Using Games to Solve Challenging Multimedia Problems

Oge Marques

Abstract There are many challenging problems in multimedia research for which state-of-the-art solutions fall short of performing perfectly. The realization that many of these tasks are arduous for computers yet are relatively easy for humans has inspired many researchers to approach those problems from a human computation viewpoint, using methods that include crowdsourcing and games. This paper offers a discussion and advices how to use games to supplement traditional content analysis techniques and assist in the solution of hard multimedia problems.

Keywords Games · Multimedia · Crowdsourcing · Human computation · Computer vision · Image processing

1 Introduction

There are many contemporary problems in multimedia research for which fully automatic solutions are still not available, and the performance of state-of-the-art algorithms is far inferior to humans performing comparable tasks [8]. In the past few years, following the success of the ESP game [11] and its ability to assist in the solution to the problem of image tagging (i.e., associating keywords to an image that describe its contents), a growing number of research projects in the intersection between games and multimedia have emerged. These so-called serious games (or "games with a purpose") belong to a category of approaches that include human users in the loop and attempt to extract valuable information from the user input, which can then inform and improve the underlying algorithms designed to solve a particular problem. In addition to games, human input (and knowledge) can also be leveraged through the use of *crowdsourcing*, where inputs from large numbers of human participants are processed to serve as a basis for statistical analysis and inference [8].

O. Marques (✉)
Florida Atlantic University, Boca Raton, FL 33431, USA
e-mail: omarques@fau.edu

Games and crowdsourcing belong to the broader field of *human computation*, whose main idea consists of collecting users' input and interactions and using them to progressively refine a computational model designed to solve a particular problem. In crowdsourcing efforts, such input is captured in the form of very specific tasks, for which users are often paid a modest amount. In games, the goal is to keep the user entertained and—after the game has been played a significant number of times—mine information from the game logs and use them to feed the associated computational model.

In this paper, we postulate that games—if properly designed—have the potential to assist in the solution of challenging multimedia research problems. We note, however, that there are some problems with the vast majority of "gamification" approaches to the solution of research problems in the field of multimedia proposed over the past decade. More importantly, we offer advice for researchers who want to overcome those problems and build successful games for solving relevant scientific problems.

2   Background and Motivation

The success of using games to solve scientific problems in the field of image analysis, computer vision, and visual scene understanding can be traced back to the pioneering work of Luis von Ahn and his *Extra Sensory Perception (ESP)* game [11]. This game was eventually made more popular in the form of a Google Labs project (Google Image Labeler) and has helped collect an enormous amount of useful tags that describe the contents of a large collection of images crawled from the Web at large. The ESP game involves two players, matched at random, which are shown the same image and asked to type words that they believe could be used as tags to describe the image contents. When the game finds an agreement between the words used by both players, they are rewarded with points. A label (or tag) is accepted as *correct* when different pairs of users agree about it.

The ESP game has inspired many efforts, including ARTigo, a Web-based platform (http://www.artigo.org) containing six artwork annotation games as well as an artwork search engine in English, French, and German. Funded by the German Research Foundation (DFG), the ARTigo project has—during the period between 2008 and 2013—successfully collected over 7 million tags (mostly in German) to describe the artworks in the collection and engaged more than 180,000 players (about a tenth of whom are registered). It remains active at the time of this writing.

The seminal work by von Ahn also coined the expression "games with a purpose" and its acronym, GWAP. Until 2011, the research group responsible for ESP and several other games used to maintain a Website (http://www.gwap.com), which contained links to a variety of games developed to assist in the solution of specific problems, such as *Peekaboom* [13] for object detection, *TagATune* [5] for music and sound annotation, and *Verbosity* [12], a game for collecting common sense knowledge for the Open Mind Common Sense project. More recently, von Ahn has devoted

most of his efforts to create, improve, and promote Duolingo (http://www.duolingo.com), an extremely successful multi-platform app that gamifies the process of learning foreign languages.

Outside of the realm of multimedia research, prominent examples of games designed to solve scientific problems include FoldIt (http://fold.it) (a 3D puzzle used to assist in understanding protein folding and potentially leading to the discovery of cures for diseases such as AIDS and cancer) and EyeWire (http://eyewire.org), a game for finding the *connectome*—a comprehensive map of neural connections—of the retina.

Games (with a purpose) remain a valid, meaningful, and not yet fully explored avenue for assisting in the solution of challenging scientific problems. Games can be particularly useful when the research question at hand that can be mapped to *tasks* that satisfy one or more of the following criteria: (i) are easy for humans and hard for computers; (b) require intensive labor; (iii) enable noble scientific pursuits; and (iv) improve human life.

