

Using GWAP to Generate Informative Descriptions for Artwork Images on a Live Streaming Platform

Ngoc Cuong Nguyen
Graduate School of Information
Science and Engineering
Ritsumeikan University
Japan

Zhenao Wei
Graduate School of Information
Science and Engineering
Ritsumeikan University
Japan

Pujana Paliyawan
Ritsumeikan University Research
Organization
Ritsumeikan University
Japan

Hai V. Pham
School of Information and
Communication Technology
Hanoi University of Science and
Technology
Hanoi, Vietnam

Ruck Thawonmas
College of Information Science and
Engineering
Ritsumeikan University
Japan
ruck@is.ritsumei.ac.jp

Tomohiro Harada
College of Information Science and
Engineering
Ritsumeikan University
Japan

Abstract—This paper proposes a method for creating informative descriptive sentences for artwork images by using “Games With a Purpose,” or GWAP on a live streaming platform. In existing studies, automatic annotation of images did not perform well, in particular, for artwork images such as Ukiyo-e. On the other hand, existing studies on GWAP, the concept of using games for human to address problems that computer cannot solve, demonstrated how games can exploit human intelligence in labeling images. With a huge number of audiences watching games live streaming on Twitch and other similar services nowadays, we propose a solution that makes chatting of audiences become an effective means for generating valuable descriptions for artwork images, Ukiyo-e images in this study, as well as promoting the interaction between the audiences and the game.

Keywords—GWAP, Ukiyo-e, Distributed Knowledge Acquisition, Game Live Streaming, Game and Interactive Media

I. INTRODUCTION

Ukiyo-e is a genre of Japanese artwork which depicts female beauties, kabuki actors, Japanese landscape, etc., through woodblock prints and paintings. The theme was developed during the Edo period from the 17th through 19th centuries and is now popular around the world. Informative descriptive sentences of Ukiyo-e images have many benefits in deep learning research, image retrieval or improving accessibility for the visually impaired. However, most of Ukiyo-e images lack their content descriptions.

Generating an informative description for artwork images’ content is a difficult task for computers. On the other hand, human is capable to do so but needs motivations. GWAP [1] was conceived based on an idea of using games to solve problems that computers are unable to solve by utilizing human time and energy. The first successful implementation of GWAP was the ESP game, in which two random players are paired to guess what her or his partner types as keywords for an image, proposed by Luis von Ahn [2] to obtain meaningful and accurate labels for images on the Web. However, as pointed out by Steinmayr [3], such labels are generic and not specific enough to distinguish similar images. As a solution, Steinmayr proposed Karido, a game designed for obtaining specific labels for artwork images. Similarly, Harris [4] introduced ClueMeIn for obtaining detailed labels for arbitrary images.

The above games only focused on creating keyword labels that only describe main features of an image, not the content of the image: for instance, keywords for an image with the content of “a man walking his dog in a park” are simply “man,” “dog,” “trees,” which do not describe the content of the image or might even lead to a wrong impression of the content like “a wild dog is chasing a man in the woods.”. Another game named “Phetch” proposed by Luis von Ahn [5] aimed to generate sentences describing image contents on web pages for visually impaired people. Nevertheless, this game relied heavily on the image search engine in use, which searched an image based on its labels previously created by the ESP game.

We propose here a novel method for generating informative descriptive sentences for Ukiyo-e images via chatting on live game streaming. This would be more interesting for audiences when they are not only watching video game live streaming but also participating in playing the game we provide. Their chat messages can be acquired using API provided by Twitch. And the resulting descriptions can be automatically improved over time.

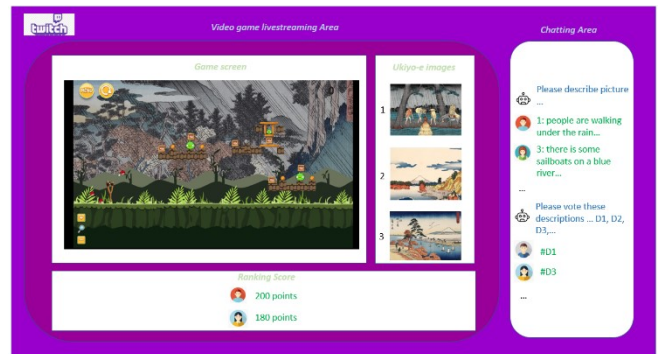


Fig. 1. Our GWAP on a Twitch live streaming site

The main contributions of this work are as follows:

- The first GWAP ever proposed for a live streaming platform (Twitch in this study).
- A proposed mechanism for generation of descriptive sentences for artwork images (Ukiyo-e in this work). Those sentences can also be further segmented into keywords for other tasks, such as keyword-based image searching.

II. SYSTEM DESIGN

A. Rules of the system

Our system is played on a Twitch live streaming site through the chatting area (this kind of game is typically called an “audience participation game”). Unlike other GWAPs, audiences do not need to guess—but just describe and vote descriptions. In the *ESP* game, a player guesses what their partner would type as keywords for a given image. In *Karido* and *ClueMeIn*, a player selects an image in a list of similar images based on their descriptions made by her or his partner.

