Multimedia Systems manuscript No.
(will be inserted by the editor)

Drawing on Mobile Crowds via Social Media

Case UbiAsk: Image Based Mobile Social Search Across Languages

Yefeng Liu · Vili Lehdonvirta · Todorka Alexandrova · Tatsuo Nakajima

Received: date / Accepted: date

Abstract Recent years have witnessed the impact of crowdsourcing model, social media, and pervasive computing. We believe that the more significant impact is latent in the convergence of these ideas on the mobile platform. In this paper, we introduce a mobile crowdsourcing platform that is built on top of social media. A mobile crowdsourcing application called UbiAsk is presented as one study case. UbiAsk is designed for assisting foreign visitors by involving the local crowd to answer their image-based questions at hand in a timely fashion. Existing social media platforms are used to rapidly allocate microtasks to a wide network of local residents. The resulting data are visualized using a mapping tool as well as augmented reality (AR) technology, result in a visual information pool for public use. We ran a controlled field experiment in Japan for 6 weeks with 55 participants. The results demonstrated a reliable performance on response speed and response quantity: half of the requests were answered within 10 minutes, 75% of requests were answered within 30 minutes, and on average every request had 4.2 answers. Especially in the afternoon, evening and night, nearly 88% requests were answered in average approximately 10 minutes, with more than 4 answers per request. In terms of participation motivation, we found the top active crowdworkers were more driven by intrinsic motivations rather than any of the extrinsic incentives (game-based incentives and social incentives) we designed.

Keywords Mobile crowdsourcing · Mobile human computation · Mobile social search · Incentive mechanisms · Mobile image translation · Mobile Q&A

Yefeng Liu (✉) · Todorka Alexandrova · Tatsuo Nakajima
63-505 Faculty of Science and Engineering, Waseda University
3-4-1 Okubo Shinjuku-ku, Tokyo 169-8555, JAPAN
Tel.: + 81 3 3207-6812
Fax: +81 3 3207-6812
E-mail: yefeng@dcl.info.waseda.ac.jp

Todorka Alexandrova
E-mail: alexandrova@aoni.waseda.jp

Tatsuo Nakajima
E-mail: tatsuo@dcl.info.waseda.ac.jp

Vili Lehdonvirta
Helsinki Institute for Information Technology
PO Box 19800, FIN-00076 Aalto, Finland
E-mail: vili.lehdonvirta@hiit.fi

1 Introduction

Nowadays, personal gadgets and other handset devices communicate wirelessly with each other and also with the various social media services in the cloud. Thus resulting in a mobile and social computing infrastructure that will benefit not only the users but also the technology providers by engaging more potential content contributors. The purpose of this line of research is to examine how such mobile and social computing infrastructure could be used to bring the new kinds of human computing or crowdsourcing model into a mobile context.

Crowdsourcing [16] is a recent term that describes the act of outsourcing tasks, which are traditionally performed by an employee or contractor, to a large group of the Internet population (the wise crowd) by means of an open call. The tasks are typically ones that humans are good at but machines are not, such as annotating pictures, recognizing images, or ranking search results. In the human-computer interaction (HCI) domain, crowdsourcing is also known as *Human Computation* (hcomp) [31], which is understood as the notion of solving difficult computational problems through human computing effort instead of machine algorithms. This notion is based on the fact that many cognitive tasks that are easy for humans remain extremely difficult for computers to perform.

Mixing mobile platforms and the crowdsourcing model potentially offers vast resources for computation. For instance, today, in most urban areas it is common that people allot a large amount of time for commuting or waiting for various events. Their time is actually fragmented into numerous small pieces of time and most of them are occupied with meaningless activities. We believe that, in this case, the crowdsourcing model provides a win-win solution for better use of this time by engaging people through their networked and ubiquitous mobile devices. However, a majority of current human computation and crowdsourcing systems (e.g., Amazon Mechanical Turk[1]) are passive services that are using worker-pull strategy to allocate tasks, and require relatively complex operation to create a new task. Consequently such systems fail to adapt to the mobile context where users require simple input process and rapid response.

We explore crowdsourcing platform designed for mobile users. Our design principle is simple: build the system around existing social media platforms — the popular services that inherently have the culture of sharing and participation, and are already well known and used by a large and growing number of users. In this paper, we practice the platform in the context of one specific application area – image-based mobile translation/search. This refers to camera phone applications that attempt to solve the problem of translating text written in an unfamiliar script. This kind of a system is particularly useful for travellers and short-term residents in a foreign country. Ordinary digital pocket translators are useless if the user is unable to input the text they see (e.g., Japanese or Chinese text for Western visitors). Image-based mobile translation/search systems typically employ Optical Character Recognition (OCR) algorithms to extract text out from image, and apply Machine Translation (MT) technologies to translate the text. Several systems like this have been proposed during the past decade [28,15,26]. However, only a handful of them have gone beyond pilot development. Even the state-of-the-art in this field, such as Word Lens[2] or Google Goggles[3], demonstrate very limited performance in real-world situations (i.e., complex background, dark environment, blurred photos, irregular fonts, size or formats, etc.), especially for non-Latin scripts like Japanese and Chinese.

To solve the problem, we present *UbiAsk*, a social media crowdsourcing application built on top of existing social networking infrastructure. UbiAsk provides translation services and situational advice to mobile users in unfamiliar environments. Instead of applying machine algorithms, we draw on the power of ordinary people in the cloud via social networks to solve the difficult computational problems such as image recognition and text translation. Since the workload of each task in the image based mobile translation/search service is lightweight enough to be described as a micro-task, the tasks are perfectly suitable to be distributed to large groups of casual workers.

In UbiAsk, users can issue requests via several channels that use a common API. Native mobile applications and email are the currently implemented channels. The requested task is pushed to a community of voluntary local experts in the form of an open call via different social media platforms (Twitter[4], Facebook[5], etc.) and email. The crowdsourced result data is not only returned to requesters but also visualized on location based social mapping and augmented reality (AR) platforms (Sekai Camera[6] and Ushahidi[7], etc.). This gradually results in an information pool that constitutes a public good.

To evaluate the UbiAsk system, we conducted a controlled, between-groups field experiment for the duration of six weeks (from late January 2011) with 55 participants. 19 participants were foreign visitors in Japan, the majority from North America and Europe. They served as requesters. 36 participants were Japan-based Japanese/English speakers, who served as local experts. In this evaluation, the main focus was on response speed and quantity. Quality of the answers and how to assure it was mainly left for the next stage of the project, although requesters were asked to evaluate the overall quality of the answers. In terms of user engagement, we also investigated how the crowdworker participation is affected by different incentive mechanisms. Our results show that:

- More than 90% of the questions got at least one answer. On average every request received 4.2 answers. The main causes of unanswered requests were bad timing (e.g., local experts were busy) and boring question's content (e.g., translation or explanation of a long text).
- Nearly 50% of the requests were answered within 10 minutes, and 75% of requests were answered within 30 minutes.
- In the afternoon, evening and night[8], nearly 88% requests were answered in average approximately 10 minutes.
- The most frequent participants were more motived by intrinsic incentives rather than extrinsic incentives we provided.
- For the less self-motivated users, the effectiveness of the designed extrinsic incentives (game-based incentives and social incentives) was verified. However, based on the results we have it is hard to come to the conclusion

[1] https://www.mturk.com Last checked: February 2011
[2] http://questvisual.com/ Last checked: March 2011
[3] http://www.google.com/mobile/goggles/ Last checked: March 2011
[4] http://www.twitter.com/ Last checked: March 2011
[5] http://www.facebook.com/ Last checked: March 2011
[6] http://sekaicamera.com/ Last checked: February 2011
[7] http://www.ushahidi.com/ Last checked: February 2011
[8] That is, from 12:00 to 2:00

that the proposed game-based incentive has a greater impact than the social psychological incentives.

