HUMAN COMPUTATION GAMES: A SURVEY

Markus Krause and Jan Smeddinck

Research Group Digital Media, TZI, University of Bremen
Bibliothekstr. 1, 28359, Bremen, Germany
phone: +49 (0)421 218-64417, fax: +49 (0)421 218-8751, email: {phaeton, smeddinck}@tzi.de
web: http://dm.tzi.de/en/

ABSTRACT

Motivation has been one of the central challenges of human computation. A promising approach is the integration of human computation tasks into digital games. Different human computation games have been successfully deployed, but tend to provide relatively narrow gaming experiences. This survey discusses various approaches of digital games for human computation and aims to explore the ties to signal processing and possible generalizations.

1. INTRODUCTION

Human computation combines human mental abilities and computational systems at a large scale. Even though human computation projects such as reCAPTCHA [4] show that the power of a human computation grid can be substantial, it remains a challenge to motivate humans to contribute. Different strategies have evolved to approach this issue. One promising solution is to use a new kind of serious games. The basic idea of these human computation games is the integration of human computation tasks into digital games. Therefore players solve problems by playing the game. More generally, human computation games let humans create data for computational systems. That data presents the result of complex computations performed by the human contributors. The following section 2 gives an overview of current human computation games with a focus on visual or audio processing and describes interesting game designs.

Since various fields of computer science are dealing with human computation and crowdsourcing approaches, current literature on these topics is found in various areas such as Human Computer Interaction, Artificial Intelligence, Data Mining, Computer Graphics, and Audio Processing. Projects that investigate the paradigms of human computation and crowdsourcing from different perspectives frequently are not aware of each other. This is problematic for researchers trying to understand the current state of the art within this field [1]. The aim of this work is to provide an overview of the current human computation landscape. Therefore, it identifies four common challenges of human computation systems. The first relevant aspect describes the problem spaces which can be handled with human computation systems. This aspect is described in general in section 3. The second common aspect is concerned with the collection of valuable data from the contributors by observing their actions. The section 4 gives an overview of this aspect. Another crucial aspect of every human computation system is data reliability. Humans are expected to be unreliable in unsupervised situations and generate false data either on purpose or for other reasons. This is especially true in gaming situations where players tend to explore the system. This exploration includes giving false answers to see how the system reacts. The Section 5 explains different strategies to enhance data quality in human computation games. As mentioned before, a central challenge for human computation is to motivate contributors to participate in a human grid. The section 6 describes possible strategies to use digital games and their mechanics to motivate people to contribute.

2. STATE OF THE ART

A common way to get contributors to participate in a human computation grid is to use extrinsic motivation. Extrinsic motivation comes from an outer source, such as granting access to special web resources, or simply by paying the contributors. Systems such as *Mechanical Turk*, *Microtask*, or *Crowdflower* allow customers to upload small tasks such as reviewing a website or tagging images or sound files. The customers then pay other users, so-called workers, to solve these tasks. Another project that uses motivation not through the system itself is reCAPTCHA. This project serves the protection of publicly available web services from abuse by automated systems. A typical reCAPTCHA is an image containing several distorted characters. Users type these characters to prove that they are indeed human. The system generates these images from scanned documents. The solutions entered by humans improve the digitization process [2]. In contrast to digital human computation games, where players are motivated to spend cognitive effort wholly out of their own interest, all these systems provide motivation through secondary elements.

Even though, the systems and services mentioned above are easy to use and the implementation of tasks is relatively simple, other projects demonstrate the power of digital games in the domain of human computation. Common tasks for human computation games are relation learning or resource labeling. Well-known examples in this regard come from the *Games with a Purpose (GWAP)* series [3]. It consists of puzzle games for different purposes. *ESP* [4] for instance, aims at labeling images. The game pairs two users over the internet. It shows both players the same picture and lets them enter keywords that describe the content of that image. If both players agree on a keyword, they both score and the next picture is shown. *ESP* produced 1.3 million labels with around 13,000 players in a four month period. Another game of the *GWAP* series is *Verbosity* [5]. The main idea of this game is to collect commonsense knowledge about words. *Verbosity* lets the first player describe a word by commonsense knowledge. For instance let the word to describe be milk then the describer could type: *It is* white. Similar to other games two players are paired at random. The players take turns describing and guessing a word. The *Describer* provides words to describe the given input by using the provided templates (e.g., "*it looks like*" and "*it is a type of*"). The templates simplify the game play and provide a better control over the input. The *Guesser* in turn has to guess the input word based on the *Describer*'s outputs.

