Crowdsourcing for Traffic Survey Work on Amazon Mechanical Turk

Takahiro Koita and Shota Suzuki

***Abstract*—The main current method for obtaining traffic volume data in surveys involves survey personnel using counters. However, at least two persons are required for each lane to obtain a high level of accuracy, because the counting is done in real time. As a result, work cost is high and reducing cost while maintaining the same level of accuracy is a problem. As a cost reduction measure that addresses that problem, the use of crowdsourcing for measuring traffic is proposed and the usefulness of that measure is examined. Specifically, an evaluation of accuracy when data is acquired by an unspecified large number of workers shows that the proposed method can reduce survey cost while maintaining the same accuracy as the existing method by eliminating unproductive time. The results demonstrate the feasibility of applying crowdsourcing, which uses human processing capability as opposed to big data processing.**

***Index Terms*—Crowd sourcing, Amazon mechanical turk, traffic survey, worker processing.**

I. Introduction

Traffic surveys have attracted attention as a technology closely connected to the automobile-related industry, which has grown significantly in recent years. A traffic-volume survey counts the number of vehicles (by the type of vehicle) passing through a specific road. Such surveys are highly demanded and used for understanding current situations and forecasting in the future in cases such as mitigation of traffic jams and planning of constructed of new roads. These surveys are therefore indispensable for automobile administration targeting the rich future society heralded by the Japanese government. Currently, the demand for traffic surveys is high, and it is required to conduct traffic surveys more frequently than under the current situation. Moreover, the level of accuracy of a traffic survey must be high. This is because the level of accuracy of a traffic survey needs to be high so that current situations can be understood and future projections can be made in regard to tasks such as alleviating traffic jams and planning new roads. However, a traffic survey is mainly conducted by investigators using counters at survey points, and the cost of that work is high. In this study, this mainstream survey method is used as an existing method, and "work cost" is defined as the cost of the labor of the investigators on the premise that the investigators are ready to measure.

Traffic surveys are used to understand the current traffic flow through specific roads, make forecasts to reduce traffic congestion, and plan road construction [1], [2]. Vehicles are classified into five categories: passenger cars, buses, small freight trucks, ordinary freight trucks, and two-wheeled vehicles. High accuracy of survey results is required to achieve those purposes. Traffic surveys are frequent throughout Japan, and the high demand for surveys requires a reduction in work cost to enable more surveys to be conducted.

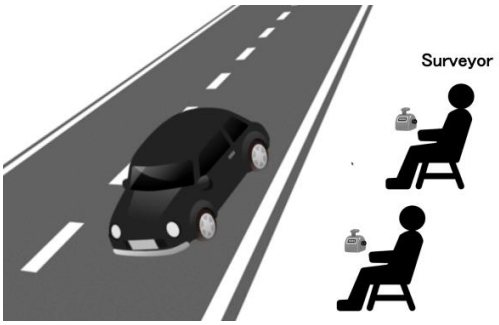

Fig.1. Current traffic survey method.

A method in which traffic volume is measured with a mobile traffic counting instrument that has an infrared sensor, an ultrasonic sensor, and a magnetic sensor has also been proposed [3]. However, the cost of installing such counters at all of the 36,000 or so survey points throughout Japan would be very high [4]. Another traffic survey method that uses a GPS history and is referred to as Probe Person [5] can classify moving things as trains or pedestrians. However, that method cannot classify vehicles into types, so the necessity of maintaining high accuracy for each vehicle type remains.

The current traffic survey method is illustrated in Fig. 1. Because vehicles must be counted visually in real time at the survey point, at least two surveyors are required at each survey point to maintain high accuracy, so the cost is high. Therefore, existing methods do not allow an expansion of traffic surveys.

Currently, surveyors measure traffic volume with counters at about 24,000 places, which is two thirds of the 36,000 survey sites in Japan [4]. The existing method requires payment of many surveyors who count traffic in real time. Because the counting is done in real time, there are times when there is no traffic and survey accuracy is not affected even if no counting is done during those times. Nevertheless, the surveyors must be paid for that unproductive time. We therefore focused on reducing cost by regarding time that does not affect the accuracy of the survey as time when surveyors are not needed and eliminating that time to reduce cost. With the current method, however, surveyors are paid by the hour whether the time is productive or not, so it is difficult to reduce cost by eliminating unproductive time. We therefore propose a method of paying for work that actually produces results rather than paying by time, thus reducing

Manuscript received February 5, 2019; revised June 12, 2019.