3   Our Work

In this section, we present a brief description of three of our recent research projects involving the use of games and crowdsourcing for solving challenging multimedia research problems, namely

- Ask'nSeek: A two-player game for object detection, segmentation, and labeling.
- Click'n'Cut: An interactive intelligent image segmentation tool for object segmentation.
- Guess That Face: A single-player game for face recognition under blurred conditions.

At the end of the section, we share some of the lessons we have learned from these projects.

3.1   Ask'nSeek

AsknSeek [2] is a two-player Web-based guessing game that asks users to guess the location of a hidden region within an image with the help of semantic and topological clues. One player (*master*) hides a rectangular region somewhere within a randomly chosen image, whereas the second player (*seeker*) tries to guess the location of the hidden region through a series of successive guesses, expressed by clicking at some point in the image. Rather than blindly guessing the location of the hidden region, the seeker asks the master for clues (i.e., indications) relative to the objects present in the image (e.g., "above the cat" or "partially on the dog" in Fig. 1). The game is

Fig. 1 Examples of object
detection and labeling results
obtained with AsknSeek:
two objects (*cat* and *dog*)
were detected and labeled
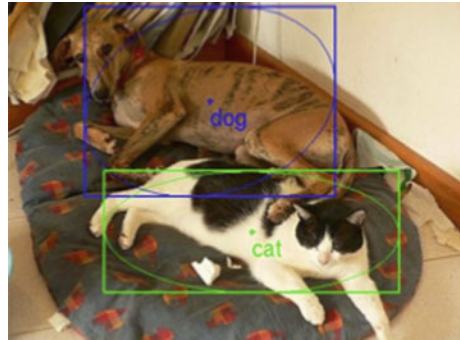
cooperative—i.e., both players score more points when the hidden region is found
quickly—which leads to the master providing accurate clues, which are stored into
the game logs for further processing.

The information collected from game logs is combined with the results from con-
tent analysis algorithms and used to feed a machine-learning algorithm that outputs
the outline of the most relevant regions within the image and their names (Fig. 1). The
approach solves two computer vision problems—object detection and labeling—in
a single game and, as a bonus, allows the learning of spatial relations (e.g., "the dog
is above the cat") within the image [8].

3.2 Click'n'Cut

Ask'nSeek was extended to address the object segmentation problem [9], where the
game traces are combined with the ranked set of segments generated by the con-
strained parametric min-cuts (CPMC) algorithm [4].

In a parallel effort, an intelligent interactive tool for foreground object segmen-
tation (*Click'n'Cut*) was created, where the users are asked to produce foreground
and background clicks to perform a segmentation of the object that is indicated in a
provided description (Fig. 2). Every time a user produces a click, the segmentation
result is updated and displayed over the image with an alpha value of 0.5. This seg-
mentation is computed using an algorithm based on object candidates [1] and aims
at guiding the user to provide information (i.e., meaningful clicks) that will help
improve the quality of the final segmentation result.

Several experiments were performed using the Click'n'Cut tool on a set of 105
tasks (100 images to be segmented plus 5 gold standard tasks, to control for errors),
with two distinct groups of users:

- Experts: 15 computer vision researchers from academia, both students and pro-
  fessors.
- Workers: 20 paid workers from the platform https://microworkers.com/; each
  worker was paid 4 USD for annotating 105 images.

Click'n'Cut (1/105)

Extract the fish.

Left click on the Foreground
Right click on the background
To reset your clicks, please click "Clear Points"
Click on any point to remove it
Use the slider to modify the mask transparency
Once you are satisfied with the mask, click 'Done' to go to the next task

Clear Points

Transparency

Show points ● Yes ○ No

Done

Fig. 2 Click'n'Cut user interface

The results obtained by each group were also compared against the segmentation output produced by using clicks collected from 162 Ask'nSeek players (mostly students) who played the Ask'nSeek game on any number of images they wanted to.

The goal of such experiments was to assess: (i) the "crowdsourcing loss" between experts and less-skilled workers; and (ii) the "gamification loss" incurred by replacing an interactive crowdsourcing tool (Click'n'Cut) with a game (Ask'nSeek) as a generator of foreground and background clicks. Detailed results and discussions can be found in [3].