In this paper, we introduce the user-centered design of the UbiAsk service and the system architecture, as one case study of a mobile crowdsourcing platform. We describe a field user study in Japan and report the results of the study on the overall performance (i.e., response time and response quantity) and how different incentive mechanism may affect crowdworker's participation, and we discuss some interesting findings.

In the following, we first give the background overview of the research in Section 2. Then we describe the design iterate of the proposed mobile crowdsourcing platform, and the pilot user study of the design research in Section 3. In Section 4, a controlled user study, and the results are described. The future directions are given in Section 5. Finally, in Section 6, the conclusion and a number of interesting research findings are discussed.

2 Background

2.1 Mobile Image Search and Translation

There has been a number of image-text translation systems proposed over the past decades (see [14] for an overview). Most of them applied ORC technologies to recognize images and then use machine translation technologies to translate the text into desired language. Masashi Koga et al. [19] discussed a camera based mobile image translation application using Kanji OCR. Their main target source text is machine-printed documents. This study suggests that users are also interested in LED displays and other non-printed texts, and more important, in deeper contextual information and advice as opposed to merely the literal meaning of a word or character. The latter is especially important when the cultural distance between the source and target languages is great. Other relevant studies can be found in the information augmentation system research stream like *InfoScope* [15]. InfoScope was a mobile application for camera phones; it captures images from real world, extracts information (e.g., text in Chinese) for the image, processes (e.g., translation) the information in the digital world, and augments the processed information (e.g., text in English) back to the original scene location via display. Some latest smartphone applications like *Word Lens* and *Goggle Goggles* is putting Infoscope's information augmentation concept into practice. Word Lens is a mobile application that is able to translate Spanish or English phrases instantly on the screen when you point your mobile device's camera at the text. Google Goggles, on the other hand, is a mobile visual search tool that support photo based searching for text, logos, books, landmarks, and so forth. It compares the objects in a photo took by mobile against the items it can find in its image database, and return the ordered results.

However, due to the computational difficulties of OCR and translation, the machine-based technologies still provides very limited performance in the real-world conditions such as complex background, dark environment, blurred photos, irregular fonts, size or formats, etc.

2.2 Mobile Crowdsourcing

Crowdsourcing and human computation are approaches that seek to use human intelligence to overcome the limitations of computing technology. Extending this model to mobile devices and users can potentially further enhance the availability of contributors. Eagle N. is one of the pioneers of promoting mobile crowdsourcing systems. He built a system called *txteagle* [12], and first deployed it in Kenya in 2009 and, in some other poorest parts of the world afterwards. The txteagle system distributes micro tasks to mobile phone users from the developing countries via mobile phone text messages (SMS) or audio clips, and provides a small amount of money as a reward for each task. The typical tasks in txteagle system include software localization, filling out surveys, and rating the local relevance of search results. On the job provider end, txteagle makes the tasks more economical; on the other end, it also offers a welcome source of income for the participants. T. Yan et al. proposed a crowdsourcing-based approach to improve the quality of real-time image search on mobile phones in [32]. Their system combines automated mobile image search with the real-time human validation of search results. Crowdworkers from Amazon Mechanical Turk perform the validation tasks. Moreover, *CrowdSearch* algorithms were proposed to optimize for the delay and the money constraints. Their experiment reported a result of 95% search precision for several categories of images. *Ushahidi* platform, a Google map based mash up tool to visualize crowdsourced information by letting participants submit information trough text messaging using mobile phones, emails or the Web, is another successful mobile crowdsourcing initiative. After Haiti earthquake on January 2010, people and organizations in Haiti posted thousands of their needs and requests on the Ushahidi Haiti map; volunteers who have the ability to answer the request then picked up the requests.

2.3 Social Search

Typical crowdsourcing and human computation systems treat contributors as exchangeable "cogs in the machine". In contrast, services that can be categorized as social search systems seek to leverage each contributors unique skills and

knowledge. They help users identify and connect with relevant experts who can offer solutions in a timely and human manner. Craigg Macdonald and Ladh Ounis proposed a data fusion based approach in [23] for predicting and ranking candidate expertise with regards to a given question. The effectiveness of using adapted data fusion techniques was demonstrated by their evaluation results. *Aardvark* [9] is currently one of most popular community-based social search engines. Aardvard users ask questions similar to other on-line question and answer sits, but the Aardvard server "routes the question to the person in the requesters extended social network most likely to be able to answer the question".

2.4 Incentivizing Contributions

A major challenge for crowdsourcing as well as social search services is how to motivate individuals to contribute work. Previous studies in social and computer science have identified a list of approaches to motivate people in on-line system [6,11]. In this section, we introduce the background of two of them: the *social* incentives and the *game-based* incentives.

We do not discuss *monetary* incentives here because in the particular situational context of question answering, the non-monetary motivations were represented successfully in mega examples like Yahoo!Answers[9] and Answers.com[10]. Moreover, once with money being involved, quality control becomes a major issue due to the anonymous and distributed nature of crowdworkers. Although the quantity of work performed by participants can be increased, the quality cannot, crowdworkers may tend to cheat the system in order to increase their overall rate of pay. Another drawback with economic incentives is that they can destroy pre-existing intrinsic motivations in a process known as *"crowding out"* [10], so they are better used when other motivations are not likely to exist. Additionally, for micro tasks the performance difference between paid mechanism and free mechanism was significant less than complex tasks [13].

2.4.1 Social Psychological Incentives

One widely harnessed set of non-monetary approaches to promoting increased contributions in digital services can be found in the literature of social psychology. Social facilitation and social loafing are perhaps the most commonly cited effects. *Social facilitation* effect [33] refers to the tendency of people perform better on simple tasks while under someone else's watching, rather than while they are alone or when they are working alongside other people. On the other hand, the *social loafing* effect [5] is the phenomenon of people making less effort to achieve a goal when they work in a group than when they work alone, since they feel their contributions do not count, are not evaluated or valued. This is seen as one of the main reasons group are less productive than the combined performance of members working alone.

Ways of taking the advantage of the positive social facilitation and avoiding the negative social loafing in on-line data collection systems were suggested in [3]: individuals' efforts should be prominently displayed, individuals should know that others can easily evaluate their work, and the unique value of each individual's contribution should be emphasized. Cheshire, C. et al. [8] conducted a series of quantitative field experiments to examine the effects of social psychological incentives and their results demonstrated social psychological incentives like historical reminders of past behavior or ranking of contributions can significantly increase repeat contributions.

2.4.2 Game-based Incentives

More recently, digital designers have begun to adopt ideas from game design to seek to incentivize desirable user behaviors. The idea of taking entertaining and engaging elements from computer games and using them to incentivize participation in other contexts is increasingly studied in a variety of fields. In education, the approach is known as *serious game* [35] and in human computing it is sometimes called *games with a purpose* [30]. Most recently, digital marketing and social media practitioners have adopted this approach under the term *gamification* [34]. The idea is to make a task entertaining, like an on-line game, thus making it possible to engage people to conscientiously perform tasks. The valuable output data itself is actually generated as a byproduct by the game. However, on the other hand, the difficulty of designing such a game is also a well-known problem. In many situations, the tasks can be too boring or complicated to turn into any game that is actually enjoyable or fun to play.

