What Will Others Choose? How a Majority Vote Reward Scheme Can Improve Human Computation in a Spatial Location Identification Task

Huaming Rao[*,+] and **Shih-Wen Huang**[+] and **Wai-Tat Fu**[+]
[*]Department of Computer Science & Technology, Nanjing University of Science & Technology
[+]Department of Computer Science, University of Illinois at Urbana-Champaign
[*,+]{huamingr,shuang51, wfu}@illinois.edu

Abstract

We created a spatial location identification task (SpLIT) in which workers recruited from Amazon Mechanical Turk were presented with a camera view of a location, and were asked to identify the location on a two-dimensional map. In cases where these cues were ambiguous or did not provide enough information to pinpoint the exact location, workers had to make a best guess. We tested the effects of two reward schemes. In the "ground truth" scheme, workers were rewarded if their answers were close enough to the correct locations. In the "majority vote" scheme, workers were told that they would be rewarded if their answers were similar to the majority of other workers. Results showed that the majority vote reward scheme led to consistently more accurate answers. Cluster analysis further showed that the majority vote reward scheme led to answers with higher *reliability* (a higher percentage of answers in the correct clusters) and *precision* (a smaller average distance to the cluster centers). Possible reasons for why the majority voting reward scheme was better were discussed.

Introduction

Crowdsourcing has been shown to be useful for solving problems that can be decomposed into micro-tasks. In addition to capitalizing on the massive number of crowd workers, research on human computation has also shown that leveraging human knowledge or cognitive processes can be effective for accomplishing tasks that are beyond the capabilities of current automated approaches (Law and von Ahn 2011). Typically, researchers identify task components that are known to be difficult for machines, either because they are computationally expensive (Lasecki et al. 2012) or because they require semantic information that are difficult to be extracted in forms that can be processed automatically (Bernstein et al. 2010).

This paper focuses on methods that improve human computation using a spatial location identification task (SpLIT), in which humans examine a three-dimensional camera view of an environment to infer its spatial location on a two-dimensional schematic map (e.g. a floor plan). The SpLIT requires (1) detection of salient cues in the 3D camera view image that are relevant for spatial inferences (e.g. landmarks, orientation, etc.), (2) mapping of the salient cues between the 3D image and the 2D map, and (3) spatial inferences based on integration and mapping of cues in both the 3D image and the 2D map. These components are fundamental to the success of applications that support spatial or location-based tasks such as navigation (Afyouni, Ray, and Claramunt 2012), remote collaboration (Billinghurst and Kato 2002), spatial designs (Choi et al. 2007), and location-based image retrieval (Zhang et al. 2011).

Although they are fundamental to many spatial or location-based applications, these components in SpLIT are often difficult to automate because the identification of salient cues often requires semantic features that are challenging to recover from the image and the map. In addition, in many cases the cues are either ambiguous or do not provide sufficient information to pinpoint the exact location. On the other hand, humans seem to frequently encounter these ambiguous or underspecified situations and can manage them "naturally" — e.g. when one attempts to locate oneself in a shopping mall in a "you-are-here" map, when doctors attempt to infer the location of a tumor on an X-ray negative, or when interior designers collaborate with architects to create new floor plans. In many of these cases, humans are able to perform some variations of a SpLIT by "filling in" with semantic information that ranges from commonsense knowledge (e.g. what are usually in a shopping mall) to domain-specific knowledge (e.g. medical or architectural expertise). This form of semantic information, however, is often difficult to be extracted and used in fully automated systems.

One natural question to ask is how one can leverage human computation to perform a typical SpLIT. In order to find the answer, we designed a study using Amazon Mechanical Turk to investigate how turkers can perform the task in two kinds of reward schemes: ground truth and majority vote, even when they were not familiar with the environment. Even though previous research (Shaw, Horton, and Chen 2011) proved how majority voting could lead to better results for tasks, such as content analysis, it is still not clear whether and how it could lead to better results in a task such as SpLIT, which had an objective answer and required identification and integration of spatial cues — which made the SpLIT unique compared to other tasks. So the goal of the current study was to investigate to what extent human computation could be utilized in situation such as the SpLIT and how the different reward schemes might affect the performance in such task.

Related Work

There has been much work in crowdsourcing and human computation research introducing methods that leverage human knowledge to solve problems that are too difficult for machines. Human computation has been shown useful in image labeling and interpretation (Sorokin and Forsyth 2008), word processing (Bernstein et al. 2010), and protein folding (Eiben et al. 2012). Research on methods that improve the quality of solutions from humans has focused on introducing structures to task assignments (Kulkarni, Can, and Hartmann 2012), having workers evaluate each other (Dow et al. 2012; Huang and Fu 2013b), or leveraging social dynamics among workers to increase their motivation (Huang and Fu 2013a). These methods often rely on our understanding of how humans are motivated to put more effort into a certain task, how to provide more or better feedback to humans to help them learn to do the task better, or how to leverage social dynamics to guide humans to reflect on their cognitive processes to generate better answers.