The Games *Squigle* [6] and *Peekaboom* [7], both let players, identify parts of images. Again two players are paired randomly. While one player gets an image along with a word related to the image the so called *Boom*. The other player called *Peek* gets no image. *Boom* reveals parts of the image to *Peek*. The goal is for *Peek* to type the word associated to the image. The *Boom* player reveals areas of the image so that *Peek* can guess the word associated to it. Once *Peek* guesses the correct word, the two players switch roles and play another round. In a one month period *Peekaboom* attracted over 14,000 different players. The game generated during this time over 1 Million entries. *KissKissBan* [8] is another interesting game for image annotation. In contrast to the previous ones the game involves a direct conflict between players. KissKissBan connect three players instead of two as the other ones. Two players became the couple they try to find consensual descriptions about an image. This mechanics works similar as the *ESP* game. The third player becomes the so called blocker. The blocker tries to prevent the couples from reaching consensus by defining taboo words for the couple.

Another interesting application domain of human computation games is in natural language processing (NLP) applications to enhance web search engines. One example is *Webpardy* [9], a game for the annotation of websites. It aims at gathering natural language questions about web page fragments from its players. The game is similar to the popular *Jeopardy* quiz. *Phrase Detectives* [10], tries to collect anamorphic annotated corpora through a web game. In this game players try to get high scores by submitting annotations to win different small prices. Yet another interesting field for human computation games is audio analysis. *HeardIt*, for instance, lets players annotate audio files in a playful environment. This game is a multiplayer game with at least 10 players at a time. The same music clip is played to all players. Different sub games ask players on their opinions about the music. These games ask the player to select a musical sub-genre or the most prominent instrument. After each round the player is awarded points for consensus with the Herd [11]. The game attracted 1049 players in its 2 week alpha phase which produces over 9,000 labels.

All games presented so far share a common human computation mechanic in that they pair players to verify the validity of the input through mutual agreement. Furthermore they also have similar game mechanics and form a distinct subcategory of puzzle games. Similarity of these games can also be found in the way they are designed. These games are designed to make a actual boring task more interesting. This is in current research called gameification. In addition to this design paradigm it is also possible to integrate a human computation task into an already successful game design. *OnToGalaxy* [12] for instance integrates human computation tasks such as ontology population or image labelling into a common game design. The game is similar to games such as Asteroids or Starscape. It attracted around 500 players in the first 10 hours of its release.

Apparent from the design style there are some examples of games that go beyond simple game mechanics and require substantial commitment by the players. *FoldIt* [13], for instance, is a game that presents simplified three-dimensional protein chains to players, and provides a score according to the predicted quality of the folding done by the player. All actions by the player are performed in a three dimensional virtual world. It requires training to solve complex open protein-puzzles which in turn requires a lot of commitment by the players. This sort of tasks would be expensive to get solved by paid workers, as the payment would need to reflect the required effort. A game of similar complexity is *Plummings* [14]. This game aims at reducing the critical path length of field programmable gate arrays (FPGA). Unlike other games, the task in this game is separated from its actual story. The game is about a colony of creatures called *Plummings* who need adequate air supply. By keeping the length of the air tubes as short as possible the player saves the colony from suffocation. As these projects show, the application domains for human computation are indeed versatile and many of them are related to the area of signal processing in general. In order to emphasize the potential challenges for games in this field, the following sections will describe the key aspects of human computation games.

3. PROBLEM SPACES OF HUMAN COMPUTATION

In general, four substantially different problem domains can be defined in which human computation is most useful: aesthetic judgment, making intuitive decisions, contextual reasoning, and free interaction with the physical world. However these categories are not necessarily exclusive of each other, meaning that some human computation tasks can match more than just one category.

3.1 Intuitive Decisions

Combinatorial optimization tasks are common problems in computer science. Many of these tasks, for instance packing problems, are known to be NP-hard (nondeterministic polynomial time) [15]. In some cases, humans are able to solve these problems in an intuitive manner as Corney demonstrated for 2 dimensional packing problems [16]. Therefore, human computation allows for utilizing human mental abilities to find better solutions or algorithms in order to handle puzzle-such as combinatorial problems. Human computation systems such as *FoldIt* [13], *Plummings* [14], and others [16] are examples that show how to exploit this human talent to solve different NP-hard problems.

