

A Crowdsourcing Based Mobile Image Translation and Knowledge Sharing Service

Yefeng Liu, Vili Lehdonvirta[‡], Mieke Kleppe[‡], Todorka Alexandrova,
Hiroaki Kimura and Tatsuo Nakajima
Department of Computer Science, Waseda University
Helsinki Institute for Information Technology[†]
Eindhoven University of Technology[‡]
{yefeng,toty,hiroaki,tatsuo}@dcl.info.waseda.ac.jp
vili.lehdonvirta@hiit.fi[‡], m.kleppe@student.tue.nl[‡]

ABSTRACT

Travelers in countries that use an unfamiliar script cannot use pocket translators or online translation services to understand menus, maps, signs and other important information, because they are unable to write the text they see. Solutions based on optical character recognition provide very limited performance in real-world situations and for complex scripts such as Chinese and Japanese. In this paper, we propose an alternative image translation solution based on crowdsourcing. A large number of human workers on mobile terminals are used to carry out the tasks of image recognition, translation and quality assurance. Compared to purely technical solutions, this human computation approach is also able to account for context and non-textual cues, and provide higher level information to the end-user. In this paper, we describe a preliminary user study to create a model of end-user requirements.

Categories and Subject Descriptors

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; H.1.2 [User/Machine Systems]: Human Factors

General Terms

Design, Human Factors

Keywords

Crowdsourcing, mobile image translation, image-text recognition, human computation, knowledge sharing

1. INTRODUCTION

The increasing market penetration of mobile phones with Internet connectivity, high processing power and integrated sensors such as cameras [2] has given rise to a new ubiquitous computing platform. As is typical with new technologies,

entertainment applications such as mobile games are leading the growth. The purpose of our research is to examine ways in which the platform could be used for new kinds of human computation or crowdsourcing applications – applications that make use of the distributed and always-on nature of the mobile phone, while borrowing ideas from gaming to attract contributions from users.

The specific application area addressed in this paper is mobile image translation [19], which refers to camera phone applications that attempt to solve the problem of translating text written in an unfamiliar script; for example, Chinese, Japanese, Thai or Arabic. Traditional digital pocket translators and online translation services are useless if the user is unable to input the unfamiliar characters into the device. Mobile image translators typically attempt to solve this by recognizing the characters optically.

Image translation services (e.g., Ta With You¹, Interlecta Translator for BlackBerry smartphones², RantNetwork Communicator³, [19]) generally consists of two steps. The first step involves extracting text from an image taken by the mobile phone's camera. The second step involves translating the text. The first step is often considered the more difficult one, and the bottleneck of current image translation applications. This is because text and handwriting recognition are still very difficult and computationally expensive tasks for computers. Thus mobile image translators are mostly limited to languages written in regular Latin scripts on simple backgrounds. Figure 1 shows examples of typical texts that such applications usually cannot handle.

In this paper, we explore a crowdsourcing based solution for mobile image translation. Crowdsourcing [15] is a way of outsourcing tasks, traditionally performed by an employee or contractor, to a large group of people (a crowd), through an open call. Here, the image-text translation tasks are outsourced to a translator community. A difference to traditional crowdsourcing models is that the work assignments also originate from a crowd of numerous end-users with differing needs. The approach can also be described as a type of human computation [25], which is understood as the no-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MUM'10, December 1-3, 2010, Limassol, Cyprus

Copyright ©2010 ACM 978-1-4503-0424-5/10/12 \$10.00

¹<http://www.tauiyou.com/en/image.html>

²<http://home.interlecta.com/paypal-blackberry/>

³<http://www.rantnetwork.com/communilator.aspx>

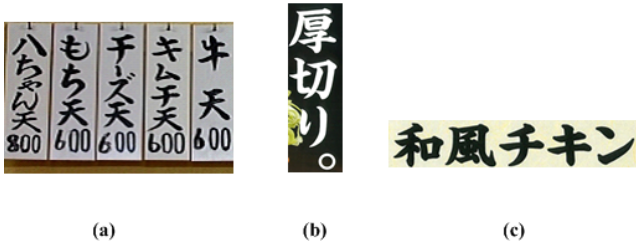


Figure 1: Menu and signs from a Japanese restaurant as the examples of image-texts which are difficult for machine to recognize

tion of solving difficult AI-related computation problems by obtaining help from (ordinary) humans instead of applying machine algorithms. Compared to a purely machine based solution, it is harder for a human computation solution to provide real-time or on-demand service, but the quality of the outcome is matchless because humans are as of yet much better than machine solutions in many tasks. Another significant feature of human workers is that they can provide richer interpretations and responses to tasks in addition to literal answers. In mobile image translation, this can mean that the system works not only as a translator, but also as a knowledge broker that allows users to share higher level information such as advice, instructions and suggestions pertinent to the situation at hand. The result can be seen as a mixture of social search engine[14] and mobile image translator capabilities.