Previous studies have shown that situating human workers in some forms of games can motivate workers to participate and encourage coordination among workers. For example, the ESP game (von Ahn and Dabbish 2004) encouraged humans to provide labels to images that match other's labels. Results showed that this form of coordination game was in general useful to improve the quality of outcomes. Given that there are usually plenty of cues (and labels) for humans to select in the image labeling task, each of these labels can be considered a potential focal point (Schelling 1960). Based on Schelling's theory, "a focal point (also called a Schelling point) is a solution that people will tend to use in the absence of communication, because it seems natural, special or relevant to them". It is therefore likely that the game would encourage the selection of most salient labels in the image, as these were more likely selected by others. It is, however, still unclear whether and how the focal points that emerge are related to the saliency of the cues or labels, and how the labels generated would be different from those generated without the game environment. In addition, the image labeling task is fundamentally different from SpLIT, as SpLIT has only a single correct answer which requires more interpretation and integration of cues to infer. The SpLIT also allows more direct manipulation of the level of ambiguity of cues to test how it impacts performance.

Automatically identifying locations have been an active area of research, as it is a fundamental process for a wide range of applications (Zamir and Shah 2010; Hansen et al. 2009; Mulloni, Seichter, and Schmalstieg 2011), such as navigational guidance, remote collaboration, or augmented reality displays. In particular, given the relative lack of unique environmental cues, location identification is essential for applications that support indoor navigation as people often lose their sense of direction. Indeed, despite the fact that Global Positioning Systems (GPS) has been widely used in outdoor navigation systems, they are not suitable for the indoor use because buildings may block satellite signals, making them unreliable, if not unusable. A number of researchers have provided new mobile methods to capitalize on the ubiquitous wireless connections for location identification (Afyouni, Ray, and Claramunt 2012; Huang and Gartner 2010). Research on remote collaboration has also identified unique challenges and solutions for location identification in different task scenarios (Billinghurst and Kato 2002; Gelb, Subramanian, and Tan 2011; Balakrishnan, Fussell, and Kiesler 2008). While research on developing autonomous systems has made significant progress, challenges remain. Some proposed that an optimal mix of computing and human agents can provide a cost-effective approach for many practical problems (Rao and Fu 2013). The main challenge for designing such a system is how one can more effectively use human computation to complement computing agents.

The Current Study

The Spatial Location Identification Task

We used a Spatial Location Identification Task (SpLIT) to test how people could make use of spatial cues. During the task, participants were required to identify the location of where a picture was taken in a two-dimension floor map. Figure 1 shows the experimental interface. On the right hand side of the screen (area B in Figure 1), a camera view of an indoor location was presented. On the left hand side of the screen (area C in Figure 1), a floor map was presented, on which the workers were asked to place a marker on the location at which they believed the camera view was taken. Throughout the experiment, the same floor map was used; but different camera views were presented.

There were a number of pre-defined markers for the floor map. These markers represented the locations of the classrooms, stairs, exits, elevators, and restrooms. In each of the camera views, a varying number of these pre-defined markers were present. Workers not only had to associate the markers between the camera view and the map, but also had to derive their spatial relations and use them to infer the location at which the camera view was taken. As we will explain later in the experimental design, the pre-defined markers in some of the camera views did not provide sufficient information to determine their exact location. In such cases, workers had to extract other spatial cues (e.g. a corner, a long corridor, the width and/or height of the space, etc.) to identify the location of the camera views. Once workers determined the location, they could drag-and-drop the solution maker to the location on the map. Before they submitted the answer, each of them were asked to provide a brief (one sentence) explanation on how they came up with the answer.

Experimental Design

In order to explore how differently these two schemes (ground truth and majority vote) would impact workers' performance in the SpLIT, we created five pictures with different numbers of pre-defined markers (stairs, classrooms, restrooms, elevators, exits, and intersections). As shown in

Table 1: Summary of the camera views designed with different combination of cues

Camera View Num.	Stairs	Classroom	Restroom	Elevator	Exit	Intersection	*Total*	*Difficulty*
1	0	0	1	1	1	1	4	1
2	0	1	0	1	0	1	3	2
3	1	0	0	0	1	0	2	3
4	0	0	0	0	0	1	1	4
5	0	0	0	0	0	0	0	5

