.

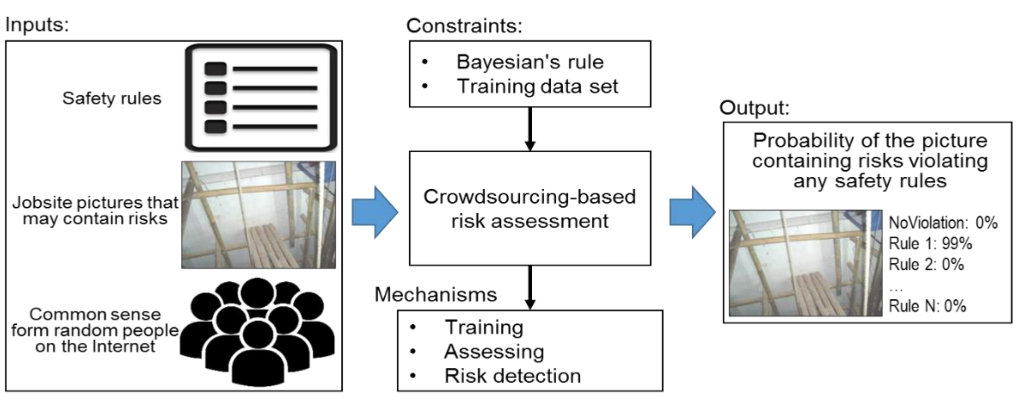

Figure 2 IDEF0 model of crowdsourcing-based risk assessment

The assumptions we made in this researches are as follows:

- Assessors are independent, i.e., the assessment on one design from one assessor will not be affected by the assessment of other assessors.
- The probability of each rule is violated in a given image is dependent, which is obviously not true in the real jobsite. We will explore the correlation that different rules are violated in one jobsite picture in future research.
- Each jobsite picture may contain at most 1 violation.

Figure 3 shows the process of the crowdsourcing-based risk assessment. Jobsite pictures that may contain violations will be uploaded to the online crowdsourcing platform. Then the platform will set up a violation assessment game played by anonymous participants to identify the violations of safety rules based on images collected on the jobsite. The game consists of three steps: the training process, the assessing process, and the risk-detection process. In the training process, the participants are divided into groups and then asked to learn about a sections of safety rules using the training materials. In the testing process, these participants will become "risk assessors" and identify potential violations against each safety rule they have learned in the training process. Finally, the decision making process will aggregate the answers from the assessors and then uses Bayesian network to determine whether a jobsite picture contains the violation against any safety rules. Among answers from different groups of assessors, the safety-rule violation that received the highest probability in the picture is considered as the actual violation captured in the picture.

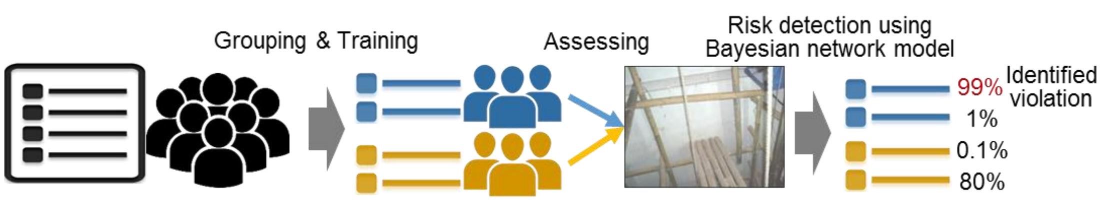

Figure 3 The process of the crowdsourcing-based risk assessment

The training process on the crowdsourcing platform

The purpose of the training process is to provide the anonymous participants basic understanding about the safety rules in a short period of time. The training process should be informative and easy to understand for normal people. In the training process of the crowdsourcing platform, all the safety rules will be divided into sections and learned by different group of assessors in order to reduce the cognitive workload of the assessors. The training material of each safety rule will show the dangerous and safe scenes. Figure 4 shows the training material of an example safety rule.