In this paper, we introduce the human mobile image translator concept and report on the results of a preliminary user study conducted to yield a model of end-user requirements, a first step towards concrete implementation [7, 10]. The user study took the form of a simulation or experiment, where real potential end-users sought help to their actual translation problems and real potential translators responded to them. More generally, this research aims to address the relative scarcity of real-world studies on mobile and ubiquitous *human computation* and *human sensor* applications, as opposed to purely machine-based solutions. As argued above, the chosen application area is also important in itself, and a successful translation system would be a major contribution. The main contributions of this paper are introducing the concept of human mobile image translation and sketching an architecture based on real user requirements and issues identified in the user study.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed concept in detail. In Section 3, we describe the results of the user experiment. In Section 4, a number of interesting research findings are discussed. In Section 5, some important related researches are described. And finally, the future direction of our project is discussed in Section 6.

2. HUMAN MOBILE IMAGE TRANSLATION

In this section we present the design of the crowdsourcing model for mobile image translation and knowledge sharing. We start with a scenario to illustrate the problem considered in the paper.

2.1 Scenario

“Daniel is a western traveler who has just arrived in Japan. He cannot speak Japanese and of course he has no idea how to write or read Kanji (the Japanese version of Chinese characters). When traveling in Europe, Daniel would not worry about the language problems, as his smart-phone can provide enough support for translating. However, the situation in Japan is somewhat different. Though he can still access on-line translators and use off-line dictionaries easily, the real problem is how to input the Japanese text? The following situation is an example of one of the many cases when he would be happy to use translation help.

Daniel notices an interesting sign (see Figure 2) on the menu board in front of a Japanese restaurant. From the image on it he understands it is related to ties but he has no idea about the exact meaning. He gets really curious whether you are allowed or not to wear a tie inside this restaurant? He cannot use a digital dictionary as he has no idea how to input the characters. He tries to ask the waiter about the meaning of the sign but he cannot answer in English.”

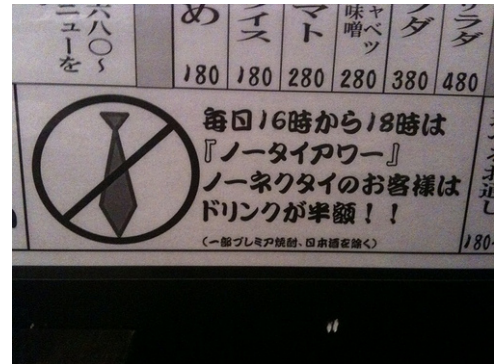


Figure 2: A menu board outside a restaurant

The above scenario explicitly describes the problem of the existing text-based translators. Logographic script (e.g., Chinese and Japanese) input methods are totally different from Latin-based languages, since users cannot type the character unless they know how to pronounce it. Handwriting input method could be one solution, however it is admittedly time consuming, especially when the characters are unfamiliar to the user. Furthermore, it also requires many hardware/software prerequisites, e.g., handwriting recognition tool, support for foreign language, touch screen, and so on.

2.2 Crowdsourcing Platform Overview

In this paper we propose a mobile crowdsourcing solution to the problems described in the previous section. Since the workload of each job in the mobile image translation service is lightweight enough to be described as a *micro-task*, the tasks are perfectly suitable to be outsourced to large groups of casual workers.

The platform utilizes a server-client architecture. The difference with the conventional server-client structure is that

there are two types of clients: the users who make requests are called *client users*, and the translators are named *work users*. On the other end, the server plays a proxy role. It receives translation requests from client users, assigns these tasks to appropriate work users, and forwards translators' final answers to the original requesters.

Figure 3 illustrates the basic work-flow of the proposed translation model and a detailed description of it is given below.