The ESP game [1] is one of the most famous examples of this kind. Players are paired in the game and they need to give relevant descriptions for a given image. If the description matched with other player's answer, players win and score the points, otherwise lose. The real purpose of the game is to rapidly collect annotations for a large number of images. Thereafter, various games have been devised in the style of ESP game. Arase et al. [4] proposed a web-based multi-player game to collect knowledge on the geographical relevance of images, in order to better represent certain images' geographical context for searching and browsing. Other than the casual games, Markus Krause et al. [20] implemented a relatively complex action game *OnToGalaxy* in the context of human computation. In their design, the task

[9] http://answers.yahoo.com Last checked: March 2011
[10] http://answers.com Last checked March 2011

is hidden in the game play that it is no longer perceived as a dominant element of the game.

In the following sections, we describe a mobile crowdsourcing application dubbed as UbiAsk, which provides (semi) real-time social and image search functionality and utilizes non-monetary incentive mechanisms to motivate crowd workers.

3 System Design and Implementation

In this section we present the design of the UbiAsk platform. Parallel to the technical development we have studied the application possibilities of UbiAsk by conducting a formative user study in the form of proof-of-concept simulation. Future UbiAsk services have been illustrated to the potential users. The users' feedback has been analyzed to identity the users' requirements to the UbiAsk system. Thus, we have been able to influence the technical developments well before the actual application development stage. Such a *user-centered* design process ensured that the forthcoming platform could support the features and characteristics needed by the end users.

3.1 Initial Design

The platform makes use of client-server architecture. There are two different types of clients: the users who make requests are called *requesters*, and the crowdworkers are named *local experts*. This feature differentiates UbiAsk from other typical server-client systems. The UbiAsk server plays a proxy role. It receives requests from client users, assigns these tasks to appropriate local experts, and finally forwards the answers to the original requesters.

Figure 1 illustrates the basic work-flow of the proposed translation model and a detailed description of it is given below:

- A client user makes a request by taking a picture using a mobile phone's camera, and submits the image to the server. Additionally, a short text message should be attached to the photo in order to clarify what exactly the requester wants to know. This short message extends the application possibilities, since users can now ask questions with no direct relationship with any object in the image. Location information and time will be automatically attached to the request, although the availability of such context information may depend on the client's terminal's functionality (e.g., if the locationing module is embedded). Such context information, together with a work user's profile, is useful for assigning tasks appropriately.
- For enhancing the response time, each task is assigned to multiple local experts simultaneously.

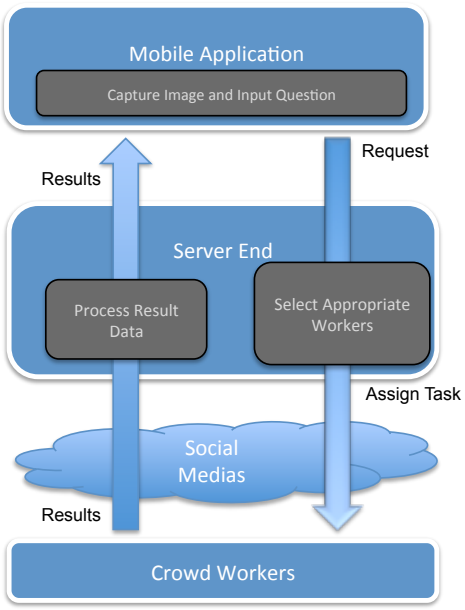


Fig. 1 System Basic Work-flow

- The original request is sent to translators via email. Translators are encouraged to reply in "key words" or "short message" style.
- For the response time considerations, the first answer to be received from local experts is forwarded to the requester immediately. For the rest of the replies, the client user can set a maximum waiting period and reject any responses received after that.
- Eventually, the client user receives the results.

3.2 Formative Experiment

For this platform to work, we have to know if the translators are able to deliver the desired responses to the requesters. Therefore, we designed a qualitative experiment in laboratory conditions to answer the research question of *how should the user's questions be presented to the local expert in order to provide the preferred results for the user?* We held a series of meetings for discussing the experiment design. Participants included Japanese and foreigners (that can be seen as our potential users) from different background areas such as technology, design, economics, user experience, and psychology. The original intention of the project has been to design a human based image translator. However, through discussions with potential users we found out that what their requirements are more than a simple translator. Instead of just knowing the semantic meaning of the words, users are much more interested in a service that can answer their questions related to the photo, which is more like an image based mobile social search across languages and cul-

tures. Based on this finding, in the later meetings we have improved our concept idea and extended the design from simple translation to a mobile social search service. Eventually the experiment work-flow has been realized as follows.

Q: ***What divination result I got?***

A1: the oracle says your fortune is 'bad luck' Knock on wood nearby as soon as you can.
A2: bad fortune :(there is a description at the bottom of the picture. But it takes long time to translate :(

Fig. 2 One Sample of Formative Experiment

First, we collected around one hundred sample images (took by mobile phone) with corresponding questions from five foreigners who were currently staying in Tokyo. Together with end users we selected fifteen characteristic cases from different types of requesters (or we could also say, from users who had different needs). Then, we interviewed the photos' providers, questioning in what situation they took the photo and what kind of answers they were expecting. In the next stage we sent the requests to a group of seven invited local experts (all male, mostly students from School of Science and Engineering, Waseda University, Japan). After receiving the answers, we interviewed the translators for collecting their feedback on the usage of the service. Finally, we compared the results from the translators with the expected answers of the requesters, examined whether they matched and discussed the possible reasons for mismatching.

Figure 2 shows one example that was observed in the pilot study. From the collected questions we found out that in many cases (i.e., approximately 47%) requests were actually driven from curiosity rather than real problems or troubles the visitors were facing (i.e., approximately 53%). A typical situation is when the users obtain partial information for something interesting but are unable to figure out the whole information that they want to know. For example, this requester knows what this piece of paper is, but cannot understand the exact information in the content.

3.3 Impressions of Pilot study

The quality of the results in crowdsourcing is a hot issue, and has been discussed widely recently [21,17]. In this study, we found that the way to input requests and answers is an important factor, which affects not only the usability but also the quality of the outcome. The simplest way of making a request is just sending the picture directly, but it can hardly be an option because translators can be easily confused about the real purpose of the request. In other words, client users must clarify what exactly they want to ask by adding more information. On the other hand, the translators often use English as their second or third language and thus, they may misunderstand the question if the text is too complicated.

There are different ways to lower the possibility of misunderstanding. One way is to limit the complexity of the messages by e.g., instructing a client user, setting maximum size of message, etc. Moreover, depending on the question, the local experts may need to reply in different ways in addition to the text, e.g., for questions like *"which button should I click?"* or *"which one will you recommend?"*, it is better to provide them an image editor thus they can simply circle the corresponding part in the picture rather than giving a description by words. However, a local expert (as a human) always makes mistakes no matter how perfect the instructions are. In the study we observed that sometimes even if the question was clear and simple enough, "bad" replies still appeared due to the fact that some translators began answering before finishing reading the request message. Even worse, we also have to consider the possible existing of a malicious reply. As single reply can hardly be trusted, another possibility is to provide multiple results to the client user. The users can compare the different replies by themselves, and make their own decision (e.g., choose the majority answer). Nevertheless, if we consider the response time, this approach might be expensive.

