The PhotoSlap Game: Play to Annotate

Tsung-Hsiang Chang and **Chien-Ju Ho** and **Jane Yung-jen Hsu**
Department of Computer Science and Information Engineering
National Taiwan University
{vgod,b90090,yjhsu}@csie.ntu.edu.tw

Abstract

This paper presents PhotoSlap, an intelligent system for semantic annotation of photos. The system contains a semi-automatic face detector, a bulk annotation tool, and a multi-player online game, PhotoSlap. By exploring the design principles of gameplay and applying game theoretic analysis, PhotoSlap is designed as a fun and productive game, which adapts itself to different players to produce the desired output. Experiments involving four focus groups showed the game to be fun and effective in annotating people metadata for personal photo collections.

Introduction

Despite impressive advances in Internet search technologies, multimedia content still presents significant challenges for the state-of-the-art search engines. Semantic annotation of images can greatly improve the accuracy and efficiency of image search. While content-independent metadata can be gleaned from a picture by computers easily, semantic information such as *who* were there or *what activity* was going on, is not yet available. In general, automatic image annotation is a difficult task for computers. Despite successful applications of computer vision in specialized problem domains, no general solution can match the performance of humans in image recognition and understanding. The concept of *human computation* proposed in (von Ahn & Dabbish 2004) demonstrated a fun alternative.

In this paper, we present PhotoSlap, an intelligent system for semantic annotation from an online game. The core of PhotoSlap is a productivity game that generates face photo clusters as a result of player actions. Semantic annotations can then created with a bulk annotation tool.

To keep players engaged in the game, PhotoSlap adapts itself in order to cater to each individual player by observing his/her behavior. To ensure data quality, game theoretic analysis has been applied to show that PhotoSlap reaches *subgame perfect equilibrium* with the target strategy when players are rational. That is, players who would like to obtain the highest score in the game would take the actions prescribed by the optimal playing strategy.

PhotoSlap Annotation System

PhotoSlap Annotation System (PAS) is a hybrid system for efficient photo annotation, which integrates a semi-automatic face detection tool and a bulk annotation tool with the PhotoSlap game.

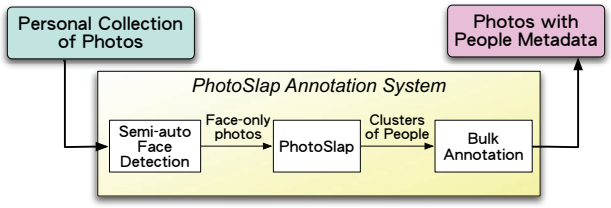

Figure 1: The block diagram of system.

While users are uploading their collection of photos, PAS performs face detection using OpenCV to extract face-only photos at the same time. The user could use a face-editing tool to fix the detection errors at any time.

Upon finishing playing the PhotoSlap game, the face photos are clustered into photos for individual persons. Users can use the bulk annotation tool to label the clusters quickly. Users are able to search the photos using the tags.

PhotoSlap

PhotoSlap is designed as a multi-player online game, whose rules are similar to the popular card game *Snap*. In this game, the cards of photos are dealt to each player in face-down stacks. Players take turns to take the top card from their stacks and place it face-up in a central pile. If two cards placed consecutively on the pile are *matching* in that they contain photos of the same person (alternatively, object, event, or location), then the first player slaps on the central pile wins the round. To ensure data quality, random slapping is discouraged by the mechanism of "objection" and setting "traps". The screenshot of PhotoSlap is shown in Figure 2.

Game Actions

Each player in PhotoSlap may perform four possible actions: 1) **Flip**. Each player flips a card in turn. The photos are chosen by the game server adaptively. 2) **Slap**. To

Figure 2: The trap page and the game screen.

get high scores, a rational player will try to be the first to slap when matching cards are presented. 3) **Object**. When a player slaps, the other players have a option to challenge the slapped result by flagging an "objection." If the objection is successful, the objector would gain points while the slapper would lose points. If the objection failed, i.e., falls into the trap, then the objector will be penalized by losing large points. 4) **Trap**. While the "Object" action is used to prevent random slapping, the "Trap" mechanism is designed to prevent random-objection. At the beginning of a new game, each player will be presented with a subset of photos, in which he/she should set traps by identifying photos containing faces of the same person.