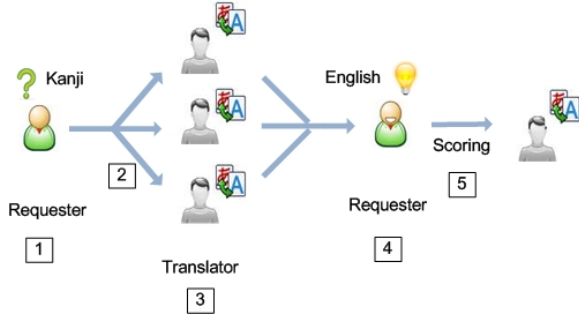


Figure 3: Basic work-flow of the proposed model

1. A client user makes a translation request by taking a picture of the text using a mobile phone's camera, and submits the image to the server. Optionally, a short text message can be attached with the photo to clarify what exactly they want to know. Context information, such as location and time, will also be automatically attached to the request, although the availability of the context information may depend on the client's terminal's functionality (e.g., if GPS module is embedded or not). Such context information, together with a work user's profile, is useful for assigning tasks appropriately.
2. For the purpose of enhancing the response time, each task is assigned to multiple workers, simultaneously.
3. The original request is sent to translators via email. Translators are encouraged to reply in "key words" or "tag" style. In many cases what the requester wishes to know is not the semantic meaning of the word, but the real meaning, e.g., in the case of a dish, "pork, spicy, Chinese food, famous" is a better answer than "Twice Cooked Pork (Huigu rou)" as far as understandability is concerned.
4. For the response time considerations, the first answer to be received from translators 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.
5. Eventually, the client user receives the final result, and optionally scores the participants according to the quality of the translation outcome. Before the client scores, a default amount of points is given to each translator. The client has one day to decide the points, however they cannot change the score once given.

The translation service is designed as a mobile touch screen application (e.g., iPhone/ Android/ Symbian apps, etc.).

The main interface of this application consists of three main parts. i) After opening the application the first interface is shown, which is a camera display view. Clients take pictures of what they want to ask using this view. ii) Then, a simple image editing tool is provided, by which the users can emphasis the important points in the photo that they are interested in (e.g., by circling them). iii) The last part is a text editor, which is used for inputting a short message to the translator, that describes what exactly the requester wants to know. As for translators, they are first asked whether they are willing to accept the task. If the answer is positive, an image editor and a text editor are provided as translation tools that translators may use depending on the request content.

3. USER 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 to answer the following research question: *How should the user's questions be presented to the translator 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, 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 they require is much 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 a *image translation service* mixed with an *community based mobile semi-realtime Q&A service*. Based on this finding, in the later meetings we have improved our concept idea and extended the design from simple translation to a knowledge sharing service. Eventually the experiment work flow has been realized as follows.

First we collected around one hundred example images (taken by mobile phone) from foreigners who are currently staying in Tokyo, together with their corresponding questions. From these images, we selected fifteen characteristic cases from different types of users (or we could also say, from users who had different needs). Then, we interviewed the photos' providers, questioning what kind of answers they were expecting. In the next stage we sent the requests to a group of seven invited translators (mostly Bachelor or Master students from School of Science and Engineering, Waseda University). After receiving the answers from the translators, we interviewed the translators for collecting their feedback on the usage of the system. Finally, we compared the results from the translators with the expected answers from the requesters, identifying whether they matched or not and discussed the possible reasons for mismatching.

In this paper, we only present four typical user study cases that were observed in the experiment.

3.1 User Study

Picture I The picture in Figure 4 is received from a traveler who has come to Tokyo for a short term vacation. He had no Japanese language skills at all and that is why he

also tended to ignore most of the information surrounding him, being only concerned about the information directly related to him. He took this photo when he was waiting for a train at the station a bit longer than expected, when he noticed that the information display was showing some special information instead of the arrival time of the coming trains.

“Why isn’t there any train? How long do I have to wait?”