The assessing and risk-detection process

The risk detection process consists of the assessing process and the decision making process. The assessing process will collect the assessments of jobsite pictures. Then the assessors need to assess whether each picture violates any safety rules they just learned in the training process. The test pictures may have violations of the safety rules that the assessors had not learned in the training process, in which case the assessors were asked to select "no violation" to avoid ambiguity. Then the decision making process will use a Bayesian network model to calculate the probability of each safety rule is violated using the answers from the assessing process. Section 3 will introduce the Bayesian network model in detail.

Figure 4. Training materials of a safety rule "improper working platforms on scaffold structure".

A BAYESIAN NETWORK MODEL FOR CROWDSOURCING RISK DETECTION

The Bayesian network model aims at determining the probability that a violation of a rule exists in a jobsite picture, given all the assessments from one group of assessors who learned one section of safety rule and then assessed that image on the crowdsourcing platform. Without losing the generality, the whole crowdsourced risk assessment platform involves *I* jobsite pictures, *N* anonymous online image assessors, and *K* safety rules. The input of the Bayesian network, which is the assessment of a picture *i* from *N* assessors, is:

$$A_i = \{a_{i0}, a_{i1}, \dots\ a_{ik}, \dots\ a_{iK}\}, \sum_{k=0}^{K} a_{ik} = N$$

where *k* denotes a certain safety rule. Notice that the assessors can also choose "no violation" in the picture, noted as $k = 0$. Therefore, a_{i0} denotes the number of people that think the picture *i* contains no violation against the rules they have learned in the training process, while a_{ik} $(k \neq 0)$ denotes the number of people that think the picture *i* contains violations against safety rule *k*.

We define r_{nik} as the assessment of picture *i* from the assessor *n* for rule *k*. $r_{nik} = 1$ means that the assessor *n* thinks picture *i* is violating rule *k*, while $r_{nik} = 0$ means no violation. The ground truth of picture *i* capturing a violation *k* or no violation is defined as $\Phi_i = 0,1,2,\dots,K$. We assume that the probability of choosing the wrong options in this image assessment game are the same. The assessment r_{nik} is modeled as a random variable following a categorical distribution, as detailed by Eq. (1):

$$P(r_{ni}) = f(r_{ni} = k\,|\boldsymbol{p}) = p_k = \begin{cases} p_{tp}, & when\ k = \Phi_i \neq 0 \\ \dfrac{1-p_{tp}}{K}, & when\ k \neq \Phi_i\ an\&d\ \Phi_i \neq 0 \\ p_{tn}, & when\ k = \Phi_i = 0 \\ \dfrac{1-p_{tn}}{K}, & when\ \ k \neq \Phi_i = 0 \end{cases} \tag{1}$$

where $\boldsymbol{p} = (p_{1,}p_{2,\dots,}p_K)$**.** p_{tp} is true-positive rate of an assessment r_{nik}, which is the probability of an assessor correctly identifying the violation of a safety rule in the jobsite picture. Similarly, p_{tn} is the true-negative rate, which is the probability of an assessor correctly answering that there is no violation in the jobsite picture. The authors assume that the categorical distribution parameter $\boldsymbol{p}$ is the same for any *n, i, k*. p_{tp} and p_{tn} can be estimated using maximum likelihood estimation through training data set.

Now we would like to know the probability of the event "the picture contains a violation of rule k" when M out of N assessors chose rule k as the potential violation in the picture i. We denote this probability as $P(\mathrm{A}\,|B)$:

$$P(A\,|B) = P(B|A\,) \cdot P(A)/P(B) \quad (2)$$

where event A is that the option *k* represents the truth in the picture *i*; and event B is that M out of N assessors chose the option k for picture i. We will discuss how to calculate the probability of violating certain rule (k≠0) or no violation (k=0) based on (2) in the following subsections.

