A General-purpose Crowdsourcing Platform for Mobile Devices

Ariel Amato[1], Felipe Lumbreras[2,3] and Angel D. Sappa[3,4]

[1]*Crowdmobile S.L. (Knowxel), Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain*

[2]*Dept. of Computer Science, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain*

[3]*Computer Vision Center, Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain*

[4]*Escuela Superior Politécnica del Litoral (ESPOL), Campus Gustavo Galindo, Km 30.5 vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador*

Keywords: Crowdsourcing Platform, Mobile Crowdsourcing.

Abstract: This paper presents details of a general purpose micro-task on-demand platform based on the crowdsourcing philosophy. This platform was specifically developed for mobile devices in order to exploit the strengths of such devices; namely: *i*) massivity, *ii*) ubiquity and *iii*) embedded sensors. The combined use of mobile platforms and the crowdsourcing model allows to tackle from the simplest to the most complex tasks. Users experience is the highlighted feature of this platform (this fact is extended to both task-proposer and task-solver). Proper tools according with a specific task are provided to a task-solver in order to perform his/her job in a simpler, faster and appealing way. Moreover, a task can be easily submitted by just selecting predefined templates, which cover a wide range of possible applications. Examples of its usage in computer vision and computer games are provided illustrating the potentiality of the platform.

1 INTRODUCTION

The large amount of mobile devices, together with the ever increasing computation capabilities, offers to the crowdsourcing model a powerful tool for tackling a large set of problems where the human feedback is needed. In most of the urban areas, commuters expend large amount of time travelling or waiting from place to place. Most of this fragmented time is wasted with meaningless activities; the current work presents a tool that takes advantage of that fragmented time; everybody and anywhere can contribute with his/her small piece of time and work using the proposed platform.

Current 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. On the other hand, although there are some proposals already devoted to exploit the combined use of mobile and crowdsourcing they are intended to a specific task. We can find applications for image translation (Liu et al., 2010), drawing (Liu et al., 2012), collaborative sensing (Demirbas et al., 2010), text transcription and surveys (Eagle, 2009), just to mention a few. On the contrary to previous proposals the current paper introduces a general purpose crowdsourcing platform oriented to mobile devices, where the users can find different kind of tasks that engage their participation.

During last decade crowdsorucing solutions have been adopted in a large set of problems in the computer vision domain. For instance in (Vondrick et al., 2010) a user interface has been proposed for dividing the work of labeling video data into micro-tasks. A specific architecture that take care of minimizing bandwidth overhead is developed. This user interface has evolved to a more complete framework that has been recently presented and evaluated in (Vondrick et al., 2013). Applications such as nutrition analysis from food photographs (Noronha et al., 2011), facial expression and affect recognition (McDuff et al., 2011), image annotation (Moehrmann and Heidemann, 2012) and multimedia retrieval (Snoek et al., 2010) have exploited the power of massive data categorization. All these approaches are oriented to a specific and single task hence all of them rely on developing a standalone tool; by sure that tools devel-

[1]https://www.mturk.com/

Amato A., Lumbreras F. and Sappa A..
A General-purpose Crowdsourcing Platform for Mobile Devices.
DOI: 10.5220/0004737202110215
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 211-215
ISBN: 978-989-758-009-3

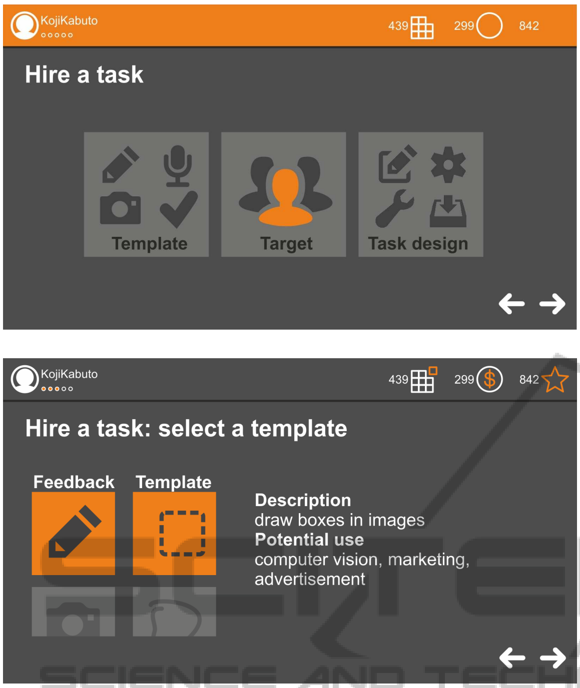

Figure 1: Screenshots of the main window (*top*) and template selection (*bottom*) during the hiring a task process.