3.2 Aesthetic Judgment

The creation of computational systems that can handle tasks which require a human-level perception and understanding of aesthetics, like judging the quality of motion, a sound, or an image, remains a largely unsolved challenge. Humans are very good at interpreting various perceptions in this regard. This ability of aesthetic judgment makes tasks such as identifying unnatural motions or images easy for us. Aesthetics in this regard means perception by means of the senses and judgment means the interpretation of these impressions. Different approaches such as the systems of Talton [17] and Dawkins [18] explore this field. They make use of human aesthetic judgment in order to create natural looking lightning of virtual environments or to model objects in two and three-dimensional space.

3.3 Contextual Reasoning

Modern computer science often deals with human reasoning tasks. Solutions for these semantic problems require contextual information about the environment and the objects in it to resolve ambiguities and allow for useful results. Examples for the application domain of contextual reasoning are tasks such as image [4,8,19] or audio annotation [11,20,21].

3.4 Embodiment Issues

The ability of a computational system to act in the physical world is usually limited. Humans, of course, can easily interact with their physical environment. Examples utilizing human interaction with the physical environment are given by Matayas [22] and Tuite [23]. Both approaches try to reconstruct real world locations as detailed 3D models by analyzing a large quantity of photographic data. Tuite et al. designed a game called PhotoCity, which is played outdoors with a camera. Players take photos to capture flags and take over virtual models of real buildings.

4. OBSERVING INTERACTION

Every human computation system needs to provide a sound survey strategy to gather data from its contributors. Human computation systems generate useful data primarily by observing human behavior and interactions with the system. Observation strategies may vary. Some tasks need large amounts of data to be solved and others tasks are more complex and need lot of effort for a single solution. Labeling tasks for example gather large amounts of data. They are relatively simple but deal with a large number of objects, as in the web image labelling task in the ESP game [4]. These systems most often use a parallel observation process as described by Chilton et al. [24]. Parallel in this regard means that contributor performs individually and in parallel on a task. Iterative approaches let workers iteratively build on each other's results. Iterative approaches are most usefull for complex tasks that necessitate a strong commitment by the contributor, as for instance in FoldIt [13]. These systems do not solve a huge amount of tasks, but the complexity of each individual task is high.

5. DATA VALIDATION

Another common challenge of human computation games is data reliability. Humans are expected to be unreliable, especially in gaming environments were a playful interaction with the system to test its borders is expected. Therefore, players may generate false data either on purpose or for other reasons. Different strategies have evolved to deal with this issue. As human computation tasks are by definition not efficiently solvable by an algorithm, it is necessary to find strategies to handle this challenge. Two main strategies have evolved in recent projects the first one again relies on humans and the second strategy on machine learning algorithms. The human based approach pairs contributors. A given reply to a certain task is accepted only if contributors agree on the same answer at the time they play [6].

The machine based approach relies on calculating reliability of contributors, or using simple classifiers to accumulate multiple answers for a consensus. Common approaches using simple methods such as Majority Vote and Naïve Bayes as presented by [25]. Others use expectation maximization as shown by Ipeirotis et al. [26]. Other methods judge the quality of an answer given by a contributor, based on the reliability calculated for that contributor as presented by Krause et al. [27]. In some situations, a computational system can validate the quality of a given vote even though it is not capable of generating the original answer by itself. Examples are *FoldIt* and *Plummings*, which both calculate the quality of a given answer with an algorithm. In some situations the time required for the evaluation is also very crucial. In a real-time gaming situation, for instance, players often have to be informed about the quality of their actions as soon as possible. In such situations it can be advisable to split the evaluation into a fast evaluation that is used to generate a response to the player and a more in-depth follow-up evaluation to enhance data quality.

6. INCENTIVE GAMES

A human computation system has to deal with a fundamental challenge: motivating a human contributor to participate in the human grid. Many concrete strategies have evolved. They can be split into two groups according to their motivational strategies: intrinsic and extrinsic motivation. Intrinsic motivation is inherent in the system itself. This is usually the case in games, as games are played for the sake of playing them. Extrinsic motivation comes from an outer source, such as giving access to a certain web service. Because the focus of this paper is the aspect of digital games, this section will not go into detail about extrinsic motivation but take a closer look as games as an incentive for human computation.