Takahiro Koita is with Department of Information Systems Design, Doshisha University, Kyoto, Japan (e-mail: tkoita@mail.doshisha.ac.jp).

Shota Suzuki is with Hitachi Systems, Tokyo, Japan.

DOI: 10.7763/IJCTE.2019.V11.1243

cost by eliminating the cost of the time when surveyors are not needed.

The above discussion shows the need to investigate methods of eliminating the cost of unproductive work by distributing the cost to tasks performed rather than to time. Here, we propose the use of crowdsourcing for conducting traffic surveys to reduce cost while maintaining the accuracy of the current method and investigate the utility of the proposed method.

II. Crowdsourcing

Crowdsourcing is one kind of participatory online activity in which individuals or companies allocate work to large groups of unspecified numbers of workers via a website [6]. In crowdsourcing, payment is made for tasks completed rather than by the hour (Fig. 2), so it is possible to inexpensively allocate only the necessary work to workers. It has been reported that this approach can reduce work cost by the reducing unproductive work time [7], [8]. Because crowdsourcing rewards workers by completion of assigned tasks rather than by time, the cost of unproductive time spent on the task can be eliminated.

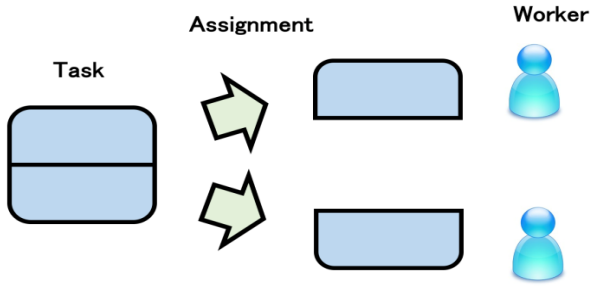

Fig. 2. Crowdsourcing.

Because typical crowdsourcing services such as Amazon Mechanical Turk can distribute tasks to workers around the world, work can be assigned to a large number of people in a short time (50 persons in one hour, etc.) [9], [10]. However, workers prefer simple tasks and many avoid difficult tasks, so it is necessary to divide work into simple micro-tasks that are readily accepted by workers and often requested cheaply [6]. Micro-tasks frequently include content that is difficult to process by computers, so tasks that cannot be performed by computers alone can be accomplished by combining the capabilities of computer programs and human intelligence.

In recent years, crowdsourcing has been used to solve a wide range of problems, but the workers who perform the work are diverse in origin, age, and occupation [6]. Some workers perform the tasks without having read the instructions, so the quality of the work performed is unclear at the time the work is requested. To obtain high quality work results, it is therefore necessary to make preparations such as dividing and © presenting work in a way that is appropriate for diverse workers.

When using crowdsourcing, it is necessary to consider which workers accept the work, because some workers do not read the specified explanations, abandon the work before it is finished, or do not finish the work by the deadline. It is also necessary to consider which crowdsourcing service to use, and select between Amazon mechanical Turk and Lancers according to which is most appropriate for the task.

Another factor that must be considered is how to appropriately distribute the work involved in existing methods by crowdsourcing. Currently, surveyors are dispatched to the survey site and are paid by time to perform the measurements, but in crowdsourcing via the Internet, the crowdsourced workers must be able to count traffic without going directly to the survey site. In the proposed method, the counting is performed using still images extracted from video acquired at the survey site, and those images are distributed as work tasks. The details of this process are described below.

III. Proposed Method

The proposed method is briefly described here. Applying crowdsourcing as a traffic survey method requires distribution of work through the Internet, so it must be possible to count traffic without actually being present at the survey site. We therefore propose acquiring video at the survey site, extracting still images from the video, and distributing the still images as work tasks by crowdsourcing, thus eliminating the need for the worker to be at the site. The use of still images for the counting task makes it easier to distribute the work in small units that are readily accepted by crowdsourced workers. Because the workers are paid only for completing the assigned tasks, the cost of unproductive time is eliminated and cost is reduced.

The proposed method is premised on video captured by a camera installed at the survey site. Still images are extracted from the video and assigned as work tasks (Fig. 3). The task is to count vehicles in five categories (passenger cars, buses, small freight trucks, ordinary freight trucks, and two-wheeled vehicles) in the same way as in the current method.