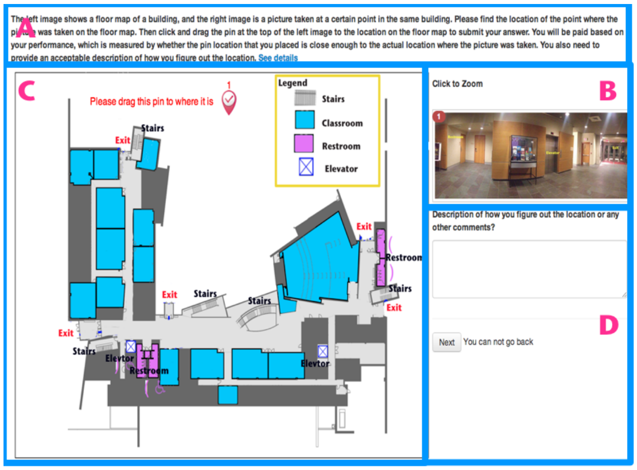

Figure 1: Experimental interfaces used in our experiment: (A) Instruction (B) Picture to identify location (C) Floor map with a pin to put where it should be (D) Textbox to enter the descriptions of how the location was identified.

Table 1, these five pictures were carefully chosen to represent different levels of ambiguity, in terms of the extent to which workers could use the pre-defined markers (cues) to infer the location of the camera view. For example, Picture 1 (see Figure 2) had 4 cues; but in Picture 5 (see Figure 3), there was no pre-defined markers.

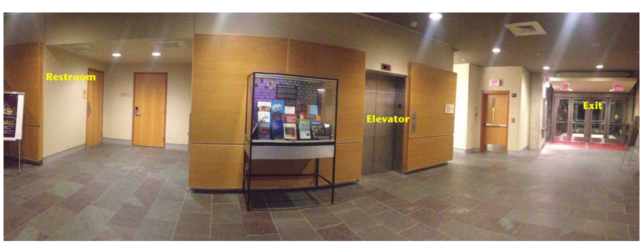

Figure 2: Camera View 1: there were four pre-defined cues that could be found also on the floor map: restroom, elevator, exit, intersection.

We created two sets of instructions for each experimental condition to explain the two reward schemes. In the *ground truth* scheme, participants were told that they would get paid if their choices were close enough to the actual location where the picture was taken; while in the *majority vote* scheme, they were told their performance was measured by

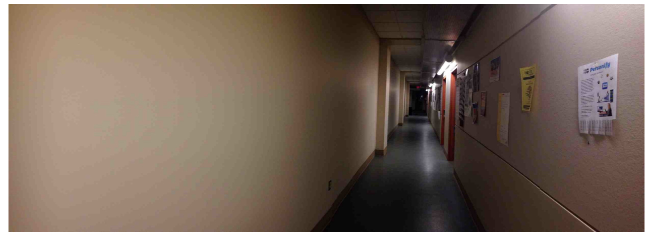

Figure 3: Camera View 5: there were no pre-defined cues that could be found also on the floor map.

Figure 4: Figure presented to workers showing them how they would be paid in ground truth scheme.

whether their choices were close enough to the majority of choices by other turkers. The exact words used in the instructions were listed in Table 2, and figures that illustrate the reward schemes (see Figure 4 and Figure 5) were also provided.

Procedure

We used the crowdsourcing platform provided by Amazon's Mechanical Turk (AMT). We published our Human Intelligence Tasks (HITs) on AMT from 4/10/2013 to 4/30/2013, each with a price tag of $0.15. Duplicate worker IDs or IP addresses would be rejected.

The workers were randomly assigned to one of the two reward schemes, and the five pictures were presented in a random order except for Picture 1, which was always presented as the first one. Given that the first picture had enough cues

Table 2: The exact instruction texts for the two reward schemes

Scheme	Instruction
Ground truth	The left image shows a floor map of a building, and the right image is a picture taken at a certain point in the same building. Please find the location of the point where the picture was taken on the floor map. Then click and drag the pin at the top of the left image to the location on the floor map to submit your answer. You will be paid based on your performance, which is measured by whether the pin location that you placed is close enough to the actual location where the picture was taken. You also need to provide an acceptable description of how you figure out the location.
Majority vote	The left image shows a floor map of a building, and the right image is a picture taken at a certain point in the same building. Please find the location of the point where the picture was taken on the floor map. Then click and drag the pin at the top of the left image to the location on the floor map to submit your answer. You will be paid based on your performance, which is measured by whether the pin location that you placed is close enough to the majority of the locations by other turkers. You also need to provide an acceptable description of how you figure out the location.