Figure 4: Picture of study case I

When we interviewed the requester, he said at that moment he understood there was some trains delay, but he didn’t know if keeping waiting was a good choice or not, since some people were leaving, while others were staying. During the translators’ interview, four of them considered this question as difficult, mainly because *“The information in the picture is too little”*. or *“The reason is hard to explain in English.”* However, the result shows most of the collected answers were actually good enough for the requester, because he was not really concerned with the reason of the delay but what he should do seeing the message on the display. Most of the replies suggested an approximate time the requester might need to wait, e.g., *“Maybe, you have to wait for thirty minutes to one hour.”*, although the exact time to wait and the exact reason for the delay were not given (which was, in fact, not possible to know), e.g., *“There is a trouble and you have to wait some more minutes. But I don’t know how long exactly you will wait.”*

Picture II From the collected questions we found out that in many cases requests were driven from curiosity rather than real problems. A typical example is when the users obtain a partial information for something interesting but are unable to figure out the whole information that they want to know. If we take a poster as an example, people (who cannot read Japanese) can see the image, understand the time and the date, but are unable to understand what the event is about, what the location is, and other detailed information. Figure 5 is an example of such a poster and the user’s question in this case is:

“What are the events between 5th and 8th?”



Figure 5: Picture of study case II

The requester said she was deeply attracted by the poster because she really wanted to experience a Japanese traditional event during her stay in Tokyo. On the poster there were events schedules for July 3 and July 4, however, when she saw the poster it was July 5, already. So she wanted to know what kind of events she would be able to attend from that day on. Unfortunately two translators misunderstood her question and their answer was *“The event is a festival of the weaver”*. But afterwards they also pointed out that if the question was written in a clearer way (e.g., *“There is only information about events on 3rd and 4th, what are the events between 5th and 8th?”*), such misunderstanding would have been avoided. Three of the other answers indicated that there were no special events during the mentioned period, and some of them suggested *“You can still enjoy a general Japanese festival.”*

Picture III Many foreigners in Japan could still have troubles with reading and writing Kanji even after staying for a few years. They usually have no urgent needs of translating/explaining everyday words, but may need help on unusual phrases or specialized vocabularies. Specialized texts such as recipes will be a good example of this kind of requests. Figure 6 has been provided by a foreigner who has been staying in Japan for three years and the picture shows one recipe in her cookbook.

“What are these two? Can you provide links of pictures?”

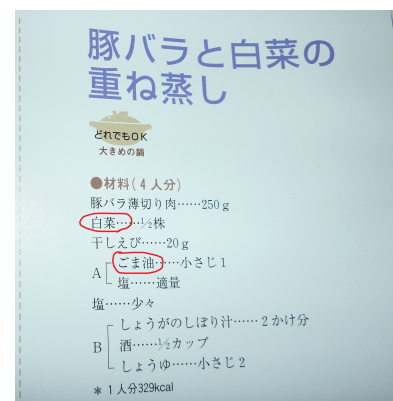


Figure 6: Picture of study case III

This kind of nouns is particularly difficult for language learners, since they can hardly be learned from vocabulary books, and are rarely used. The requester comes from a non English speaking country, which means that even if the translator provides English words, it is very likely she still has no idea of what it is. That is why she asked for links of pictures. The results show that the three translators, who understood the question correctly, provided good responses. However, another three translators appeared to start answering the question before reading all the question carefully. Instead of replying back the requested links, they were trying to give a text explanation of the ingredients, which was very difficult. Half of them ended up with somewhat useless answers, and others gave up the task.

Picture IV This example belongs to the most difficult ones. For people who have been staying in Japan for several years, ordinary or basic knowledge issues would not be a problem. What they would really like to know is mainly higher level information, often related to other background knowledge e.g., cultural understanding. Considering most of our work users are not skilled or experienced translators, providing a text-based answer for this type of requests or explanation can be quite challenging. Figure 7 illustrates a typical example of this kind, an electricity bill.

“What is the difference between 1 and 2 in my bill?”

上 記 料 金 内	基本料金
	電力量料金
	・ 1 段料金
	・ 2 段料金
	・ 燃料費調整額
	太陽光促進付加金

Figure 7: Picture of study case IV

Literally it is unlikely to understand what is the difference between the two fees, the direct translation of the two phrases is “the first level fee” and “the second level fee”, respectively. The requester said normally she would ask for help her Japanese friends the next day, but then she had to wait for at least one night, and sometimes the document contained information which she was unwilling to share with her acquaintances, e.g., “*result of medical examinations*”. The experiment results confirmed the difficulty of this type of requests. Three of the translators replied valuable answers (which satisfied the origin requester), while others gave up the task saying “*I don’t know the difference.*” or “*Even if I know, maybe I can’t explain in English.*”. However, the fact that the majority of our translators are college students could also be part of the reason for this outcome, i.e., such knowledge might be out of their life experience.