There is a third solution, which can be seen as a compromise between the above-mentioned methods. We can add a proofreader (as Figure 3 shows) to verify the correctness of the answer and to prevent from malicious replies. Moreover, the task of classifying/tagging images can also be assigned to the proofreader, for maintaining a more valuable results database. Depending on the different clients types, there is a trade-off between the accuracy and the response time of the reply. For requests that need immediate answer, timeliness is the key factor concerning clients' *quality of experiences*. We may want to skip intermediate stages and directly forward the answer from the first translator to the user. On the other

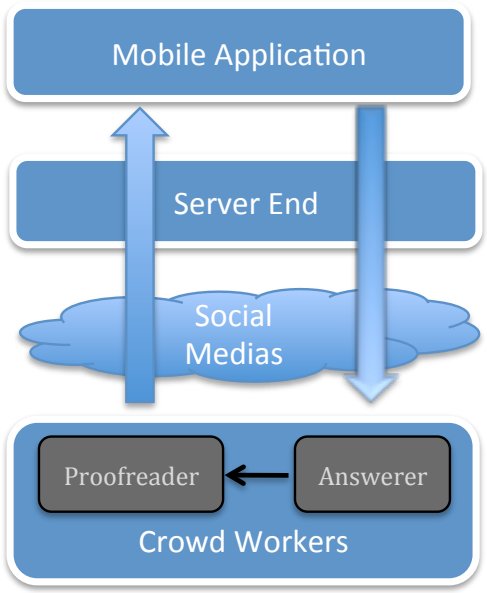


Fig. 3 System Basic Workflow with Additional Proofreading Phase

hand, for waitable type of requests, the proofreading or multiple answers should be a strict requirement. In general, we advocate the appropriate use of different request processing strategies, depending on the users' request types.

Through the experiments, we also confirmed the inherent disadvantage of the computer translation. Even if we assume the machine based image recognition and text translation technologies can always provide perfect outcome, they still have no chance of offering the desired answer for such kind of services that demand higher-level information.

3.4 Prototype Implementation

After several rounds of design iteration including formative evaluation with potential users in autumn 2010 [22], we implemented a prototype version of UbiAsk. Figure 4 illustrates the basic system structure of UbiAsk. Requesters can quick create a task and upload to server by utilizing diverse mobile applications. Proxy server pushes task to appropriate workers via regular emails as well as different social media platforms. When a task accomplished, we argue that the question-answers pairs are not only valuable for original requester but also be able to beneficial to the broader public, thereby, the location-based result data is also visualized with social mapping tool or social augmented reality services.

The proxy server was built as a REST style web service with open API to clients. A task assigner was developed as the component with responsibility for assigning task to and receiving answers from local experts. The task assigner also connects to the social media platforms APIs. All back-end programs are implemented in Java. Requester can use an

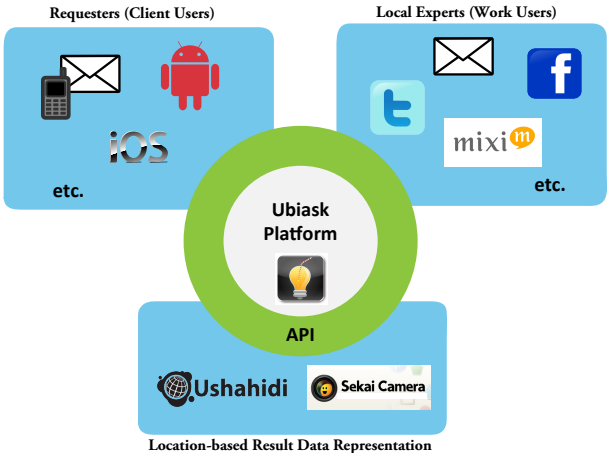


Fig. 4 UbiAsk System Overview

iPhone application (see figure 6) or mobile email to make a request. The main interface of the iPhone application consists of three main parts: an interface to show a list of existing requests, an camera display view to take picture of what requester want to ask, and a text editor to input the short question. The uploaded image was saved at the server end. The link of photo and corresponding question were pushed to local experts via email and Twitter. The local experts can submit their answer by simply replying the email or tweet. The maximum living period of a request was 12 hours and server will reject any later answers. All result data were visualized on an Ushahidi based interactive map (see Figure 7).

Fig. 6 iPhone Application for Requesters

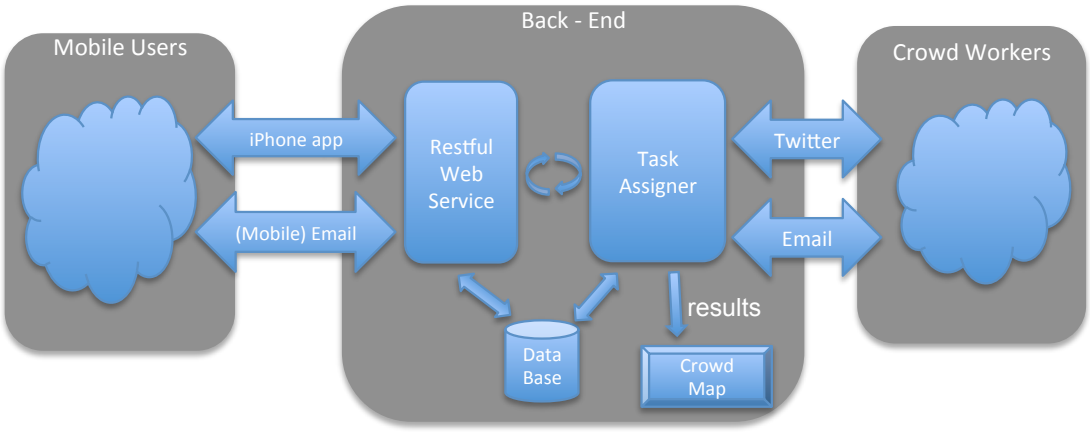


Fig. 5 UbiAsk Prototype Architectural

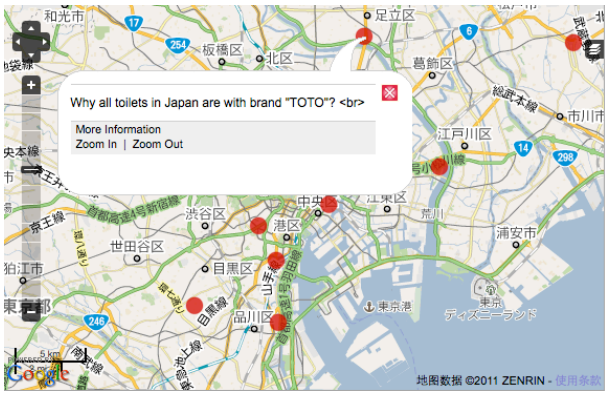


Fig. 7 Ushahidi Based Interactive Map for Visualizing Results

4 Field User Study

After UbiAsk was implemented, we launched a field user study with 55 end-users (19 requesters and 36 local experts) for six weeks[11]. On the requester end, the user study is designed to find out if UbiAsk could provide answers of satisfactory speed and quantity; on the crowdworker end, the relevance of the activating participation to different motivational mechanisms is examined. Also, the usage data is gathered to study the typical usage pattern.

4.1 Participants

The 19 foreign requesters were mainly recruited from a total of thirteen popular international travel forums such as Lonely Planet thorn tree[12], Tripadvisor[13] , Fodor's Travel Talk[14], etc. The majority of the requesters were short-term travelers, the exceptions included Japanese returnees, foreign students, foreign employees, foreign housewife and visiting scholars. All, but one, of them were from Western countries (i.e., United States, European countries, and Australia), the only exception was from Southeast Asia. Of the 36 local experts, 17 were recruited from Internet by twitter adverts or posts on local forums such as Yahoo!Chiebukuro[15] (The Japanese version of Yahoo!Answer) - English questions. The rest of the local experts were mainly undergraduate students, graduate students, and staffs from Waseda University that were recruited by emails.

However, truly active participants in the numbers listed above were significantly less. Initially 23 travelers claimed they were interested in the study and agreed to participate, however we lost contact with 4 of them before starting the user study, and 8 of them never used the service during their visiting period. We re-contacted the missing requesters during the experiment in order to identify the reasons of their absences. Not so surprisingly, most of them said it was because overseas 3G/4G usage fee is too expensive, and they could not find any free Wi-Fi spot when they want to use the service [footnote]. Although alternatively it was possible to save the photo and send it later when there was Wi-Fi available (e.g., in their hotels), the users did not do so. Thirteen local expert testers were likewise the *"lookyloos"*, who never reply any question during the user study. One possible interpretation is the experiment duration happened to be during the exam season for the Japanese schools and the student participants might have been caught up in examination preparation and final reports writing.