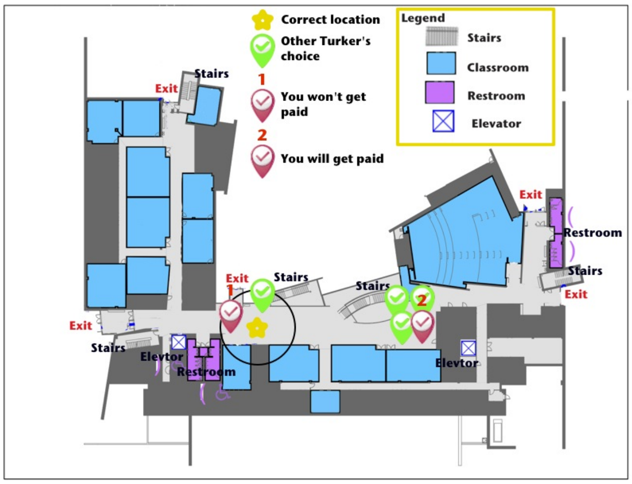

Figure 5: Figure presented to workers showing them how they would be paid in majority vote scheme.

to pinpoint the exact location of the camera, it allowed the workers to practice the task. Another purpose of the first picture was to test whether the workers understood the task, and to potentially catch spammers who did not put in the minimal amount of effort to finish the task.

As shown in Figure 1, workers would be presented with (A) the instruction (also see Table 2) as well as a link to a figure showing the details of the reward scheme (see Figure 4 and Figure 5), (B) one of the five camera views of a spot in the indoor environment (there was also a hint above the picture telling them to click on the camera view, after which the picture would be magnified to occur the whole screen), (C) the floor map with signs and a legend of the pre-defined cues, and (D) a textbox for workers to write a short description of how they did the task. The red pin at the top of the floor map beside the legend could be dragged to any location on the floor map to indicate their answers. When workers finished one task, they clicked the "Next" button to proceed to the next camera view. The system would not proceed and would pop up a warning message if it detected that workers never moved the pin or if the text box was empty. The coordinates of the pins on the map and the explanations workers provided in the textbox were recorded in our database.

Results

Pre-processing of Data Since low quality answers are common in AMT, in addition to rejecting repeated turkers by tracking their worker IDs and IP addresses, we also removed workers who had answers that were outside the building. We also removed workers whose answers were almost exactly the same for all five tasks. 103 workers were recruited from AMT, of which 50 performed SpLIT on the five pictures in the instruction scheme of *ground truth* and 53 did it in the instruction scheme of *majority vote*. After pre-processing these data according to the above criteria, there were 41 valid workers for the *ground truth* scheme and 43 for *majority vote* scheme.

Accuracy of Location Identification For each correct location of the five tasks, we manually determined an area within which the points chosen by turkers were counted as correct. The size of the area was based on the width of the space as captured by the corresponding camera view. In other words, we considered any answers on the floor map that corresponded to any of the visible space in the camera view as correct ones, whereas any answers outside the camera view would be considered incorrect.

We calculated the accuracies of each task in each scheme by dividing the total number of correct answers by the total number of answers given by the workers. To test whether the accuracies between the two reward schemes were statistically significant, we performed a paired t-test to test the difference. Because accuracies were calculated as percentages and did not often follow a normal distribution (they are limited in the range between 0 and 1), we performed a standard arcsin transform to the accuracies to correct for possible violation of normality. Results showed that the difference of the accuracies between these two schemes were significant $t(4) = 8.29, p < 0.001$.

As shown in Figure 6, accuracies drop as the level of ambiguity increases. This suggests that the lack of spatial cues did make the task harder. And it is obvious from the figure that performance in *majority vote* scheme was better than that in the *ground truth* reward scheme.

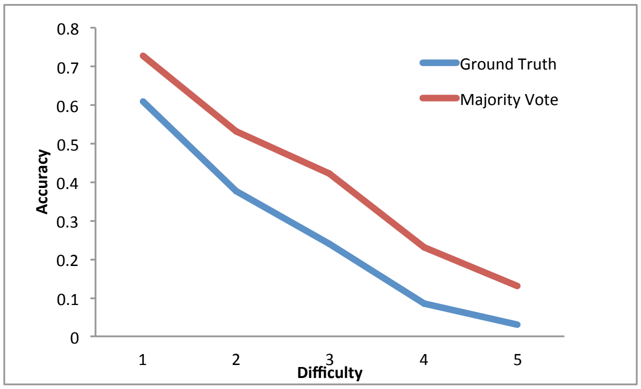

Figure 6: Accuracies of SpLIT in each of the tasks for the two reward schemes.