3.2 Summary of Findings

First of all, as one of the most important requirements of the proposed model, we consider the clear statement of both the requester’s question (to translator) and the translator’s answer (to requester). Translator needs to understand the real question of the requester, in order to be able to provide the

desired answer for the requester. The system should support the functionality of avoiding the misunderstanding between the two parties in order to achieve an accurate result.

We found that in most cases, once the translators have the correct understanding of the question messages from the requesters, they can give reasonable and useful answers. The results also show that unclear writing (by requester) and hasty reading (by translator) are the major reasons for misunderstanding of the question. To overcome this problem one of our suggestion to the requesters is to always ask questions with choices as long as possible, since it is the clearest way to express what they mean. Moreover, requesters should always better clarify to what level of details they want to know the answer, e.g., “*Which station/ district/ city?*” is a better question than “*Where?*” for translators. The lack of enough language skills to explain the answer is another reason that may lead to unfavorable results. The translators are (normally) not native English speakers, that is why they often face the problem of being unable to explain clearly their thoughts in English. Therefore, we encourage translators to reply a short text message rather than long sentences.

“Because I’m not fluent in English, tweeting-style (short message style) is good for me.”

Another effective workaround of the lack of language skills problem is the idea of replying with a link which can explain the question. This is a clear and distinct solution for both parties. For some situations, the translator can even only response with a link of the Wikipedia⁴ or Google Images⁵ search results. This would be a useful and satisfactory information for the requesters that are not able to type the Japanese characters and search on Internet by themselves.

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.

4. DISCUSSION

In this section we focus on the discovered design implications from the results of the user study.

4.1 Useless Answers due to Misunderstanding

The quality of the results in crowdsourcing is always a hot issue in human based systems, and has been discussed widely recently [3, 12, 16, 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

⁴<http://www.wikipedia.com>

⁵<http://www.google.com/imghp>

second or third language and thus, they may also 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, or other mechanisms. Moreover, depending on the question, the translator 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 simply circle the corresponding part in the picture rather than giving a description by words. However, a translator (as a human) always makes mistakes no matter how perfect the instructions are. For example in our study case in Picture III, 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.

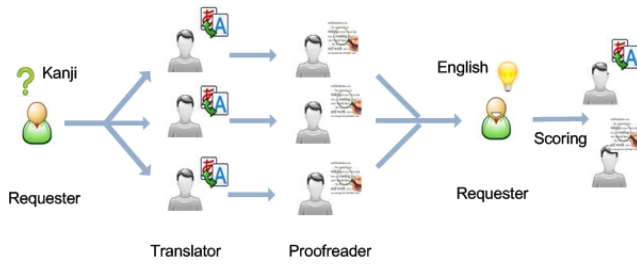


Figure 8: Work-flow with an additional proofreading phase

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 (see Figure 8) 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 the purpose of maintaining a more valuable results database.

4.2 Accuracy vs. Timeliness

Based on the experiment results, we identify different types of users. Clients can first be divided into two classes based on their period of stay in the foreign country, i.e., *short-term* (e.g., tour, business, visiting, etc.) and *long-term* (e.g., study, work, training, etc.). Then, users can be further classified according to the time requirements on their requests. Some of the translation tasks expect *immediate* answers (e.g., menu, instruction, sign, etc.), in most cases the requesters are blocked in the middle of their on-going activity and they have to wait for this answer to continue. The other type of requests are so called *waitable* questions (e.g., documents, posters, etc.), which often involve more complicated question contents. The translator may need to add extra explanation or related background information

to avoid meaningless/useless answers or misunderstandings between the two parties rather than simply translating the semantic meaning.

Eventually, four basic types (A, B, C and D) of clients were defined (see Figure 9).

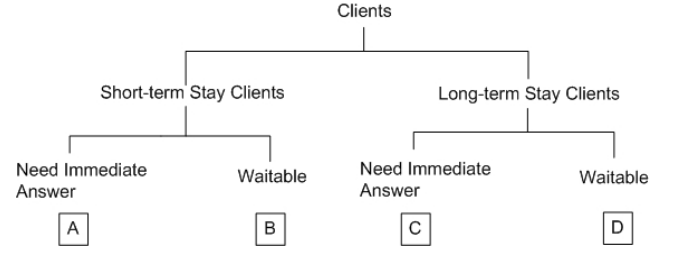


Figure 9: Four basic types of clients

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 with regards to 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 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.