Overall, this user study reported 37.5% *no show rate*, which was considerably higher than the quantitative user studys average number of 11% [27]. On the other hand,

[11] From middle of January 2011 to End of February 2011
[12] http://www.lonelyplanet.com/thorntree/ Last checked: February 2011
[13] http://www.tripadvisor.com/ Last checked: February 2011
[14] http://www.fodors.com/community/ Last checked: February 2011
[15] http://chiebukuro.yahoo.co.jp/ Last checked: March 2011

however, considering the nature of on-line system, it is also quite normal to have a power law distribution, with only a small minority contributing. For instance, E. Adar and B.A. Huberman [2] reported in the P2P music sharing service Gnutella, two-thirds of users share no music files and ten percent provide 87% of all the music. In the open source software community, 4% of participants likewise contribute 88% percent of the new code and 66% percent of the bug fixes [24].

4.2 Procedure

For collecting demographic information, all participants completed a pre-test questionnaire before the requesters left their countries for Japan. The iPhone application with the usage instructions and the information of the user study were emailed to the requesters before their departures. When requesters arrived in Japan, they were told to use the service freely.

4.2.1 Overall Performance

In the pre-test questionnaire two important questions were asked: the users *expected* response time, and the users *bearable* response time. Based on the answers we can further identify three time periods with regards to user satisfaction: we assume a user will be satisfied if the first answer comes before his/her expected response time; a user will be unsatisfied if the first answer comes after his/her bearable response time; and it is acceptable if the first answer come between the two time points (see Figure 8 for an example). The survey result is shown in Figure 9, if the response time is less than 10 minutes, we can satisfy all users. It is beyond our expectations that even if it takes 30 minutes to have the response, only 25% of user will be unsatisfied. But if the waiting time becomes more than one hour, it will be unacceptable for most of the users.

4.2.2 Incentive Mechanisms Comparison

In terms of participation motivation mechanism design, local experts were randomly placed in three incentive groups, among which there were two experimental groups and one control group.

Group A - derived from the GWAP concept, a location based mobile social game was designed and implemented as participation incentive. The main interface of the game (see Figure 10) was a Google map based real world map, which was divided into non-overlapping hexagons. The goal of the game is to conquer territories. Every hexagon has one owner, who is the player with the highest number of "task-done" in the area. In other words, crowdworkers need to compete with each other to get the "Local Expert" title of an actual location on the map.

Group B - a simple feedback mechanism - social psychological incentive that was known to be effective - was applied: when a local expert provided an answer, the system will rapidly reply him the number of existing answers of that question and previous answers' content.

Group C - control group, no additional motivational method was applied.

Fig. 10 the Interface of the Location Based Ranking Game

For performance measures, we instrumented our prototype to log the timestamp and the question/replies ratio. To measure user satisfaction, ease of use, and overall experiences, we administer a post-test questionnaire with Likert scales. The questions covered typical usage patterns, content's quality, features preferences, likes, dislikes, and suggestions.

4.3 Study Results

In this subsection we present the main findings from the user study.

4.3.1 Performance Results

In all, the system recorded 180 interactions, covering 33 questions and 147 answers. We expected to see a much bigger number of requests. However eventually there were only 11 (58%) requesters that submitted their questions to the system. On the other end, 23 (64%) local experts answered at least once question. Of the 147 answers, the local experts that recruited from Internet provided 93 of them. The seven most active local experts accounted for nearly 70% of the answers. The requests are relatively equally distributed across the day expect early in the morning and mid-night.

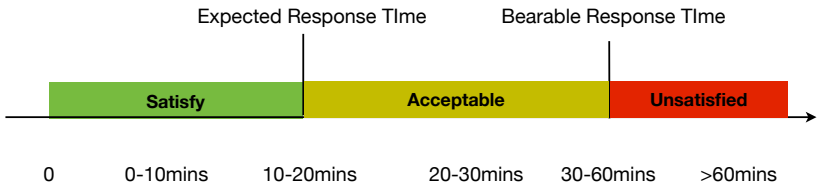


Fig. 8 Sample of One Requester's Requirements on Response Time

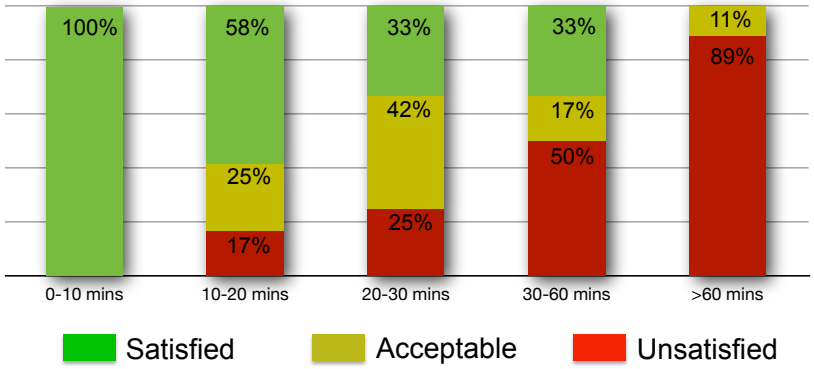


Fig. 9 Requesters' Overall Requirements on Response Time Results

Figure 11 shows the overview of the response time (the first answer). Approximately 50% of the questions were answered within 10 minutes, and 27% of them were responded within just 5 minutes. Three-quarters of the first answers arrived within half an hour. Only 6% of the answers were arrived after 60 minutes, and most of them were asked during midnight or in the morning. The 9% of the requests were never being answered, the main reasons were bad timing (e.g., local experts were busy) and boring question's content (e.g., translation or explanation of a long text).

According to the pre-test questionnaires result, we set a strict rule for the response time: if the first answer of a question arrives after 60 minutes, we consider this question has no answers. Based on this rule, we show a time-of-the-day breakdown of the average response speed, average response rate, and answers per request in Figure 12. Generally speaking, the overall performance is reliable: except midnight and early morning, nearly 88% of the requests could be answered in approximately 10 minutes, with more the 4 answers per request.

It is true that results data were collected under an experiment setting and the end-users may behave differently if the service was deployed in the real world environment. But it is also necessary to point out that the system was intentionally designed without any additional incentive mechanism due to the fact that the incentives comparison is one of our research questions. A number of the fundamental components of the Q&A systems motivation module were not provided, e.g., global view of all questions status, contribution histories, others contributions, etc. Additionally, the above-mentioned no-show-rate also indicates, this experiment was not intended to access the quality of data that can be achieved in optimal (strictly coordinated) conditions.

4.3.2 Incentives Comparison Results

The incentives comparison results were more complicated. Figure 13-I and Figure 13-II shows the response time and number of answers by groups. Local experts from Group A demonstrated a much more active engagement of the system, 67% of the first answers were provided by them. In terms of quantity, more than half of the total number of answers was likewise produced by people from group A. In the meanwhile, the control group produced a better outcome than group B. However, the web access logs did not support these findings. The access history reveals that the page of the location-based ranking only drew participants attention in the beginning stage of the study. The access number shows an obvious decrease over time, and eventually dropped to zero per day in the last phase of the experiment. Moreover, in the post-test questionnaire there was no participant that claimed that he/she answered those questions because of the location-based game.