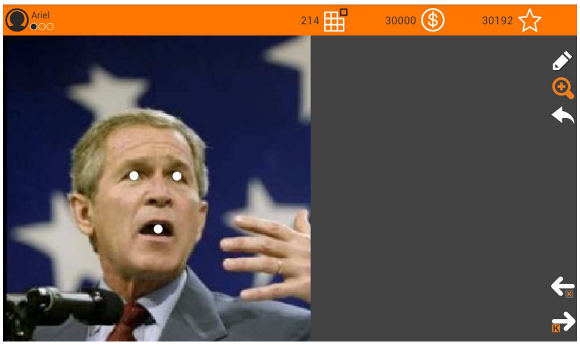

Figure 2: Working space with image edition tools in a *face alignment* problem (mark eyes and mouth) http://vis-www.cs.umass.edu/lfw/index.html.

oped in this way are the best option from the HCI and efficiency point of view. However, neither HCI nor efficiency are the most critical elements in a crowdsourcing platform, but having a large crowd is the key point. In other words, in most of the work presented above (except those based on the use of services like Amazon Mechanical Turk) after developing the tool a big effort for having a large crowd of workers is needed, which sometimes become critical for a successful result.

Having in mind the key point mentioned above the current work presents some details of a general purpose crowdsourcing platform, which was initially conceived for computer vision problems, but recently found applications in other domains. The manuscript is intended to share our experiences on developing and using such a platform. Section 2 briefly presents the platform. Then, Section 3 introduces two case studies and finally Section 4 summarize conclusions and future work.

2 THE PLATFORM

This section summarizes the most representative characteristics of the proposed platform. The whole platform has been implemented to be used in mobile devices taking advantage of all their capabilities. The task management and user profile administration is implemented in a database management system available through local servers. The most relevant elements of our platform are presented here. First the template based task proposal is illustrated and then the user interface for solving a task is depicted.

2.1 Templates for Task Proposing

The platform is conceived in a Client-Server philosophy for both the client that hire a task and the users that provide the feedback for a required task using his/her own mobile device and the provided tools. Note that since everything is intended for mobile devices the embedded sensors can be used. There is a large set of templates already designed to use with the different sensors embedded on mobile devices; this factor allows a large scalability and simplicity. Figure 1 shows the main screen with the different options for hiring a task. First the template according to the required feedback is selected (e.g., cameras, GPS, microphone, speaker, touch screen); then, the target population is defined; and finally the task is designed in a fill-in-a-form scheme. A brief summary of the most representative templates is given below.

Draw: these templates exploit the potentiality of touch-screen devices. The objective of tasks based on these templates is to draw something over a specific object on the given image (e.g., detect where the object is, draw the contour of a given object, etc.), which in the computer vision literature is usually referred to as image segmentation.

Annotate: templates related with the previous ones (draw); in this case the templates allow obtaining the feedback from the user for a semantic description of the object in the scene.

Survey: a set of predefined templates that can be easily customized according to the need of the client. These templates include the possibility to create classical text based surveys, image and sound based sur-

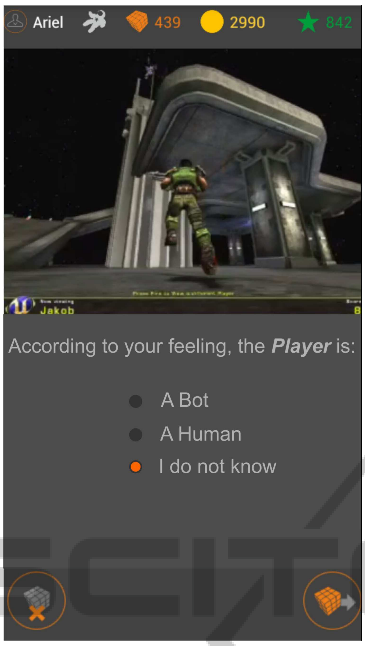

Figure 3: Illustration of a video-based survey for a *Bot competition* http://botprize.org.

veys as well as video based surveys.