3.3 Guess That Face!

Guess That Face! [7] is a single-payer Web-based face recognition game that reverse engineers uses the human biological threshold for accurately recognizing blurred faces of celebrities under time-varying conditions (Fig. 3). The game combines a successful casual game paradigm with meaningful applications in both human- and computer-vision science. Results from preliminary user studies conducted with 28 users and more than 7,000 game rounds supported and extended preexisting knowledge and hypotheses from controlled scientific experiments, which show that humans are remarkably good at recognizing famous faces, even with a significant degree of blurring [10]. A live prototype is available at http://tinyurl.com/guessthatface.
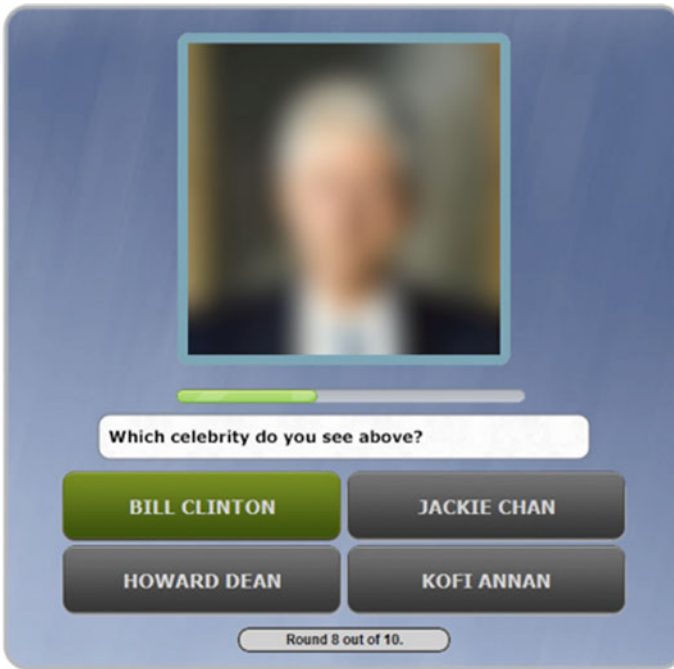
Fig. 3 Guess That Face! screenshot: The player is presented with a severely blurred image (de-blurring progress is indicated by the *green bar* below the image) and four options (*buttons*). Here, the correct answer has already been chosen (*green*)

3.4 Lessons Learned

The experience with the games and crowdsourcing projects described in this section has taught us several valuable lessons that we could not have anticipated if we had not developed and deployed the games and tools and conducted associated user studies and experiments, among them:

- Most of our users played the game for *extrinsic*—rather than *intrinsic*—reasons; for example, the highest response rates for *Ask'nSeek* came as a result of assigning bonus points in a course for students who played 30 or more rounds of the game.
- On a related note, for *Guess That Face!*, a different reward system was adopted, in which students placed in the "High scores" table were assigned extra bonus points. As a result, students played many more games than expected, driven by intrinsic reasons—placing themselves among the top scorers and beating their friends' scores. As a side effect of playing the game for many more rounds than originally expected, some players ended up "memorizing" the dataset, which led to additional work when processing and cleaning up the game logs.
- None of our games has reached a level of engagement remotely close to "viral" or "addictive."

4 Discussion

4.1 Problems

In this paper, I postulate that most of the efforts aimed at creating games for assisting in the solution of multimedia research problems suffer from two main problems:

1. The design process is often reversed, i.e., rather than designing a game with the player in mind, researchers (including the author) usually follow these steps: (i) start from a problem; (ii) think of a crowdsourcing solution; (iii) create a tool; and (at the end) (iv) "make it look like a game."
2. Our terminology is not exactly inspiring: expressions such as "serious games," "Games with a purpose (GWAP)," "Human-based computation games," or "Non-entertainment focused games" do not convey the idea that such games can (and should!) be fun to play.

4.2 Possible Solutions

I was recently asked to give a talk on this topic and provide advice for (young) researchers who are interested in the intersection of games and multimedia research. While reflecting upon the message I wanted to convey to the audience, I realized that the worst advice I could give was "Gamify everything!" Gamification has been overused and misused during recent years; it should not be seen as a "cure-all" solution, but rather a creative way to engage users in meaningful human computation tasks, while being driven by the intrinsic motivation evoked by a properly designed game.