To better interpret this phenomenon, we further analyzed the top active users. We found four of the top five most frequent local experts were from group A, and the exception is from group C. Based on their replies in the questionnaire, all of them have demonstrated strong intrinsic motivations,

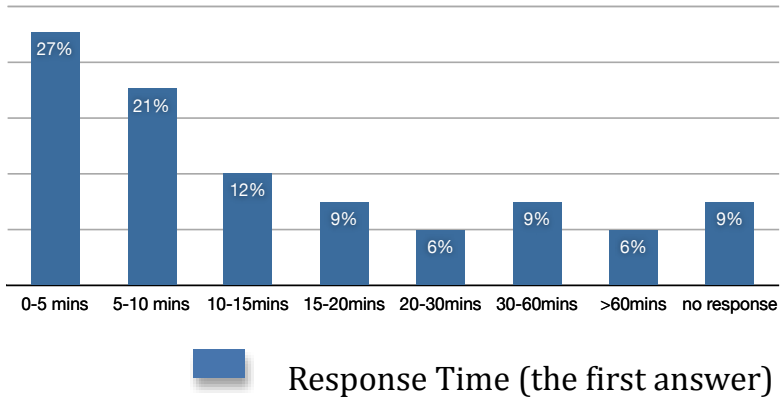


Fig. 11 the Performance Results: Response Time

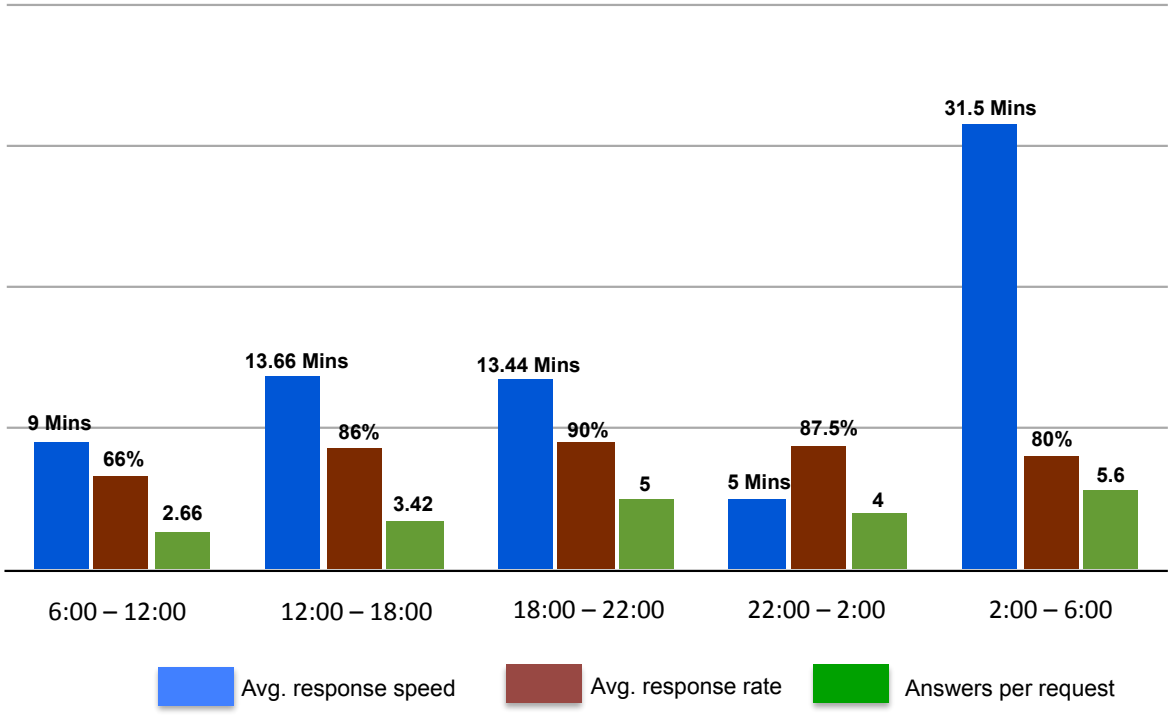


Fig. 12 the Time-of-the-Day Breakdown of the Average Response Speed, Response rate, and Answers per Request

i.e., *"I want to help the people in trouble"* or *"I want to introduce Japanese cultural to foreigners"*, which may have a stronger impact than the extrinsic motivations we provided.

On the hand, however, if we measure the user's participation as the number of users in the system (i.e., users who answered at least one question), and give equal weight to every user no matter the amount of the answers they provided (i.e., minority highly active users data will have less power of influence to the final comparison result), the result will be more explainable. In this case (see Figure 13-III), we observed the same number (38%) of active users were from group A and B, and fewer users (24%) were from the control group.

Overall, we argue that although the most frequent participants might be more motived by intrinsic incentives, the effectiveness of the designed extrinsic incentives to the rest of less self-motived users was still verified. Nevertheless, base on the results, we could hardly come to the conclusion

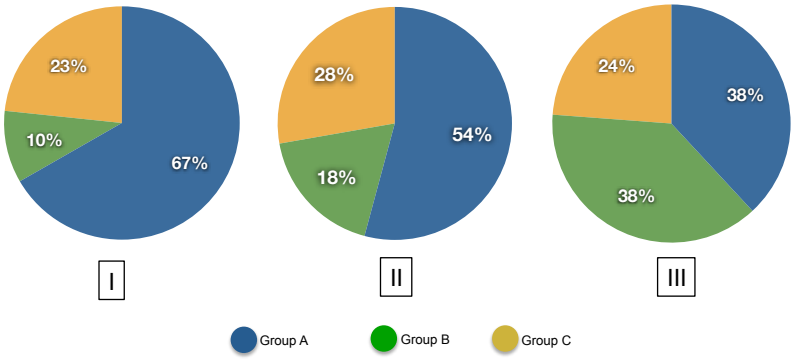


Fig. 13 I: Response Time by Groups; II: Number of Answers by Groups; III: Number of Active Users by Groups

that the proposed game-based incentive has a greater impact than the social psychological incentives. We believe it is not only because the game may not very interesting for the local experts, but also because the expected fierce competition did not happen due to insufficient requests' number.

4.3.3 Types of Requests

Three main types of questions were identified: translation, explanation, and cultural (see Figure 14). The number of answers of each type of questions is also shown in Figure 15. As most of the requesters cannot speak Japanese, the majority of the requests were translation related, including pure *translation* questions (42%) and requests for to *explanation* of a thing or a piece of text in English (36%). But as for local experts, results clearly shows they were more interested in answering explanation questions (44%) than pure translation questions (39%), because pure translation can be extremely boring work, e.g., translating the meaning of every button on a controller or a washing machine. In contrast, providing higher layer information is much more interesting. When asked the main reason for not answering a question, one of the top 5 active local experts also said:

> *"Sometimes the entries were not question, actually they are translation request."*

Cultural related questions were also popular. It may be quite difficult since normally such question requires answerers with some extra domain or local knowledge, but it is more meaningful for both parties, sometimes it can even be a quite interesting interaction:

> *"Q: Why all toilets in Japan are with brand TOTO?*
>
> *A: I searched the topic on the Internet. I found that TOTO used a beautiful woman as their advertisement while the competition company used a gorilla. Thats why TOTO toilet is widely used."*