The probability of a certain rule being violated in a jobsite picture

In (2), $P(B|\mathrm{A}\,) = f(N, M, p_{tp})$ is the probability that M out of N people are choosing the rule k is violated when k≠0. $P(A)$ is the probability that rule k is violated in a jobsite picture which is provided by the historical safety assessment data. $P(B)$ is the overall probability that M out of N people are choosing rule *k* no matter rule *k* is violated or not, which includes three circumstances: 1) rule k is violated in the picture (event A, $k = \Phi_i$); 2) any other rule is violated in the picture, represented by the event A' ($k \neq \Phi_i, \Phi_i \neq 0$); 3) no violation in the picture, represented by A''($k \neq \Phi_i = 0$). We use $P(B|A') = f(N, M, p_{tp})$ and $P(B|A'') = f(N, M, p_{tn})$ to denote the probabilities that the assessors mistakenly chose rule k when the truth is "other rule is violated" and "no violations", respectively. As a result, $P(B)$ can be calculated by:

$$P(B) = P(B|A\,) \cdot P(A) + P(B|A') \cdot P(A') + P(B|A'') \cdot P(A'')$$

where $P(A')$ and $P(A'')$ are the probability of "other rules are violated" and "no violations", respectively, which are obtainable from historical risk assessment database.

The probability of "no violation" in a jobsite picture

In (2), $P(B|\mathrm{A}\,) = f(N, M, p_{tn})$ is the probability that M out of N people are choosing "no violation" when correct. $P(A)$ is the probability that the picture contains no violation. $P(B)$ is the probability that M out of N people are choosing rule k (k =0 means choosing "no violation") in both circumstances: 1) no violation in the picture ($k = \Phi_i = 0$); 2) violation exists in the picture, noted as event A' ($k = 0 \neq \Phi_i$). $P(B)$ can be represented with the help of $P(B|A') = f(N, M, p_{tp})$:

$$P(B) = P(B|A\,) \cdot P(A) + P(B|A') \cdot P(A')$$

where $P(A')$ is the probability of violating any rules. In sum, $P(A|B\,)$ can be calculated using $P(B|\mathrm{A}\,)$, $P(A)$, and $P(B)$:

$$P(A|B\,) = f(k, M, N, p_{tp}, p_{tn})$$

VALIDATION AND DISCUSSION

In the experiment, the authors use the safety rules and jobsite pictures of the elevator installation process to validate the effectiveness of the proposed risk assessment method. The training process chose the 12 rules that influence the elevator installation safety most. The 12 safety rules are divided into 6 rules/section* 2 sections and 3 rules/section*4 sections in experiment set 1 and 2, respectively to test how cognitive workload will influence the assessment performance. The assessing process collected answers from 389 anonymous participants using the crowdsourcing platform Amazon Mechanical Turk ("Amazon Mechanical Turk" n.d.).

Table 1 compares the true-positive rate p_{tp} and true-negative rate p_{tn} between two experiment sets. The true positive rates between two experimental sets are close. The difference between the true-negative rate may due to the fact that the number of choice in Set 2 are significantly fewer than that of Set 1. This result supports the hypothesis that: if the assessors are confident when they identify a violation in the picture; the assessors are less confident when they didn't find any violations in the pictures. This pattern enables the crowdsourcing method to identify the jobsite pictures containing violations of safety rules.

Table 1 Parameters of 2 experiment sets

Parameters	True-positive rate	True-negative rate
Set 1, 6*2	0.7152	0.3517
Set 2, 3*4	0.7884	0.5353

Table 2 shows the crowdsourcing-based risk assessment results of six jobsite pictures. The Bayesian network-based crowdsourcing assessment result is correct for all six texting jobsite pictures in both experiment sets. On the other hand, the majority voting result made 2 and 1 false-positive mistake in experiment set 1 and 2, respectively. Future research will explore the logical relationship between safety rules and how the relationship will influence the performance of crowdsourcing-based risk assessments.