In our system, players are assigned to one of two groups (c.f. Fig. 2 for a scenario example). The first group, called “Describers,” are tasked with describing, say, three Ukiyo-e images shown by our system through live streaming within, say, one minute. Describers can describe as many shown images as they want, but each Descriptor cannot give more than one description per image. Then, the second group, called “Voters,” are tasked with voting the quality of descriptions provided by the first group within a limited time slot, say, 15 seconds. Each Voter can only vote for one description. A Voter will become a Descriptor when an image -- whose latest description he or she had previously voted won -- is shown again; such a description will automatically be associated to this Descriptor (former Voter).

B. Scoring

The description of each image that obtains the highest votes from Voters will win, and the Descriptor who has created this description will obtain points according to the number of votes. The latest set of Voters who have voted for the winning description also earn points and will be Describers when the image appears again next time where the system will set the description to be their description for the image unless they submit a new description. If they do not submit a new description and the automatically associated description does not receive any votes, their points will be decreased as a penalty of for previously voting a bad albeit winning (at that round) description; if this is the case, the points of the original Descriptor who actually created the description will also be decreased. These are mechanisms for preventing Describers and Voters from intentionally creating poor descriptions and not seriously making votes, respectively. In theory, the quality of the description of an image will be improved every time the image appears on the system.

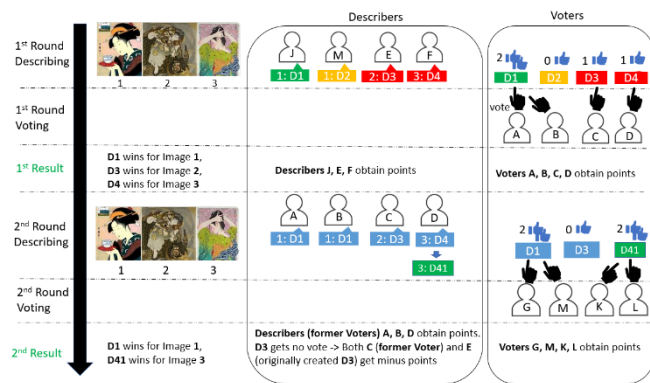


Fig. 2. Describing and Voting Mechanisms

C. Description Quality Assurance

To ensure a description is good for an image, we set a threshold after the image is shown the second time. When the description’s total votes, which are accumulated from the second display, reach the threshold, it will be the image description and the system will stop showing the image.

III. IMPLEMENTATION

The screen of the game system has three parts (Fig. 1): a game screen, an area for displaying Ukiyo-e images, and an area for showing a ranking of viewers based on their scores. The game is streamed on Twitch using StreamLabs OBS [6], one of the most popular streamer tools for live streaming sites, such as Twitch, YouTube and Mixer.

We created a chatbot using Twitch API for asking audiences to describe images they prefer in a format of “*image ID: description*,” for example: “*1: a Japanese man is fishing on a river next to a mountain*.” The description time is set to one minute; after this one-minute period, new coming descriptions are invalid and not processed by the system. Then, a chatbot will ask other players to vote generated descriptions by typing “*#dID*” within 15 seconds, where *dID* is a description identifier shown in the list of descriptions. We set the time for voting to only 15 seconds to prevent vote spamming (cheating) from Voters who have many accounts. We expect Voters to rush to vote descriptions they consider best representing a respective image, by which they cannot see the voting trend beforehand.

IV. CONCLUSIONS AND FUTURE WORK

We presented a novel mechanism of Games With a Purpose that creates informative descriptive sentences for Ukiyo-e images through chatting on a game live streaming platform. Informative descriptions of Ukiyo-e images achieved from this system can be further used for future work, e.g., a module recommending Ukiyo-e images as a game background to a group of audiences watching live streaming.

This work is currently in progress, yet promising. Our future plans include a full-scale system evaluation, both in terms of quality assurance of generated descriptions and effects on player experience.

REFERENCES

- [1] L. von Ahn and L. Dabbish, “Designing games with a purpose,” *Communications of the ACM*, vol. 51, no. 8, pp. 58–67, 2008.
- [2] L. von Ahn and L. Dabbish, “Labeling images with a computer game,” in *CHI ’04: Proc. of SIGCHI conf. on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2004, pp. 319–326.
- [3] B. Steinmayr, C. Wieser, F. Kneißl and F. Bry, “Karido: A GWAP for telling artworks apart,” in *16th International Conference on Computer Games (CGAMES)*, Louisville, KY, 2011, pp. 193–200.
- [4] Christopher Harris, “ClueMeIn: Enhancing the ESP Game to Obtain More Specific Image Labels,” in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY ’18 Extended Abstracts)*. ACM, New York, NY, USA, 2018, pp. 447–452.
- [5] L. von Ahn, S. Ginosar, M. Kedia, R. Liu, and M. Blum, “Improving accessibility of the web with a computer game,” in *CHI ’06: Proc. of SIGCHI conf. on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2006, pp. 79–82.
- [6] StreamLab OBS – Retrieved on <https://streamlabs.com/>