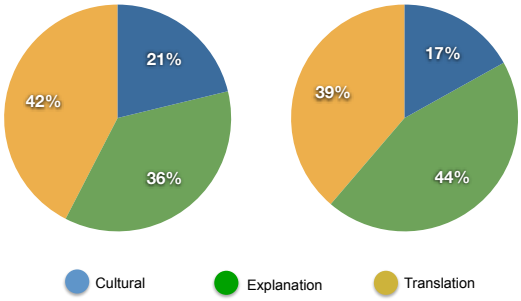
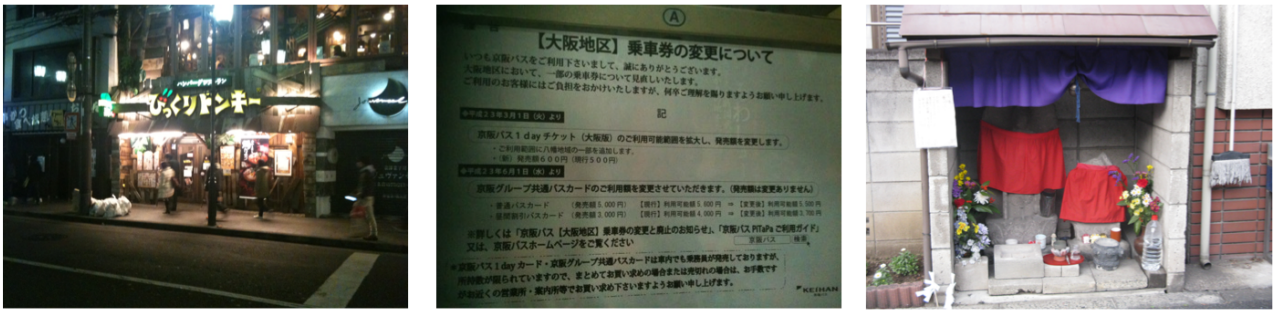


Fig. 14 left: Number of Questions by Question Types; right: Number of Answers by Question Types

5 Conclusions

In this paper we introduced a mobile crowdsourcing platform built on top of existing social and mobile computing infrastructural. UbiAsk, a mobile system for supporting foreign traveller by involving local user in the cloud to answer their image-based queries in a timely fashion, have presented as one case study. We have described the user center design circle of UbiAsk and the findings of a qualitative formative experiment. Moreover, we presented the results of six weeks quantitative field study of UbiAsk.

The user study results demonstrated a reliable overall performance on response speed and response quantity: half of the requests were answered within 10 minutes, 75% of requests were answered within 30 minutes, and on average every request had 4.2 answers. Especially in the afternoon, evening and night, nearly 88% requests can be answered in average approximately 10 minutes, with more the 4 answers per request. We also investigated the participation motivation of crowdworkers. We found the top active users were



Explanation
Q: What is this restaurant? What kind of food is it?

Translation
Q: What does this sign say? Is it a day pass?

Cultural
Q: Why the statues are with red cover?

Fig. 15 Examples of Different Types of Requests

more driven by intrinsic motivations rather than any of the extrinsic incentives (i.e., game-based incentives and social incentives) we provided. But the effectiveness of designed extrinsic incentives to the less self-motived users was still verified. However, based on this study we could not come to any conclusion that can suggest the comparison result between the game-based incentives and the social psychological incentives. In this section we discuss the implications of our findings.

Existing machine-based image search or translation applications (e.g., [14, 19], *Word Lens* and *Goggles*) provide very limited performance in real world conditions (such as complex background, dark environment, blurred photos, irregular fonts/handwritings), and cannot satisfy users who want to ask concrete questions. Whereas the proposed crowdsourcing based approach showed the potential to deliver reliable service under above-mentioned situations and answer user's concrete questions on demand. Also, most of the existing social search or human search system (like [9, 23]) are not able to gratify mobile user whom requires an immediate response. UbiAsk system, on the other hand, is designed for mobile context thus can deliver rapid answers in (almost) real time. Moreover, in contrast to paid mobile crowdsouring systems such as [12, 32], UbiAsk demonstrated a reasonable good performance by engaging with (free) active crowd workers from existing social media.

5.1 Users Desire Rich Information Instead of Plain Translation

Real-time image translation is one of the classic application ideas for the mobile argument reality systems. But how useful it really is in the real world situations? For example, one popular usage scenarios of such system are the translating of important signs, but one of our pilot testers has said:

> *"No, I can almost always make a reasonable guess of the meaning of such signs, just based the contextual information, or simply follow what other (local people) do."*

Based on the user study results, we confirmed that the majority of users were actually more interested in higher-level information rather than literature translation. That suggests even if the OCR and MT technologies can eventually always provide 100% correct results, they still cannot meet most of the users needs. Travelers without language skill normally tend to ignore most of the information surrounding them, and are only concerned about the information directly related to their activities. Most of the needed information is a rich explanation (or the real meaning) of a text rather than simple translation of the semantic meaning, e.g., in the case of a dish, *"pork, spicy, famous Chinese food"*, is a better answer for the requester than *"Huiguo rou"* (the actual name of the dish) as far as understandability is concerned.

5.2 Image Q&A vs. Text Q&A

Participants were asked their preference on picture-based questions and text-based questions in the questionnaires. A quite even number of users voted to *"depends on the situation"* and *"picture-based question"*, only one answerer preferred *"text-based question"*. This results well support the common setting of the mainstream on-line question answering services — have uploading picture as an optional choice.

In the context of mobile application we believe the picture-based question is even more useful and important. The questions people asked in the mobile context are often directly related to their current situations and the surroundings, which mean there is almost always something for them to "picture", and a picture can normally express a situation a lot more easier than explaining it in a text. Besides, via mobile devices interface it is normally harder to input text compared to computer keyboard, and thus even if user can describe the context in text, it is still easier to just attach an image.

Moreover, our application involves users from different countries using different languages. In many cases both parties communicated using their second or third languages. Hence the picture also becomes a very important component of the quality assurance mechanism to reduce the misunderstanding between the asker and answerer.

5.3 User-User Communication

Our current design does not involve any means for establishing a direct link between translators and requesters, but the necessity of such a communication link is worthy to be discussed. From the study results we noticed a trend of requiring worker-to-requester communication. When local experts cannot comprehend the meaning of the questions, some of them wish to confirm what they have understood with the requester. On the other end, the requesters may need to ask further questions related to an answer they received. In the user study design we actually took this issue into consideration. We provided the answerers email address or Twitter id together with their answer, but did not instruct the requesters to use the contact information or embassy on this design. Via post-test questionnaire and unobtrusive observation we found none of the requesters further contacted the answerers, although in the questionnaire many of the requesters expressed their desires to ask further questions related to the answer and to thank the answerer. But there is one answerer tried to communicate with requester via his answer:

> *"I think that attached picture is wrong. Please attach again."*

However, building a communication link brings obvious drawbacks as well. Serial and continuous questions heavily increase one local experts workload, which is against our original intentions to outsource micro-weight tasks to a large number of work users. When single task becomes heavier, active user participation and engagement likewise becomes much more difficult to achieve.

6 Future Directions

In this section we discuss the future directions of the proposed mobile crowdsourcing platform.

6.1 Cultural Differences and Incentives

Many studies have been conducted on the topic of designing incentive mechanisms in on-line systems, however only a handful of research have insight into how the difference of the people' cultural backgrounds may influence their preferences and decisions.

D. Chandler et al. [7] conducted a natural field experiment that investigates the meaningfulness of a task and people's willingness to work. They compared the outcomes from crowdworkers from different countries (e.g., U.S. and Indian). The results reported USA workers were induced to work at higher proportion when given cues that their task was meaningful, while Indian users were not. J. Ross et al.'s research [29] further explained the results by exploring the worker population and usage behaviors in Amazon Mechanical Turk. They found that, unlike moderate-income U.S. based workers who were more likely doing tasks for fun, there were increasing numbers of International workforce from lower-income countries (e.g., Indian) who treat the micro tasks as one of their major income source and actually rely on it to make ends meet.