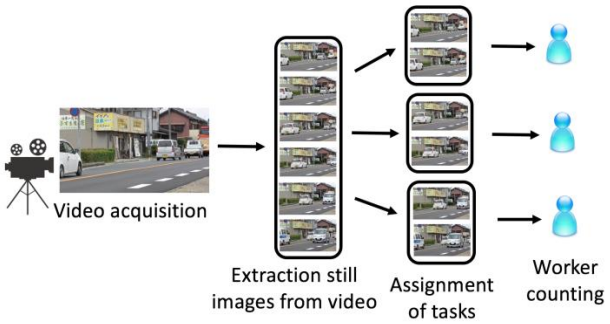

Fig. 3. Proposed method.

There are two major advantages of using an unattended portable video camera at the survey site. One is that video can be acquired without a surveyor present. The camera can be controlled remotely and crime prevention techniques are effective, so there is no need for the surveyor to go directly to the survey site. Another advantage is that portable cameras can be easily moved to different locations, thus facilitating the conduct of surveys at approximately 36,000 sites each year [4]. We therefore chose to use portable cameras in the proposed method.

Extracting still images from the video acquired at the survey site at specified time intervals (Fig. 4) makes it possible to assign work on a task basis rather than a time basis.

The task of counting traffic can thus be assigned by crowdsourcing.

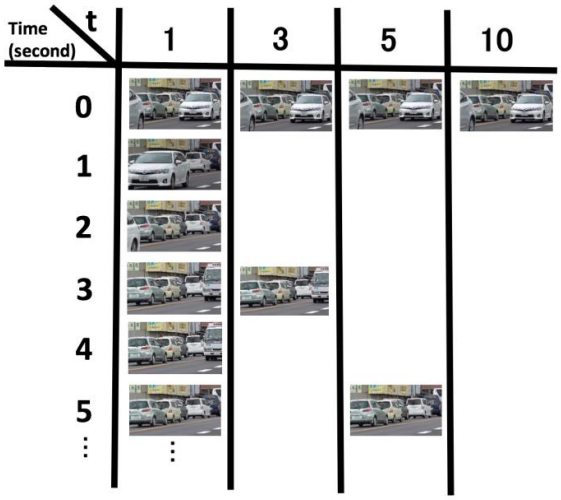

Fig. 4. Still image extraction.

IV. Evaluation

In this section, a preliminary experiment for determining the task of crowdsourcing is described, and the effectiveness of the proposed method is demonstrated by this experiment. In the experiment, a five-minute video was captured at an actual survey spot, and the level of accuracy when an unspecified number of workers measured still images extracted from the video at 1-, 3-, 5-, and 10-second intervals was evaluated . The level of accuracy is given as follows:

$$\text{Correct Answer Rate (\%)} = \left\{1 - \left(\frac{\text{Total error}}{Total\ number\ of\ correct\ answers}\right)\right\} * 100 \quad (1)$$

The total number of the vehicles (by the type of vehicle) were counted before the experiment and were compared with the numbers of vehicles counted by the worker. (The total number of the type of vehicles (by the type of vehicle) were considered the "accurate" number of vehicles.) The level of accuracy was expressed in a percentage and indicated how accurate the worker counted the vehicles. Each time the worker made a mistake in classifying the vehicle type, it was considered an error and was subtracted from the total number of vehicles (by the type of vehicle).

A. Preliminary Experiment

A preliminary experiment was conducted to evaluate, on the basis of information in video data, the possibility of maintaining the same degree of level of accuracy as attained by the existing method. In the preliminary experiment, crowdsourcing was assumed, and similar tasks were assigned to multiple subjects. And tasks were allocated to eight workers, and level of accuracy in regard to "still-image extraction interval" and "still-image allocation number" was evaluated. Also, since tasks were distributed to multiple workers, it is possible that a vehicle measured by the first worker will also be measured by the second worker. To prevent such a situation, in regard to the still images, the condition that only the front vehicle is to be measured (i.e., the car behind is not to be measured) was stipulated.