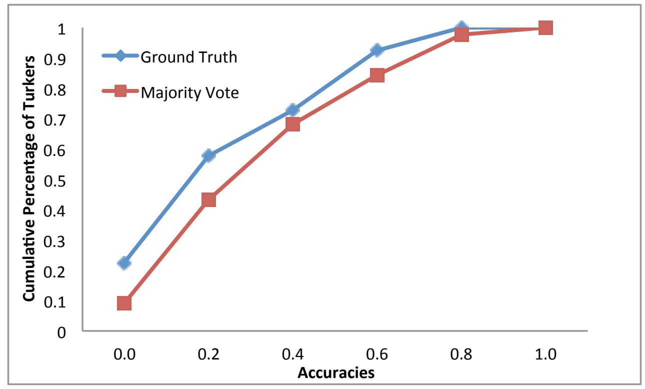

Figure 8: Cumulative histogram of the accuracies of individual workers.

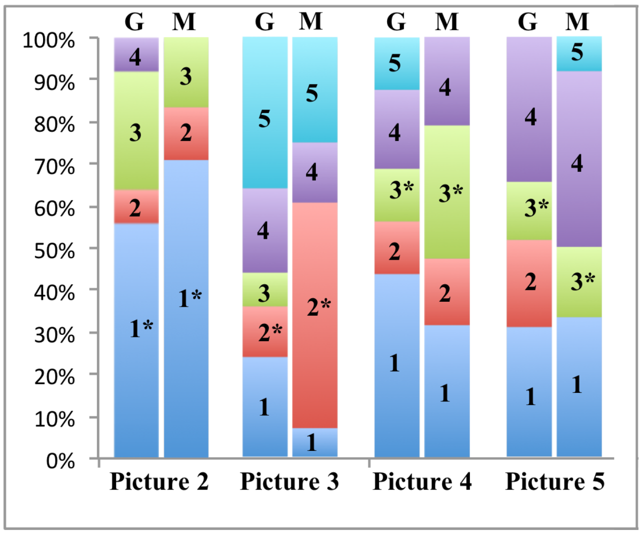

Figure 7: Percentage of points in each cluster: the asterisk indicates the correct cluster.G=ground truth scheme and M=majority vote scheme.

Figure 9: Clustering illustrations for Picture 2 in ground truth scheme: the size of the circle represents the mean inner distance of the cluster and the stroke width of the circles represents the percentage of points in that cluster.

Accuracies of Individual Participants As each participant performed the five tasks, we calculated their accuracies in terms of the percentages of these tasks completed correctly. The best workers could perform all five tasks correctly, lead to an accuracy of 1.0; while the poorest workers would perform all tasks incorrectly, leading to an accuracy of 0.0. For each reward scheme, we counted the percentages of the workers with a specific accuracy level (i.e. how many of them were correct in one, two, etc. tasks correct). The cumulative histogram for the percentages is shown in Figure 8. As it indicates, there were more turkers that at least chose one correct location in the *majority vote* scheme than in the *ground truth* one. In addition, the percentage with high accuracies (0.8 and 1.0) in the *majority vote* scheme is also higher than that in the *ground truth* one.

Cluster Analysis After visualizing the points on the floor map for each picture, we noted that most of the points were clustered around certain locations, although not all clusters were close to the correct location. There were also a small number of answers scattered all over the map, however most of them did not offer meaningful descriptions about how they had been chosen. These points (less than 10%) therefore were not included in our cluster analysis.

By applying the standard clustering method PAM (Partitioning Around Medoids), we obtained several clusters for each picture in the two schemes. To test the difference in precision in each cluster, the average distances within each clusters were calculated for each tasks for each instruction condition. Results are listed in Table 3. It shows that the answers in the *majority vote* scheme are significantly more precise than those in the *ground truth* scheme, as confirmed by the paired t-test $t(4) = 2.8885, p < 0.05$.

We also counted the number of answers in each cluster and calculated the percentages of these answers with respect to all answers in each scheme. The result is shown in Fig-

Figure 10: Clustering illustrations for Picture 2 in majority vote scheme: the size of the circle represents the mean inner distance of the cluster and the stroke width of the circles represents the percentage of points in that cluster.