Table 2 Risk assessment results

	Picture Number		A	B	C	D	E	F
Violating which safety rule ("0" means no violation)	Ground truth (pre-defined by experienced professionals)		7	0	12	10	0	0
	Set 1, 6*2	Bayesian network	7	0	12	10	0	0
		Majority vote	7	0	12	10	1	3
	Set 2, 3*4	Bayesian network	7	0	12	10	0	0
		Majority vote	7	0	12	10	7	0

CONCLUSION

This research shows the capability of using crowdsourcing-based assessment method to identify violations of safety rules based on images collected in complex construction workspaces. Comparing with the results of majority votes, the Bayesian network-based crowdsourcing risk assessment is reliable in eliminating the jobsite pictures which contains no violation against any safety rules while identifying the pictures capturing violations. The experiment result also validates the hypothesis that the assessors are confident when they identify a violation in the picture; the assessors are less confident when they didn't find any violations in the pictures. The high true-positive and true-negative rate of the proposing risk assessment method shows the potential of applying such risk assessment method in real-word project management in order to reduce the cost and improve the performance of safety control. The safety inspection outcomes can help management team to process on-site risk identification effectively and cost-efficiently.

ACKNOWLEDGEMENT

This material is based upon work supported by the U.S. Department of Energy (DOE), Nuclear Engineering University Program (NEUP) under Award No. DE-NE0008403 and Natural Science Foundation of China (Award No.: 51578317). Their supports are acknowledged. Any opinions and findings presented are those of authors and do not necessarily reflect the views of DOE and NSFC.

REFERENCES

"Amazon Mechanical Turk." (n.d.). <https://www.mturk.com/mturk/welcome>.

Brabham, D. C. (2008). "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases." *Convergence: The International Journal of Research into New Media Technologies*, 14(1), 75–90.

Burnap, A., Ren, Y., Gerth, R., Papazoglou, G., Gonzalez, R., and Papalambros, P. Y. (2015). "When Crowdsourcing Fails: A Study of Expertise on Crowdsourced Design Evaluation." *Journal of Mechanical Design*, 137(3), 31101.

Chang, P. C., and Liao, D. (2012). "Image-based structural damage assessment with sensor fusion." *Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications - COM.Geo '12*, ACM Press, New York, New York, USA, 1.

Kittur, A., Chi, E. H., and Suh, B. (2008). "Crowdsourcing user studies with Mechanical Turk." *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems*, 453.

Lagasse, P. F., Clopper, P. E., Pagan-Ortiz, J. E., Zevenbergen, L. W., Arneson, L. A., Schall, J. D., and Girard, L. G. (2009). *Bridge Scour and Stream Instability Countermeasures: Experience, Selection and Design Guidance. Volume 1*. techreport.

Lattanzi, D., and Miller, G. R. (2014). "3D Scene Reconstruction for Robotic Bridge Inspection." *Journal of Infrastructure Systems*, 21(2), 4014041.

Liu, K., Asce, S. M., Golparvar-fard, M., and Asce, A. M. (2015). "Crowdsourcing Construction Activity Analysis from Jobsite Video Streams." 141(11), 1–19.

Moore, M., Phares, B., Graybeal, B., Rolander, D., and Washer, G. (2001). *Reliability of Visual Inspection for Highway Bridges. volume I: Final report.*

Papaelias, M., Cheng, L., Kogia, M., Mohimi, A., Kappatos, V., Selcuk, C., Constantinou, L., Mu??oz, C. Q. G., Marquez, F. P. G., and Gan, T. H. (2016). "Inspection and Structural Health Monitoring techniques for Concentrated Solar Power plants." *Renewable Energy*, 85, 1178–1191.

Poetz, M. K., and Schreier, M. (2012). "The value of crowdsourcing: Can users really compete with professionals in generating new product ideas?" *Journal of Product Innovation Management*, 29(2), 245–256.

Ranard, B. L., Ha, Y. P., Meisel, Z. F., Asch, D. A., Hill, S. S., Becker, L. B., Seymour, A. K., and Merchant, R. M. (2014). "Crowdsourcing--harnessing the masses to advance health and medicine, a systematic review." *Journal of general internal medicine*, 29(1), 187–203.

Zhang, C., and Tang, P. (2015). "Visual Complexity Analysis of Sparse Imageries for Automatic Laser Scan Planning in Dynamic Environments." *Computing in Civil Engineering 2015*, American Society of Civil Engineers, Austin, TX, 271–279.

Zhu, Z., and Brilakis, I. (2010). "Machine Vision-Based Concrete Surface Quality Assessment." *Journal of Construction Engineering and Management*, ASCE, 136(2), 210–218.