1) Still-image-extraction interval

Level of accuracy regarding still-image extraction interval during evaluation in the preliminary experiment is plotted in Fig. 5. The accuracy for still-image extraction intervals of 1 s second and 3 s are both 80% or more, whereas that for still-image extraction intervals of 5 s is about 55%, and when the interval is 10 s, it drops to 25%. As for the preliminary experiment, the level of accuracy for 1-s and 3-s intervals is close to that for the existing method; however, in the case that the interval is 5 s or more, the level of accuracy drops greatly. It is necessary to shorten the still-image extraction interval to maintain the same degree of accuracy as the existing method.

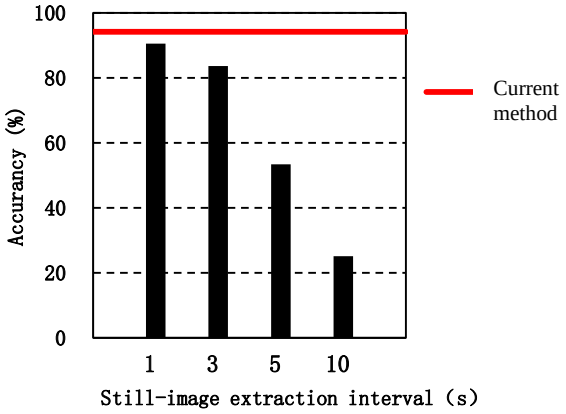

Fig. 5. Level of accuracy versus still-image extraction interval.

2) Still-image distribution number

The level of accuracy for the number of still images distributed during the evaluation in the preliminary experiment is plotted in Fig. 6. In the figure, "the number of still images" is the number of still images allocated to one worker. Although 25, 50, 100, and 300 still images per second were assigned to the worker, the influence on the level of accuracy due to the variation in the number of still images allocated side by side was small.

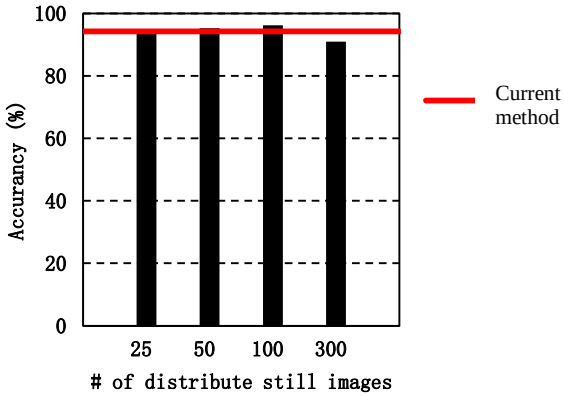

Fig. 6. Level of accuracy versus number of distributed images.

B. Main Experiment

The proposed method is evaluated in the Amazon Mechanical Turk crowdsourcing environment, in which it is possible to format various tasks in HTML for distribution. The format of the traffic survey is shown in Fig. 7. Crowdsourcing workers were asked to download a still image from the specified URL and perform the counting according

to an example presented in English. Task explanations that are presented as examples and diagrams are easily understood visually by workers.

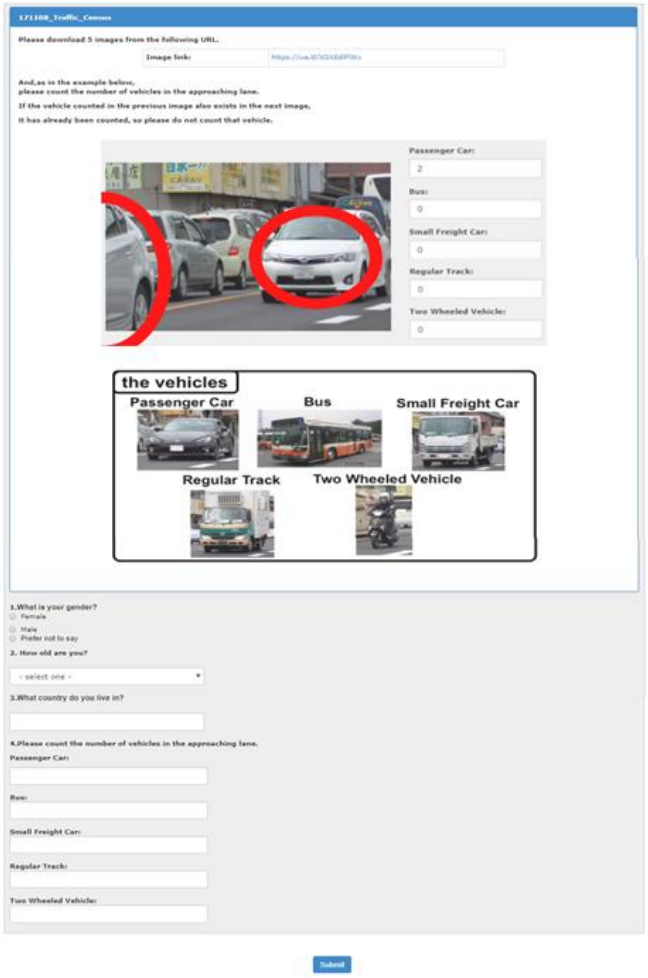

Fig. 7. Traffic survey format.