The potential power of the concept of using digital games as incentive for human computation is enormous. A look at some statistics about games on the social network platform Facebook illustrates that very nicely. By September 2010 players on Facebook played more than 900 million hours per month [28]. Considering the labelling rate reported by von Ahn [4] this would be ~19 million labels per hour. One may argue that human computation games are a very special type of games and that they have repetitive game mechanics. On the other hand games such as FarmVille also have a repetitive game play and however were very successful. However, many human computation games were designed around the task they try to solve. The main idea thereby is to make a task more appealing by adding game elements. Approaches trying to make such boring things more interesting by applying game mechanics are recently called gameification. Therefore, as already argued most human computation games provide a valuable but specific sort of game experience.

Only few human computation games utilize common game designs. Some elements of game design which are very rare in human computation games are: progressive and competitive game play, as well as the use of story lines. All these elements are valuable for player immersion and motivation [12]. As digital games are played millions of hours every day it seems to be vital for human computation games to adapt common elements of these games. However, the design of a digital game is a complex and complicated task. For this reason it might be difficult to design human computation games comparable to current digital games with this approach. Instead to gameify a Task it is also possible to augment a normal game with a human computation task, as shown by *OnToGalaxy* [12]. This approach might be much easier then the design of completely new games. On the other the integration of human computation tasks is still a challenge as human computation frameworks such as *RABJ* (http://wiki.freebase.com/wiki/RABJ) or the *Dewknow* (http://www.dewknow.com) service are either research projects or not yet publicly available.

7. CONCLUSION

As various fields of computer science are dealing with human computation games, current literature on these topics is found in various areas of computer science. This is problematic for researchers trying to understand the current state of the art within this field. This work gave an overview of the key aspects of to human computation games. It introduced common concepts and challenges of the presented games, as well as the problem spaces that human computation games have already been successfully applied to, namely aesthetic judgement, intuitive decisions, contextual reasoning and interactions with the physical world. The second aspect described how the interaction between humans and human computation systems can be analysed to filter relevant data.

The third aspect provided an introduction to methods that can be used to verify the correctness of the gathered data. The methods introduced were human evaluation by human mutual agreement or calculating reliability values for contributors and algorithmic evaluation. The last aspect described the two main strategies to motivate contributors to participate in a human computation grid – intrinsic and extrinsic motivation. The examples presented herein showed that human computation can be applied to interesting problems in the field of signal processing such as audio or image analysis. Furthermore, it was shown that even NP-hard problems can sometimes be tackled using this paradigm. This work also emphasizes that digital games can be a valuable motivation strategy in this field, even though the design of such games still holds notable challenges.

ACKNOWLEDGEMENTS

The authors of this work were scholarship holders of the Klaus Tschira Foundation (KTF).

REFERENCES

[1] A. Quinn and B. Bederson, *Human Computation, Charting The Growth Of A Burgeoning Field*, College Park, Maryland, USA: 2010.

[2] L. von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum, “reCAPTCHA: human-based character recognition via Web security measures.,” *Science*, vol. 321, Sep. 2008.

[3] L. von Ahn, “www.gwap.com (Games with a Purpose),” 2010.

[4] L. von Ahn and L. Dabbish, “Labeling images with a computer game,” *CHI ’04 Proceedings of the 22nd international conference on Human factors in computing systems*, Vienna, Austria: ACM Press, 2004, pp. 319-326.

[5] L. von Ahn, M. Kedia, and M. Blum, “Verbosity: a Game for Collecting Common-Sense Facts,” *CHI ’06 Proceedings of the 24th international conference on Human factors in computing systems*, Montréal, Québec, Canada: ACM Press, 2006, p. 78.

[6] E. Law and L. von Ahn, “Input-agreement: a new mechanism for collecting data using human computation games,” *CHI ’09 Proceedings of the 27th international conference on Human factors in computing systems*, Boston, MA, USA: ACM Press, 2009, pp. 1197-1206.

[7] L. von Ahn, R. Liu, and M. Blum, “Peekaboom: a game for locating objects in images,” *CHI ’06 Proceedings of the 24th international conference on Human factors in computing systems*, Montréal, Québec, Canada: ACM Press, 2006, pp. 64-73.

[8] C.-J. Ho, T.-H. Chang, J.-C. Lee, J.Y.-jen Hsu, and K.-T. Chen, “KissKissBan: a competitive human computa-

tion game for image annotation," *HComp'09 Proceedings of the ACM SIGKDD Workshop on Human Computation*, Paris, France: ACM Press, 2009, pp. 11-14.

[9] H. Aras, M. Krause, A. Haller, and R. Malaka, "Webpardy : Harvesting QA by HC," *HComp'10 Proceedings of the ACM SIGKDD Workshop on Human Computation*, Washington, DC, USA: ACM Press, 2010, p. 4.