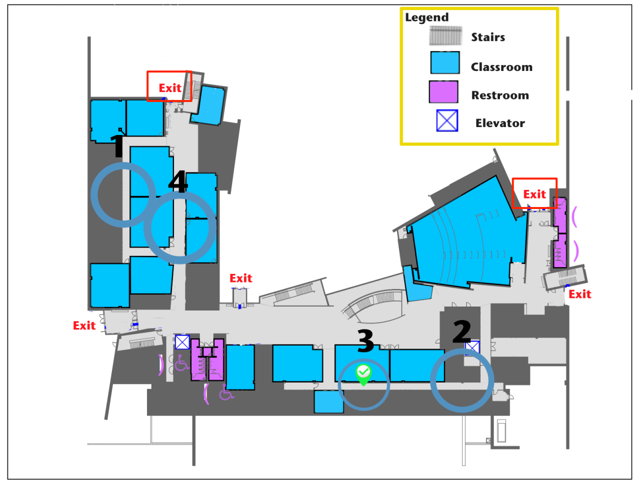

Figure 11: Clustering illustrations for Picture 5 in ground truth scheme: the size of the circle represents the mean inner distance of the cluster and the stroke width of the circles represents the percentage of points in that cluster.

Table 3: Average distances within clusters in the ground truth and majority vote reward schemes.

Pic Num.	1	2	3	4	5
Ground Truth	46.94	30.76	39.53	19.33	39.24
Majority Vote	44.03	22.51	39.71	13.13	28.44

ure 7. In the figure, percentage of each cluster identified by PAM were shown. The asterisk means that the cluster was the correct location, which we called "correct" cluster. As shown in the figure, in the *majority vote* scheme, the clusters with the largest percentage were more likely the correct ones.

To further test whether sizes of the clusters in the majority vote reward scheme were more "extremely distributed" (i.e. the largest one was much larger than the smallest one) than those in the ground truth scheme, we compared the cluster sizes of the largest clusters in each of the tasks in the majority scheme to the sizes of the corresponding clusters in the ground truth scheme. The difference in cluster sizes between the two schemes were statistically significant based on a one-tail paired t-test $t(4) = -2.4198, p < 0.05$, with the largest clusters in majority vote larger than those in ground truth. In contrast, we compared the largest clusters in each of the tasks in the ground truth scheme to those in the majority vote scheme and one-tailed t-test showed that the difference was not statistically significant. This indicated that the clusters in the *majority vote* scheme were more unevenly distributed (the largest ones were much larger than the smaller ones) than those in the *ground truth* scheme. Given that the largest clusters also tended to be the correct clusters, thus the results provided further support that the majority vote reward scheme led to more reliable results than the ground truth reward scheme. As an example, we drew the clusters as circles on the floor map for each scheme for Picture 2 in Figure 9 and Figure 10. The thickness of the boundaries of the circles represented the percentages of workers who provided answers in the clusters, while the sizes of the circles represented the average distances of the answers to the cluster centers. We can see that, generally speaking, the clusters in the scheme of majority vote had higher percentages (thicker lines) and smaller averages distances to the center (i.e. more clustered) than the corresponding clusters in the ground truth scheme.

And mentioned above, the clusters with the largest percentages were usually the correct ones in the scheme of majority vote. However, one exception was the results in Camera view 5. As seen in Figure 7, the largest cluster was cluster 4 in both schemes. Because this camera view (see Figure 3) had no pre-defined cues that could be used to identify its location, workers varied in their detection of relevant cues for location identification, as shown in the clusters on the floor map for camera view 5 (see Figure 11 and Figure 12). Some explanations provided by the workers provided hints on how they identified the locations. For example, workers reported:

> "I could see an exit symbol at a corner in the picture with that I have assumed that it must be the place."

> "I did it based on the long corridor and what looks to be an exit at the end."

Comparing the percentages of cluster 4 and other clusters between the two schemes (see Figure 7 , Figure 11 and Figure 12), we found that they were both larger in the scheme of majority vote. Cluster 4 was apparently due to the wrong cue of the wrongly interpreted"Exit" sign. In fact, it was pointing to the direction where one should turn to find the exit, but not where the exit was. However, without any other cues, many workers chose cluster 4. It was interesting to see that there were more workers in the majority vote scheme who chose cluster 4 than in the ground truth scheme. The

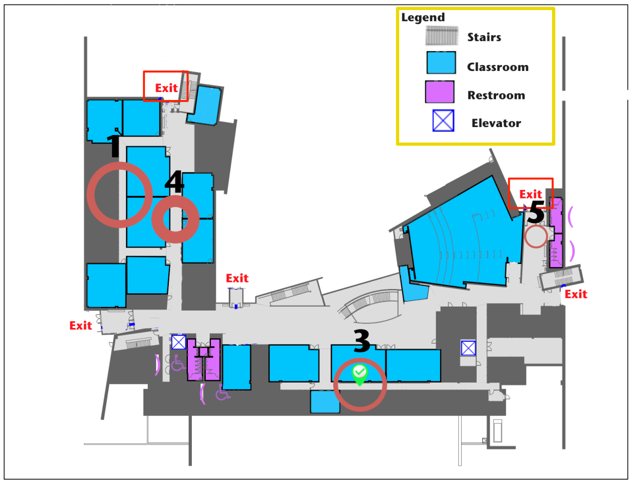

Figure 12: Clustering illustrations for Picture 5 in majority vote scheme: the size of the circle represents the mean inner distance of the cluster and the stroke width of the circles represents the percentage of points in that cluster.