4.3 Interactivity

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 translator-to-requester communication. When translators cannot comprehend the meaning of the questions, some of them wish to confirm what they have understood with the requester. On the other end, after receiving the reply, some of the requesters expressed their desires to ask further questions related to the answer.

However, building a communication link between the two parties brings a drawback as well. Serial and continuous tasks heavily increase translators’ workload, which is against our original intentions to outsource *micro-weight* tasks to a large number of work users. This point is a fundamental part of our design philosophy – if the job is heavy, it is very difficult to motivate people to participate.

5. RELATED WORK

There have been a number of image-text translation systems deployed over years (see [11] for an overview). Most of them utilize OCR technology to recognize images. 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.

There have been some earlier efforts related to language translation via crowdsourcing and mobile devices. Ambati et al. [24] proposed an Active Crowd Translation (ACT) for automatic translation of low-resource language pairs, which makes use of both active learning and crowdsourcing concepts. Callison-Burch conducted a significant study of the performance of crowdsourcing based translation quality evaluation [8]. His report concludes that compared to conventional manual evaluation, the online market is a faster, cheaper and more creative option. Neither of these studies were aimed at mobile translation and problem-solving, however. Eagle N. conducted a field study [9] in Kenya and Rwanda on a mobile crowdsourcing system *txteagle*, where the users were able to earn small amounts of money by completing simple tasks on their mobile phones. The tasks did not include translation, but the study demonstrates the use of mobile phones and economic incentives in crowdsourcing.

Other relevant studies can be found in the human computation research stream [25, 4, 5]. 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 images' geographical context for searching and browsing. Barrington et al. [5] proposed a Facebook⁶ game to collect data for building machine learning models that automatically associate music with tags. These concepts use game-like incentive systems and deal with task distribution, but do not address mobile and real-time requirements. In relation to the latter, Quinn et al. [22] proposed a toolkit called CrowdFlow for blending human computing and machine computing in order to attain tighter control over the inherent tradeoffs in speed, cost and quality.

6. FUTURE DIRECTIONS

This paper represents an early stage in the development of a human mobile image translation and knowledge sharing system. A number of research questions were discussed and addressed but most remain to be addressed. In terms of the next step of work, we look forward to addressing these research problems via a larger scale user study on a functional prototype.

6.1 Task distribution

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. Generally speaking, the former aspect mainly affects the quality of translation, and the later one may affect the quantity.

Compared to ordinary everyday task, micro-task are commonly defined as jobs with lightweight workload and low difficulty, hence should be able to easily handled by most normal people. As a consequence, the capability requirement of crowdsourcing is less strict. In this application, first of all, it is assumed that participants are people who at least have enough self-confidence of their skill in both languages. Furthermore, a work user is expected to provide a list of *familiar areas/places* when they sign-up. Server compares

⁶<http://www.facebook.com>

such data with each requester's location, and always prefers to assign a task to a worker who claim her/himself know this or nearby place well.

The other requirement, detection of a user's availability, is more difficult to deal with. It is not only about if people are free [21], 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 researches, 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), it is also possible to use human as *sensor* to perform tasks with relatively harder real-time requirements. For example, people can be employed to collect high-level context information (e.g., human activity, non-electronic object's location, identification or state, etc) of a given environment. Such rich data are extremely expensive and difficult to get via machines, but very valuable and useful for ubiquitous computing applications.

6.2 Implementation platform

Besides establishing a new model from top to bottom, we also consider the possibility of building our system atop of existing web-based platforms. On-line crowdsourcing marketplaces like Amazon Mechanical Turk⁷ support outsource tasks to the masses, though such marketplaces normally use financial rewards as incentive. Another choice is to use social networking and microblogging service Twitter⁸, since it is lightweight, easy to operate, current gained widespread popularity worldwide (registered user reaching 190 million by June 2010 [1]), and well support all needed functionalities (i.e., text/image transfer, positioning, etc.).