We investigated the reward for work performed in the Amazon Mechanical Turk environment, where payment can be set in units of $0.01, to determine what payment is effective for work acceptance by workers. Times from distribution of work with the same contents by Amazon Mechanical Turk until each reward is acceptable by the worker are listed in Table I. The result was that compensation of $0.01 or $0.05 was not sufficient for worker acceptance of work within one hour, but work was accepted when the offered compensation was $0.10. We therefore used that value for task payment.

TABLE I: REWARD AND TIME TAKEN UNTIL IT IS ACCEPTED (AVERAGE)

Reward	Time taken until being accepted
0.01 dollars	N/A
0.05 dollars	2.5 hours
0.1 dollars	20 minutes

Video of five minutes duration was acquired from cameras at survey sites and still images were extracted from the video at intervals of one second, three seconds, five seconds, and 10 seconds. Mosaic processing was applied to the still images, the processed images were distributed to an unspecified large number of workers by crowdsourcing, and the accuracy of the counting results was determined. The work was distributed to 20 workers using the Amazon Mechanical Turk service with a compensation of $0.10. Although it is unnecessary to use mosaic images of vehicles driving on public roads by law, we chose to do so in this research to avoid any possible problems [10].

The evaluated accuracies for various still image extraction intervals are shown in Fig. 8. For an extraction interval of one second, accuracy that is close to the current method was obtained, but accuracy declined as the extraction interval increased in the same way as observed in the preliminary experiment. It is therefore necessary to reduce the still image extraction interval to one second.

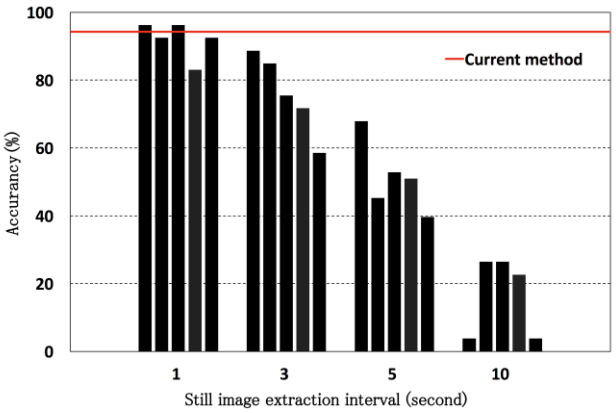

Fig. 8. Level of accuracy versus still-image extraction interval.

A box and whisker plot of the same accuracy results (Fig. 9) shows a difference of about 10% between the minimum and maximum accuracy values for the image extraction intervals of one, three, and five seconds, indicating a variance in the accuracy values, even for the same extraction interval. This result suggests that the task was not distributed as a simple task that all workers could easily understand. To reduce the observed variance, it is necessary to consider descriptions that are easily understood by any worker when creating the survey format with Amazon Mechanical Turk.

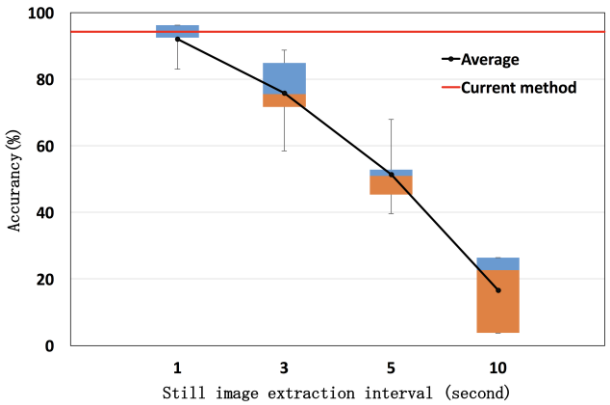

Fig. 9. Level of accuracy versus still-image extraction interval.