Convert: this family of templates is intended for converting the way in which the given data are expressed; examples of such conversions are: audio to text (listen a sound and transcribe it), text to audio (read a book), image to text (transcribe the information in a given image), etc. Additionally, these templates can be used for high level feedbacks whether the user needs to have some expertise for converting the given data (e.g., language translation tasks: text to text, audio to text, text to audio).

Capture: these final templates allow the companies and academia to collect information using mobile devices embedded sensors, from an audio, picture or video till GPS positions.

2.2 User Interface for Task Solving

Some snapshots of the user interfaces for solving different tasks are presented in this section. Figure 2 presents the working space of a face alignment problem with the basic tools (zoom in/out, edit, delete). In this task the user has to mark eyes and mouth that are used for estimating the face alignment. Figure 3 depicts a screenshot of a video-based survey; the user has to watch a short video and answer the question.

Redundancy: Like in most of the tasks involving the human being, there is a need to detect wrong inputs. The platform allows a redundancy scheme to filter out outliers in most of the computer vision based tasks (e.g., image segmentation, data categorization, target selection, etc.). Further elaborations are needed to detect outliers or wrong inputs in tasks such as data collection or surveys.

Figure 4: Original page to be processed (text segmentation).

Reward: Current version of the platform relies on volunteer effort; however, like in other crowdsoruc-ing initiatives financial incentives of workers can be easily implemented. There is a relationship between monetary payment and time to perform a given set of tasks. Larger incentive makes more appealing the tasks and as a result less time to complete them is needed. A deep study on all these effects has been presented in (Mason and Watts, 2009). For instance, the authors conclude that the accuracy of results are similar when they compare payment versus no payment at all.

3 CASE STUDIES

This section describes two applications (tasks) that have been conducted in the proposed platform. The first one is focused on ground truth generation of handwriting documents. The purpose of the second one is to obtain users' feedbacks for video game developers.

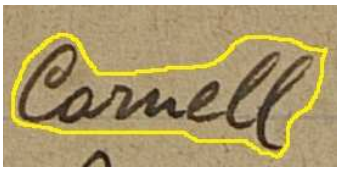

Figure 5: Example of the obtained final word segmentation.

3.1 Ground Truth Generation for Ancient Documents in Computer Vision Context

For historical documents, ground truth may not exist, or its creation may be tedious and costly, since it has to be manually performed. In such a labor, crowdsourcing has become very popular in the last years, because only with the massive help of volunteers, huge amounts of data can be manually labeled. The conducted experiment was focused on the Marriage Licenses Books (Romero et al., 2013) from the Archives of the Barcelona Cathedral. These books contain details of more than 500.000 unions celebrated between the 15th and 19th centuries. One page of the original manuscript is shown in Fig. 4. The final goal of this task was to obtain the bounding contour of every single word of the page, without overlapping neighbor words (see Fig. 5).

The process follows a recursive task atomization scheme (Amato et al., 2013). The initial task consists in extracting the page layout. This task results in columns, see Fig. 6. The outputs of this first task allow the platform to release a second task; splitting these columns of text into several boxes containing on average 6 lines of text to properly display the task in several devices (from 3.4” to 10.1” screen sizes). In the second task, the user is asked to draw the bounding box of every single word. Later on, this output is used to generate the next task. Finally, in the last task the user is asked to precisely segment the word contained in the given box (see right side of Fig. 7).

3.2 Video-based Survey for Video Game Competitions

In this task the users were asked to watch ten videos. Each one of these videos contains one minute of a recorded video game sequence. In turn, the solvers were asked to answer if the player behavior shown on the video corresponds to a human or a bot. The purpose of this task was to evaluate the performance of artificial intelligent engines on different video games. A snapshot of this task is presented in Fig. 3

Figure 6: Page layout obtained with the platform.