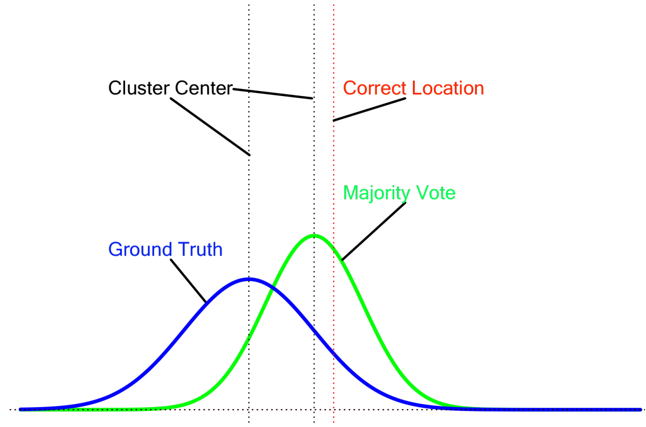

Figure 13: A notational figure showing the distributions of answers given by workers in the ground truth and majority vote schemes.

exit cue was apparently interpreted as a cue that workers believed others would notice and use to infer the location of the camera view, and thus the percentages of workers who chose the answers using this cue was higher. This also hints that focal points may not always be correct, especially when some workers believe that a wrong answer or wrong cue could be more salient than the correct answer to most other workers, they may choose the wrong one instead of the correct one. Although in our case, the majority vote scheme still outperformed the ground truth scheme, one could imagine that in certain situations the reverse could be true. Future research can focus on how to provide some forms of feedback or guidance to give workers the perception that other workers are more likely going to choose the correct answers or cues than the wrong ones, such that workers are motivated to choose the best answers that they can.

Discussions

The main effect we were interested in was the difference between the ground truth and majority vote reward scheme. As we presented in previous section, general speaking, the majority vote reward scheme led to more accurate and more reliable answers. Figure 13 shows the general results of the effect of the two schemes.

For the sake of simplicity, we assume that workers are generating answers on a line (instead of a two-dimensional map). The x-axis of Figure 13 represents the value of the answers generated by the workers, while the y-axis represents the percentages of workers who generated the answers. Assuming that answers provided by workers are generated from some probability distributions, the distributions of answers may show up as clusters of answers as in Figure 13. The peaks of the distributions represent the center of the clusters. As the result was shown that the answers given in majority vote scheme were more precise — i.e. they were closer to each other (more clustered). This is represented by the narrower distribution of the majority vote than the ground truth scheme in the figure. In addition, the result also showed that the majority vote scheme were more reliable — i.e. the percentages of workers that were in the cluster of answers that were closest to the correct answer were higher in the majority vote than the ground truth scheme. This is indicated by the taller distribution of answers that were closer to the correct answers in the majority vote than the ground truth scheme in the figure.

Although our findings showed that the majority voting scheme led to higher precision and reliability when identifying spatial locations, this experiment could not pinpoint exactly how the reward scheme could influence precision and reliability. One likely explanation was that the majority voting scheme was more successful in filtering out spammers. In other words, workers recruited by the majority reward scheme tended to be those who put more effort in the task than those recruited by the ground truth reward scheme, assuming that those who put in more effort would lead to higher precision and reliability.

Another explanation was that the majority vote reward scheme encouraged workers to reflect more on what other workers would choose, and thus led to the results we found – an effect similar to the coordination effect in Schelling's experiment (Schelling 1960).

The current experiment was limited in its ability to directly test which explanation was correct. One possible way to further test whether the first explanation is correct is to disclose the reward scheme after workers selected the task. If the majority vote reward scheme is more effective for filtering spammers, we will see a higher dropout rate. Despite of the limitation, the current findings have highlighted the nature of the differences in workers' answers in a spatial identification task induced by the reward schemes. The results have provided many possible future research directions to gain a deeper understanding of applying human computation techniques to spatial tasks.

Conclusion

We presented results from a study that tested the effects of the majority vote reward scheme on performance in the Spatial Location Identification Task (SpLIT). To summarize the results, we found that the majority vote reward scheme in general led to a higher level of precision and reliability in the answers provided by the workers. Though there were not enough data to fully understand this result, the current study is clearly a first step towards understanding how human computation can be incorporated into applications that support spatial tasks, such as navigation and remote collaboration. In addition, it does point to a promising direction of research that incorporates humans into automated methods that perform complex graphic or spatial computations. Similar to many previous attempts to incorporate human computations into system development, the current results show that it is possible to use simple methods to utilize human knowledge (in our case, general knowledge about spatial environments) to complement existing technologies.