6.3 Results database

All *good* results (i.e., answers that get full score from requesters) should be stored due to their utility value. In addition, we need to require translators to tag all the images they are translating with the aim of more efficient accessing of the data. We believe this results database, as its size grows, can be of great value to various kinds of groups such as future clients, language learners, AI researchers, tourism service provider, etc.

6.4 Motivation and incentive for workers

Another key challenge of this system is how to design participation incentive mechanisms, since the performance of the system strongly depends on whether the task rewards can activate the work user's participation. To some extent, it is also one of the most significant and fundamental challenges in user participation systems such as human computation systems, persuasive systems, open-content publishing

⁷<https://www.mturk.com/>

⁸<https://www.twitter.com/>

systems. Previous studies have identified three broad approaches to motivating contributors: economic incentives, social incentives and intrinsic incentives [6, 13, 20, 23].



Figure 10: Interface of the location-based game

For example, our preliminary design leverages social and intrinsic motivations related to game play [4, 26]. As mentioned above, work users are awarded scores by the end-users. A location-based mobile game could be constructed on the basis of these scores or points. The main interface of game (see Figure 10) is a Google map based real world map, which is divided into non-overlapping hexagons. The goal of the game is to conquer territories. Every hexagon has one *owner* or *lord*, who is the player with the highest number of points awarded inside this area. The lord's profile photo along with an identifying color are displayed in their hexagons on the map. Depending on the extent and geographic location of their territory, players hold different special titles which are shown in their profile. In the next steps of this project, we plan to experiment with this and similar concepts.

7. REFERENCES

- [1] Costolo: Twitter now has 190 million users tweeting 65 million times a day, <http://techcrunch.com/2010/06/08/twitter-190-million-users/>.
- [2] Gartner says mobile phone sales will exceed one billion in 2009, http://www.gartner.com/press_releases/asset_132473_11.html.
- [3] ADAMIC, L. A., ZHANG, J., BAKSHY, E., AND ACKERMAN, M. S. Knowledge sharing and yahoo answers: everyone knows something. In *WWW '08: Proceeding of the 17th international conference on World Wide Web* (New York, NY, USA, 2008), ACM, pp. 665–674.
- [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, p. 811.
- [5] BARRINGTON, L., O'MALLEY, D., TURNBULL, D., AND LANCKRIET, G. User-centered design of a social game to tag music. In *HCOMP '09: Proceedings of the ACM SIGKDD Workshop on Human Computation* (New York, NY, USA, 2009), ACM, pp. 7–10.
- [6] BENABOU, R., AND TIROLE, J. Intrinsic and extrinsic motivation. *Review of Economic Studies* 70, 3 (2003), 489–520.
- [7] BROWN, T. *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*. HarperBusiness, 2009.
- [8] CALLISON-BURCH, C. Fast, cheap, and creative: evaluating translation quality using amazon's mechanical turk. In *EMNLP '09: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (2009), pp. 286–295.
- [9] EAGLE, N. txt eagle: Mobile crowdsourcing. In *Internationalization, Design and Global Development*, N. Aykin, Ed., vol. 5623 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2009, pp. 447–456.
- [10] EDWARDS, W. K., BELLOTTI, V., DEY, A. K., AND NEWMAN, M. W. The challenges of user-centered design and evaluation for infrastructure. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2003), ACM, pp. 297–304.
- [11] FUJISAWA, H. Forty years of research in character and document recognition-an industrial perspective. *Pattern Recogn.* 41, 8 (2008), 2435–2446.
- [12] GYONGYI, Z., KOUTRIKA, G., PEDERSEN, J., AND GARCIA-MOLINA, H. Questioning yahoo! answers. Technical Report 2007-35, Stanford InfoLab, 2007.
- [13] HARS, A., AND OU, S. Working for free? - motivations of participating in open source projects. In *Proceedings of The Annual Hawaii International Conference on System Science* (2001), p. 163.
- [14] HOROWITZ, D., AND KAMVAR, S. D. The anatomy of a large-scale social search engine. In *WWW '10: Proceedings of the 19th international conference on World wide web* (New York, NY, USA, 2010), ACM, pp. 431–440.
- [15] HOWE, J. *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Crown Publishing Group, New York, NY, USA, 2008.
- [16] HSUEH, P.-Y., MELVILLE, P., AND SINDHWANI, V. Data quality from crowdsourcing: a study of annotation selection criteria. In *HLT '09: Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing* (Morristown, NJ, USA, 2009), Association for Computational Linguistics, pp. 27–35.
- [17] IPEIROITIS, P. G., PROVOST, F., AND WANG, J. Quality management on amazon mechanical turk. In *HCOMP '10: Proceedings of the ACM SIGKDD Workshop on Human Computation* (New York, NY, USA, 2010), ACM, pp. 64–67.
- [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] LEIMEISTER, JAN MARCO, H. M. B. U., AND KRUMAR, H. Leveraging crowdsourcing: Activation-supporting components for it-based ideas competition. *Journal of Management Information Systems* 26, 1 (Summer 2009), 194–224.
- [21] 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* (New York, NY, USA, 2010), ACM, pp. 1739–1748.
- [22] QUINN, A. J., B. B. B. Y. T. L. J. Crowdfunder: Integrating machine learning with mechanical turk for speed-cost-quality flexibility. Tech. rep., Technical Report HCIL-2010-09 (University of Maryland, College Park, 2010).
- [23] SHIRAIISHI, M., WASHIO, Y., TAKAYAMA, C., LEHDONVIRTA, V., KIMURA, H., AND NAKAJIMA, T. Using individual, social and economic persuasion techniques to reduce co2 emissions in a family setting. In *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology* (New York, NY, USA, 2009), ACM, pp. 1–8.
- [24] VAMSHI AMBATI, S. V., AND CARBONELL, J. Active learning and crowd-sourcing for machine translation. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)* (may 2010).
- [25] VON AHN, L. Human computation. In *K-CAP '07: Proceedings of the 4th international conference on Knowledge capture* (2007), pp. 5–6.
- [26] VON AHN, L., AND DABBISH, L. Designing games with a purpose. *Commun. ACM* 51, 8 (2008), 58–67.