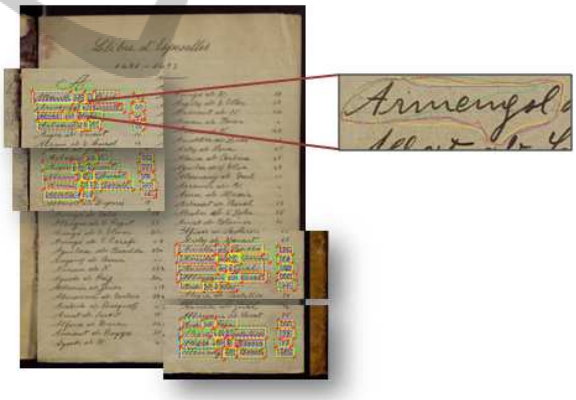

Figure 7: Illustration of the segmentation result.

4 CONCLUSIONS

The manuscript briefly presents a novel general purpose platform for crowdsourcing tasks, oriented to mobile devices. The general idea and some of the applications tackled using this scheme are introduced showing its potentiality and possibilities, mainly in the computer vision domain. Currently, the platform is been used by a small and controlled community (50 *solvers*) in order to evaluate efficiency, task atomization strategy and validation results. Future steps will be addressed to increase the size of the community in order to release a commercial version of the platform.

ACKNOWLEDGEMENTS

This work has been partially supported by the Spanish Government under Research Projects TIN2011-25606, TIN2011-29494-C03-02 and PROMETEO Project of the ”Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación de la República del Ecuador”.

REFERENCES

Amato, A., Sappa, A. D., Fornés, A., Lumbreras, F., and Lladós, J. (2013). Divide and conquer: atomizing and parallelizing a task in a mobile crowdsourcing platform. In *Proceedings of the 2Nd ACM International Workshop on Crowdsourcing for Multimedia*, CrowdMM ’13, pages 21–22.

Demirbas, M., Bayir, M. A., Akcora, C. G., and Yilmaz, Y. S. (2010). Crowd-sourced sensing and collaboration using twitter. In *IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM)*, pages 1–9.

Eagle, N. (2009). txteagle: Mobile crowdsourcing. In *In Internationalization, Design and Global Development, Volume 5623 of Lecture Notes in Computer Science*. Springer.

Liu, Y., Lehdonvirta, V., Alexandrova, T., and Nakajima, T. (2012). Drawing on mobile crowds via social media. *Multimedia Systems*, 18(1):53–67.

Liu, Y., Lehdonvirtay, V., Kleppez, M., Alexandrova, T., Kimura, H., and Nakajima, T. (2010). A crowdsourcing based mobile image translation and knowledge sharing service. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*.

Mason, W. A. and Watts, D. J. (2009). Financial incentives and the ”performance of crowds”. *SIGKDD Explorations*, 11(2):100–108.

McDuff, D., el Kaliouby, R., and Picard, R. (2011). Crowdsourced data collection of facial responses. In *Proceedings of the 13th international conference on multimodal interfaces*, ICMI ’11, pages 11–18.

Moehrmann, J. and Heidemann, G. (2012). Efficient annotation of image data sets for computer vision applications. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA ’12, pages 2:1–2:6, New York, NY, USA. ACM.

Noronha, J., Hysen, E., Zhang, H., and Gajos, K. Z. (2011). Platemate: crowdsourcing nutritional analysis from food photographs. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST ’11, pages 1–12.

Romero, V., Fornés, A., Serrano, N., Sánchez, J. A., Toselli, A. H., Frinken, V., Vidal, E., and Lladós, J. (2013). The esposalles database: An ancient marriage license corpus for off-line handwriting recognition. *Pattern Recognition*, 46(6):1658–1669.

Snoek, C. G. M., Freiburg, B., Oomen, J., and Ordelman, R. (2010). Crowdsourcing rock n’ roll multimedia retrieval. In *Proceedings of the 18th International Conference on Multimedia, Firenze, Italy, October 25-29*, pages 1535–1538.

Vondrick, C., Patterson, D., and Ramanan, D. (2013). Efficiently scaling up crowdsourced video annotation - a set of best practices for high quality, economical video labeling. *International Journal of Computer Vision*, 101(1):184–204.

Vondrick, C., Ramanan, D., and Patterson, D. (2010). Efficiently scaling up video annotation with crowdsourced marketplaces. In *11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11*, pages 610–623.