References

Afyouni, I.; Ray, C.; and Claramunt, C. 2012. Spatial models for context-aware indoor navigation systems: A survey. *Journal of Spatial Information Science* (4):85–123.

Balakrishnan, A. D.; Fussell, S. R.; and Kiesler, S. 2008. Do visualizations improve synchronous remote collaboration? In *Proceeding of the twenty-sixth annual CHI conference*, 1227–1236. New York, New York, USA: ACM Press.

Bernstein, M. S.; Little, G.; Miller, R. C.; Hartmann, B.; Ackerman, M. S.; Karger, D. R.; Crowell, D.; and Panovich, K. 2010. Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, 313–322. New York, NY, USA: ACM.

Billinghurst, M., and Kato, H. 2002. Collaborative augmented reality. *Communications of the Acm* 45(7):64–70.

Choi, J. W.; Kwon, D. Y.; Hwang, J. E.; and Lertlakkhanakul, J. 2007. Real-time management of spatial information of design: A space-based floor plan representation of buildings. *Automation in Construction* 16(4):449–459.

Dow, S.; Kulkarni, A.; Klemmer, S.; and Hartmann, B. 2012. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 1013–1022. New York, NY, USA: ACM.

Eiben, C. B.; Siegel, J. B.; Bale, J. B.; Cooper, S.; Khatib, F.; Shen, B. W.; Players, F.; Stoddard, B. L.; Popovic, Z.; and Baker, D. 2012. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nature Biotechnology* 30(2):190–192.

Gelb, D.; Subramanian, A.; and Tan, K. H. 2011. Augmented reality for immersive remote collaboration. In *IEEE Workshop on Person-Oriented Vision (POV)*.

Hansen, R.; Wind, R.; Jensen, C. S.; and Thomsen, B. 2009. Seamless Indoor/Outdoor Positioning Handover for Location-Based Services in Streamspin. In *Mobile Data Management: Systems, Services and Middleware, 2009. MDM '09. Tenth International Conference on*, 267–272. IEEE Computer Society.

Huang, S.-W., and Fu, W.-T. 2013a. Don't hide in the crowd!: increasing social transparency between peer workers improves crowdsourcing outcomes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 621–630. New York, NY, USA: ACM.

Huang, S.-W., and Fu, W.-T. 2013b. Enhancing reliability using peer consistency evaluation in human computation. In *Proceedings of the 2013 conference on Computer supported cooperative work*, 639–648. New York, NY, USA: ACM.

Huang, H., and Gartner, G. 2010. A survey of mobile indoor navigation systems. *Cartography in Central and Eastern Europe* 305–319.

Kulkarni, A.; Can, M.; and Hartmann, B. 2012. Collaboratively crowdsourcing workflows with turkomatic. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 1003–1012. New York, NY, USA: ACM.

Lasecki, W.; Miller, C.; Sadilek, A.; Abumoussa, A.; Borrello, D.; Kushalnagar, R.; and Bigham, J. 2012. Real-time captioning by groups of non-experts. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, 23–34. New York, NY, USA: ACM.

Law, E., and von Ahn, L. 2011. Human Computation, 1st edition. *Human Computation, 1st edition* 1–123.

Mulloni, A.; Seichter, H.; and Schmalstieg, D. 2011. Handheld augmented reality indoor navigation with activity-based instructions. In *the 13th International Conference*, 211. New York, New York, USA: ACM Press.

Rao, H., and Fu, W. T. 2013. A General Framework for a Collaborative Mobile Indoor Navigation Assistance System. *LAMDa'13*.

Schelling, T. C. 1960. The strategy of conflict. *Cambridge*.

Shaw, A. D.; Horton, J. J.; and Chen, D. L. 2011. Designing incentives for inexpert human raters. In *CSCW'11: Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 275–284. New York, USA: ACM Press.

Sorokin, A., and Forsyth, D. 2008. Utility data annotation with amazon mechanical turk. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, 1–8.

von Ahn, L., and Dabbish, L. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 319–326. New York, NY, USA: ACM.

Zamir, A. R., and Shah, M. 2010. Accurate image localization based on google maps street view. In *Computer Vision–ECCV 2010*. Springer. 255–268.

Zhang, J.; Hallquist, A.; Liang, E.; and Zakhor, A. 2011. Location-based image retrieval for urban environments. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 3677–3680.