These are some other pieces of advice that I hope will be helpful to readers of this paper:

- Do not try to gamify if you cannot see the world from a gamer's viewpoint. If you are not a gamer, learn more about it from resources such as the excellent book "Getting Gamers: The Psychology of Video Games and Their Impact on the People who Play Them" [6] and companion site (http://www.psychologyofgames.com) by Jamie Madigan.
- Select multimedia problems that are worth researching and can be modeled as tasks that fulfill at least two of the criteria stated earlier in this paper (and repeated here for convenience): (i) are easy for humans and hard for computers; (b) require intensive labor; (iii) enable noble scientific pursuits; and (iv) improve human life.
- Be mindful of (and try to incorporate, whenever possible) new devices and technologies, such as increasingly popular virtual reality (VR) kits, new sensors, and wearable devices and gadgets.

- Consider engaging in research on game effectiveness, e.g., creating experiments to find out if people are having fun while playing a certain game.
- Challenge the design workflow described in Sect. 4.1 (and turn it upside-down!)

5   Conclusion

In this paper, I have discussed the intersection between (serious) games and multimedia research and provided advice on how to use games to supplement traditional content analysis techniques and assist in the solution of hard multimedia problems.

As stated in an earlier paper [8], I believe that games should be designed such that the input collected from the users is as simple as possible, but carries as much meaningful information as possible. Moreover, we should not ignore the traditional approach of content analysis, which can be used to augment crowdsourcing, in order to reduce the number of participants needed to obtain meaningful results, using semi-supervised machine-learning approaches.

As a final reflection, for those readers who might want to consider engaging in this field of research and studies, a simplified SWOT analysis might be helpful:

- Strengths

    - There are many meaningful research problems waiting to be solved.
    - People love games!

- Weaknesses

    - Poorly designed games turn people away (quickly!).

- Opportunities

    - There are multiple game platforms to develop for—from Web to traditional consoles to mobile apps for iOS and Android devices.
    - There is a growing interest in games, and the trend is not likely to change any time soon.

- Threats

    - There may be better solutions for certain multimedia research problems (e.g., the increasingly popular use of *deep learning* techniques) that do not use games.

References

1. P. Arbeláez, J. Pont-Tuset, J. T. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping. In *CVPR*, 2014.
2. A. Carlier, O. Marques, and V. Charvillat. Ask'nseek: A new game for object detection and labeling. In *Computer Vision–ECCV 2012. Workshops and Demonstrations*, pages 249–258. Springer, 2012.
3. A. Carlier, A. Salvador, F. Cabezas, X. Giro-i Nieto, V. Charvillat, and O. Marques. Assessment of crowdsourcing and gamification loss in user-assisted object segmentation. *Multimedia Tools and Applications*, pages 1–28, 2015.
4. J. Carreira and C. Sminchisescu. Constrained parametric min-cuts for automatic object segmentation. In *CVPR*, 2010.
5. E. L. Law, L. Von Ahn, R. B. Dannenberg, and M. Crawford. Tagatune: A game for music and sound annotation. *ISMIR*, 3:2, 2007.
6. J. Madigan. *Getting gamers: the psychology of video games and their impact on the people who play them*. Rowman & Littlefield, Lanham, Maryland, 2016. ISBN 978-1-4422-3999-9.
7. O. Marques, J. Snyder, and M. Lux. How well do you know tom hanks?: Using a game to learn about face recognition. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 337–342, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1952-2. doi: 10.1145/2468356.2468416. URL http://doi.acm.org/10.1145/2468356.2468416.
8. W. T. Ooi, O. Marques, V. Charvillat, and A. Carlier. Pushing the Envelope: Solving Hard Multimedia Problems with Crowdsourcing. *E-letter COMSOC Multimedia Communications Technical Committee*, 8(1):37–40, 2013. URL http://oatao.univ-toulouse.fr/13004/.
9. A. Salvador, A. Carlier, X. Giro-i Nieto, O. Marques, and V. Charvillat. Crowdsourced object segmentation with a game. In *Proceedings of the 2nd ACM international workshop on Crowdsourcing for multimedia*, pages 15–20. ACM, 2013.
10. P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, Nov 2006. ISSN 0018-9219. doi: 10.1109/JPROC.2006.884093.
11. L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 319–326, New York, NY, USA, 2004. ACM. ISBN 1-58113-702-8. doi: 10.1145/985692.985733. URL http://doi.acm.org/10.1145/985692.985733.
12. L. von Ahn, M. Kedia, and M. Blum. Verbosity: A game for collecting common-sense facts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 75–78, New York, NY, USA, 2006a. ACM. ISBN 1-59593-372-7. doi: 10.1145/1124772.1124784. URL http://doi.acm.org/10.1145/1124772.1124784.
13. L. von Ahn, R. Liu, and M. Blum. Peekaboom: a game for locating objects in images. In *ACM CHI*, 2006b.