In this investigation of the utility of using Amazon Mechanical Turk, the cost for all tasks was set at $0.10, which corresponds to about ¥11 (for an exchange rate of ¥111 per US dollar on January 18, 2018). In the current method, compensation for surveyors is ¥1000 per hour, so the cost for a five minute survey is ¥83. It is therefore possible to greatly reduce work cost by applying the proposed method.

Information concerning the workers that participated in the experiment is summarized in Table II. It is clear that many of the workers were nationals of the USA and India. It is clear that there is some bias in workers participating in Amazon Mechanical Turk. In some preliminary experiments, more than 90% of the workers were from the USA. However, workers with nationalities other than those from the USA and India, such as Greece and Venezuela, also undertook work, so it is clear that work is distributed to workers around the world. However, workers around the world did not understand the contents of tasks distributed by using Amazon Mechanical Turk, and it turned out that about 20% of the workers took actions that differed from those in the explanation. Therefore, to reduce the number of such workers as much as possible, it is necessary to consider how to allocate simple tasks and how to improve work quality.

TABLE II: INFORMATION ABOUT WORKERS

Still-image-extraction interval	Gender	Age	Country
every second	male	31~40	USA
	male	21~30	India
	male	21~30	Venezuela
	male	51~60	USA
	male	21~30	USA
every 3 seconds	female	31~40	USA
	female	31~40	Botswana
	female	21~30	Turkey
	male	21~30	USA
	male	31~40	USA
every 5 seconds	female	31~40	India
	female	31~40	USA
	male	41~50	India
	male	31~40	Bosnia and Herzegovina
	male	21~30	USA
every 10 seconds	female	21~30	USA
	male	21~30	USA
	male	21~30	USA
	male	31~40	Greece
	female	21~30	India

V. CONCLUSION

We have shown that the proposed method of conducting traffic surveys using crowdsourcing reduces the cost of work while maintaining the accuracy of the current method. As issues for future work, it is necessary to investigate the evaluation of accuracy with respect to the distribution of the work and worker qualifications, and investigate the evaluation of accuracy with respect to changes in weather conditions.

REFERENCES

[1] Final Report of the Subcommittee for the Further Development of Automotive Administration for a Rich Future Society, Ministry of Land, Infrastructure and Transport, 2015.

[2] 2015 National Highway and Traffic Survey, General Traffic Survey Overview, Ministry of Land, Infrastructure and Transport, 2015.

[3] S. Itsubo, "Research on advanced and efficient use of IT in traffic surveys," Doctoral dissertation, Kyoto University, 2009.

[4] Uesaka, Hashimoto, and Kawano, "Advanced traffic surveys," Research materials, National Institute for Land and Infrastructure Management, 2009.

[5] Zenji, Horiguchi, Akahane, and Komiya, "Development of a method for estimating transportation modes with handy GPS equipment," presented at the 4th ITS Symposium, Chiba Institute of Technology, 2005.

[6] Y. Horita, "Analysis of the effects of non-monetary incentives for crowdsourced workers," Master's thesis, Kyoto University, 2015.

[7] M. Ashikawa, "Studies on quality improvement of large-scale data analysis using Microtask crowdsourcing and practical system deployment," Doctoral dissertation, University of Electro-communications, 2016.

[8] Goto, Ishida, and Lin, "Optimization of crowdsourcing workflow involving repeated and parallel processes," *IEICE Transactions*, vol. J99-D, no. 11, pp. 1094-1101, 2016.

[9] M. Buhrmester, T. Kwang, and S. D. Gosling, "Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data?" *Perspectives on Psychological Science*, vol. 6, issue 1, 2011.

[10] M. Yamamoto, "Tolerance and limitations of an automatic licence plate reading system (n-system)," *Meiji Gakuin University Revue*, no. 6, pp. 95-112, 2007.

Takahiro Koita received the PhD degree in engineering from Nara Institute of Science and Technology in Japan. He is currently an associate professor at Doshisha University in Japan. He was a visiting professor at University of Wisconsin-Madison in computer science department. His research interests include cloud computing, human computation, distributed processing and high-performance computing.

Shota Suzuki earned his BS degree in engineering from Doshisha University in Japan. His research interests include information systems, information technology, and crowd sourcing society. He has published papers in crowd sourcing and its application. He is a member of the Institute of Electronics, Information and Communication Enginners in Japan.