6.2 Crowdworker Availability Detection

How to achieve efficient and appropriate task allocation is an important topic of this work. Here appropriateness stands for two aspects: the capacity and the availability. Capacity indicates whether a worker has enough knowledge or skill to accomplish the task, and availability is about if this is a good time that the user willing to work. The former aspect mainly affects the quality of translation, and the later one may affect the quantity.

It is not only about if people are free [25], but also involves other factors like social relationship, expertise, properties of questions, etc. We will look deeper in this issue in the future. In fact, we noticed it is a common issue existing in various fields. For instance, Tejinder and Carman [18] summarized the design challenges in future domestic communication technologies and indicated that one important issue is how to represent the true availability or the "willingness" to video conference in the initiating stage. Besides existing research, we believe the user availability detection technology also opens new possibilities in ubiquitous computing research. If the availability of an individual at given time is detectable, both response rate and time of mobile crowdsourcing can be greatly improved. Thus, in addition to use people as processors (as what we do in human computation), people are also amenable to use as sensors to perform tasks with relatively harder real-time requirements, which is obviously a promoting future application area. For example, people can be employed to collect high-level context information (e.g., human activity, group emotion, non-electronic object's location, identification or state, etc) of a given environment. Such rich data are extremely expensive and difficult to collect via machines, but very valuable and useful for ubiquitous computing applications.

References

1. Esp games *http://www.espgame.org/gwap/* last checked: February 2011.
2. ADAR, E., AND HUBERMAN, B. Free riding on gnutella. *First Monday 5*, 10 (2000).
3. ANTIN, J. Designing social psychological incentives for online collective action. In *Directions and Implications of Advanced Computing; Conference on Online Deliberation* (2008), p. DIAC 2008.
4. ARASE, Y., XIE, X., DUAN, M., HARA, T., AND NISHIO, S. A game based approach to assign geographical relevance to web images. In *Proceedings of the 18th international conference on World wide web - WWW '09* (New York, New York, USA, 2009), ACM Press.
5. B LATANÉ, K. W., AND HARKINS, S. Many hands make light the work: The causes and consequences of social loafing. *Personality and Social Psychology 37*, 6 (1979), 822–832.
6. BENABOU, R., AND TIROLE, J. Intrinsic and extrinsic motivation. *Review of Economic Studies 70*, 3 (2003), 489–520.
7. CHANDLERA, D., AND KAPELNERB, A. Breaking monotony with meaning: Motivation in crowdsourcing markets. Second Draft, University of Chicago Working Papers in Economics, 2010.
8. CHESHIRE, C., AND ANTIN, J. The social psychological effects of feedback on the production of internet information pools. *Computer Mediated Communication*, 13 (2008), 705–727.
9. D HOROWITZ, S. D. K. The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web (WWW 2010)* (2010), pp. pp. 431 – 440.
10. DECI, E., AND FLASTE, R. *Why We Do What We Do: Understanding Self-Motivation.* London: Penguin, 1996.
11. DECI, E. L., AND RYAN, R. M. *Intrinsic Motivation and Self-Determination in Human Behavior (Perspectives in Social Psychology)*. Plenum Press, 1985.
12. EAGLE, N. txteagle: Mobile crowdsourcing. In *Proceedings of the 3rd International Conference on Internationalization, Design and Global Development: Held as Part of HCI International 2009* (2009), IDGD '09, pp. 447–456.
13. FREI, B. Paid crowdsourcing: Current state & progress toward mainstream business use. In *Whitepaper Produced by Smartsheet.com*. 2009.
14. FUJISAWA, H. Forty years of research in character and document recognition-an industrial perspective. *Pattern Recognition 41*, 8 (2008).
15. HARITAOGLU, I. Infoscope: Link from real world to digital information space. In *Proceedings of Ubicomp 2001: International Conference on Ubiquitous Computing* (2001).
16. HOWE, J. The rise of crowdsourcing. *Wired Magazine*, 14.06 (2006).
17. HSUEH, P.-Y., MELVILLE, P., AND SINDHWANI, V. Data quality from crowdsourcing: a study of annotation selection criteria. In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing* (2009).
18. JUDGE, T. K., AND NEUSTAEDTER, C. Sharing conversation and sharing life: video conferencing in the home. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems* (New York, NY, USA, 2010), ACM, pp. 655–658.
19. KOGA, M., MINE, R., KAMEYAMA, T., TAKAHASHI, T., YAMAZAKI, M., AND YAMAGUCHI, T. Camera-based kanji ocr for mobile-phones: Practical issues. In *ICDAR '05: Proceedings of the Eighth International Conference on Document Analysis and Recognition* (2005), pp. 635–639.
20. KRAUSE, M., TAKHTAMYSHEVA, A., WITTSTOCK, M., AND MALAKA, R. Frontiers of a paradigm: exploring human computation with digital games. In *Proceedings of the ACM SIGKDD Workshop on Human Computation* (2010), HCOMP '10.
21. LADA A. ADAMIC, JUN ZHANG, E. B. M. S. A. Knowledge sharing and yahoo answers: Everyone knows something. In *Proceedings of the 17th International World Wide Web Conference (WWW 08)* (2008).
22. LIU, Y., LEHDONVIRTA, V., KLEPPE, M., ALEXANDROVA, T., KIMURA, H., AND NAKAJIMA, T. A crowdsourcing based mobile image translation and knowledge sharing service. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia* (2010), MUM '10.
23. MACDONALD, C., AND OUNIS, I. Voting for candidates: adapting data fusion techniques for an expert search task. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management* (Arlington, Virginia, USA., 6-11 November 2006).
24. MOCKUS, A., R. F., AND ANDERSEN, H. Two case studies of open source software development: Apache and mozilla. *ACM Transactions on Software Engineering and Methodology 11*, 3 (2001), 309–346.
25. MORRIS, M. R., TEEVAN, J., AND PANOVICH, K. What do people ask their social networks, and why?: a survey study of status message q&a behavior. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems* (2010), pp. 1739–1748.
26. NAKAJIMA, H., MATSUO, Y., NAGATA, M., AND SAITO, K. Portable translator capable of recognizing characters on signboard and menu captured by built-in camera. In *Proceedings of the ACL 2005 on Interactive poster and demonstration sessions* (2005).
27. NIELSEN, J. Recruiting test participants for usability studies *http://www.useit.com/alertbox/20030120.html* last checked: February 2011.
28. REKIMOTO, J. Navicam: A palmtop device approach to augmented reality. *Foundamentals of Wearable Computers and Augmented Reality, Woodraow Barfield and Thomas Caudell (ed.), Laurence Erlbaum Associates, Publishers,* (2001).
29. ROSS, J., IRANI, L., SILBERMAN, M. S., ZALDIVAR, A., AND TOMLINSON, B. Who are the crowdworkers?: shifting demographics in mechanical turk. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems* (2010), CHI EA '10.
30. VON AHN, L. Games with a purpose. *IEEE Computer Magazine 39*, 6 (2006), 92–94.
31. VON AHN, L. Human computation. In *K-CAP '07: Proceedings of the 4th international conference on Knowledge capture* (2007), pp. 5–6.
32. YAN, T., KUMAR, V., AND GANESAN, D. Crowdsearch: exploiting crowds for accurate real-time image search on mobile phones. In *Proceedings of the 8th international conference on Mobile systems, applications, and services* (2010), ACM.
33. ZAJONC, R. B. Social facilitation. *Science* (1965), 269–274.
34. ZICHERMANN, G., AND LINDER, J. *Game-Based Marketing: Inspire Customer Loyalty Through Rewards, Challenges, and Contests*. Wiley: London, 2010.
35. ZYDA, M. From visual simulation to virtual reality to games. *Computer 38*, 8 (September 2005), 25–32.