APPENDIX

A. EXPERIMENT RESULTS OF THE PRESENTED STUDY CASES

Question I : Why isn't there any train? How long do I have to wait here?	
Translator 1	Due to the traffic accident. I'm not sure how long you have to wait but the board says the train company might prepare alternative ways to go.
Translator 2	There are some troubles and you have to wait some more minutes. But I don't know exactly how long you will wait wait .
Translator 3	(No response)
Translator 4	Because there is an accident at Tokyu-Denentoshi line. I can't tell you how long you have to wait from only this picture.
Translator 5	I don't know the matter. Maybe, you have to wait for 30 minutes to 1 hours.
Translator 6	I have no idea.
Translator 7	The reason for delay of trains is uncertain. It is also unclear how long does it takes the operation of trains is recovered. But you may be able to take alternative route by using another transportation.

Question II :What are the events between 5th and 8th?	
Translator 1	There was no special event on 5th and 8th. On 3rd and 4th, there were some special events such as parading, flea market.
Translator 2	The event is Festival of the Weaver.
Translator 3	There is no information in this picture.
Translator 4	(No response)
Translator 5	The events are festivals on 5th to 8th. There are many shops along the street.
Translator 6	A festival will take place. It's theme is "tanabata", an old story related to a couple which is separated by the Milky way.
Translator 7	There is no special event. But you can enjoy a general Japanese festival.

Question III: What are these two? Can you provide links of pictures?	
Translator 1	1: chinese cabbage: [google image link] 2: sesame oil: [google image link]
Translator 2	The first one: [google image link] The second one: [google image link]
Translator 3	[google image link] and [google image link]
Translator 4	I don't know how to answer.
Translator 5	The first word : This is a vegetable. The second word : This is a oil.
Translator 6	(No response)
Translator 7	The first one is a vegetable whose shape is like elongated cabbage. And the second one is oil made from sesame. The latter may be used for seasoning.

Question IV: What is the difference between 1 and 2 in my electricity bill?	
Translator 1	I cannot answer this question.
Translator 2	According to Tokyo Electric Power Company(Tokyo-denryoku), an unit amount of electricity is defined as following: Level-1 -> 17.87yen/kWh, applied up to 120kWh Level-2 -> 22.86yen/kWh, applied more than 120 to 300kWh.
Translator 3	Sorry, but I don't know.
Translator 4	1: basic charge. 2: metered fee.
Translator 5	1 is basic fee. 2 is the amount fee of electronic energy.
Translator 6	I do not know how to answer this question.
Translator 7	I have no idea.