[10] J. Chamberlain, M. Poesio, and U. Kruschwitz, "Phrase Detectives: A Web-based Collaborative Annotation Game," *I-Semantics'08 Proceedings of the International Conference on Semantic Systems*, Graz: ACM Press, 2008.

[11] L. Barrington, D. O'Malley, D. Turnbull, and G. Lanckriet, "User-centered design of a social game to tag music," *HComp'09 Proceedings of the ACM SIGKDD Workshop on Human Computation*, Paris, France: ACM Press, 2009, pp. 7-10.

[12] M. Krause, A. Takhtamysheva, M. Wittstock, and R. Malaka, "Frontiers of a Paradigm - Exploring Human Computation with Digital Games," *HComp'10 Proceedings of the ACM SIGKDD Workshop on Human Computation*, Washington, DC, USA: ACM Press, 2010, p. 4.

[13] L. Bonetta, "New Citizens for the Life Sciences," *Cell*, vol. 138, 2009, pp. 1043-1045.

[14] L. Terry, V. Roitch, S. Tufail, K. Singh, O. Taraq, W. Luk, and P. Jamieson, "Harnessing Human Computation Cycles for the FPGA Placement Problem," *ERSA*, T.P. Plaks, ed., CSREA Press, 2009, pp. 188-194.

[15] L. Andrea, M. Silvano, and V. Daniele, "Two-dimensional packing problems: A Surrvey," *European Journal of Operational Research*, vol. 141, 2002, pp. 241-252.

[16] J.R. Corney, I. Kowalska, A.P. Jagadeesan, A. Lyn, H. Medellin, and W. Regli, "CrowdSourcing Human Problem Solving Strategy," *CrowdConf'10 Proceedings of the 1st International Conference on Crowdsourcing*, 2010.

[17] J.O. Talton, D. Gibson, L. Yang, P. Hanrahan, and V. Koltun, "Exploratory modeling with collaborative design spaces," *Proceedings of ACM SIGGRAPH Asia 2009*, vol. 28, 2009, pp. 1-10.

[18] D. Richard, *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design*, New York, NY, USA: W.W. Norton & Company, Inc., 1987.

[19] N. Diakopoulos and P. Chiu, "PhotoPlay: A collocated collaborative photo tagging game on a horizontal display," *UIST '07: Proceedings of the 19th annual ACM symposium on User interface software and technology*, Newport, USA: ACM Press, 2007, pp. 53-54.

[20] N. Diakopoulos, K. Luther, and I. Essa, "Audio Puzzler," *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, Vancouver, Canada: ACM Press, 2008, p. 865.

[21] Y.E. Kim, "MoodSwings : A Collaborative Game for Music Mood Label Collection," *ISMIR 2008*, 2008, pp. 231-236.

[22] S. Matyas, C. Matyas, and C. Schlieder, "Designing a Location-based Game for the Collection of Geospatial Data: The CityExplorer Case Study," *ACE '08: Proceedings of the International Conference on Advances in Computer Entertainment Technology*, Yokohama: ACM Press, 2008, pp. 244-247.

[23] K. Tuite, N. Snavely, D.-Y. Hsiao, A.M. Smith, and Z. Popović, "Reconstructing the world in 3D: bringing games with a purpose outdoors," *FDG'10 Proceedings of the Fifth International Conference on the Foundations of Digital Games*, Monterey, California, USA: ACM Press, 2010, p. 232–239.

[24] G. Little, L.B. Chilton, M. Goldman, and R.C. Miller, "Exploring iterative and parallel human computation processes," *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*, New York, New York, USA: ACM Press, 2010, p. 4309.

[25] A. Kumar and M. Lease, "Modeling Annotator Accuracies for Supervised Learning," *Methods*, 2011, pp. 19-22.

[26] P.G. Ipeirotis, F. Provost, and J. Wang, "Quality management on amazon mechanical turk," *HComp'10 Proceedings of the ACM SIGKDD Workshop on Human Computation*, Washington, DC, USA: ACM Press, 2010, pp. 0-3.

[27] M. Krause and J. Smeddnick, "Human Computation – A new Aspect of Serious Games," *Handbook of Research on Serious Games as Educational, Business and Research Tools: Development and Design*, M.M. Cruz-Cunha, ed., IGI Global, 2011.

[28] AllFacebook, "Facebook Games Statistics," 2010, p. http://www.allfacebook.com/facebook-games-statistics-2010-09.