To address a potential problem of random-trap, the "trap" and "slap" mechanisms serve as mutual validation in our game design. The trap photos identified in the trap stage will be randomly selected and presented in the game stage. The player who sets the trap is not allowed to slap on it. On the other hand, he/she can get points if another player slaps on it, but will lose points if no player slaps. The game mechanism is designed to encourage players to slap and trap as accurately as possible.

Design Consideration

Self-adaptive Gameplay

The core elements of gameplay in PhotoSlap are pattern-recognition challenge and reaction time challenge; players compete with others to slap on the matching pair as soon as possible. To keep players engaged, PhotoSlap is designed to be a self-adaptive game. That is, their competitors will be carefully chosen so that the difficulty of challenges, the competitors' reaction time, and their ability can be balanced to make them stay in the flow state(Csikszentmihalyi 1990).] In other words, the players with similar reaction time will be grouped together. Furthermore, to ensure the players experience challenges in a proper tempo, the appearance of a matching pair will be dynamically adjusted to follow a tension curve that has proper frequent peaks. While flipping a card, PhotoSlap decides whether to present a challenge by a probability function $p(t) = Ne^{gt}$ where N is a normalization constant, t is the interval from the last hit to the current flipping, and g is a constant for adjusting the growing speed of tension. The current implementation presents the same card twice if $p(t)$ is over 100% and no potentially matching cards can be showed.

Game Strategy Analysis

To ensure the data quality of PhotoSlap, game theoretic analysis has been applied to examine if the design of PhotoSlap would produced the desired output, i.e., correct annotations. The whole process of the game is modeled as an extensive game without perfect information. By assuming that all players are rational, striving to maximize their scores, PhotoSlap can be shown to reach subgame perfect equilibrium with the target strategy, which would lead to desired output.

Evaluation

We have conducted small-scale experiments using four focus groups, consisting of 4 users each. For each focus group, the users played PhotoSlap for a 30-minute session continuously. Each session produced about 11 games. The test dataset used in the experiments contains 572 faces manually labeled and annotated by the authors.

Given the test dataset with the ground truths, the game is evaluated in the following aspects. 1) **Is the game fun?** Based on the gameplay survey at the conclusion of each focus group session, PhotoSlap received an average score of 7.6 on a 10-point scale. All users claimed that they would like to play again. 2) **How good is the game strategy?** The analysis of how the user conforms to the target strategy is analysed in precision and recall. While considering "object" as the elimination of "slap", the precision of the user strategy could achieve 99.84 % and the recall of it is 96.04 %. 3) **Is the game productive?** The *productivity* of the game is measured by the number of links being built and the accuracy of them. The links between face photos indicate they are *matching* based on users' agreement. In the focus-group study (8 person-hours), 1480 links are formed and 98.31% of them are correct.

Conclusion

This research aims to explore the design principles underlying productivity games and to identify meaningful evaluation metrics. This paper presented our design of PhotoSlap Annotation System, which explores the power of human computation for semantic annotation of digital photos. The experiments showed the game to be fun, conforming to target strategy, and productive in annotating people metadata for personal photo collections. In the future, this system can be extended to annotate object, event, and location, in addition to the people metadata.

Acknowledgment

This research was supported by NSC (NSC-95-2622-E-002-018) and by the Program for Promoting Academic Excellence, NTU (95R0062-AE00-05).

References

Csikszentmihalyi, M. 1990. *Flow: The Psychology of Optimal Experience*. Harper and Row, 1st edition. chapter 4.

von Ahn, L., and Dabbish, L. 2004. Labeling images with a computer game. In *Proc. CHI '04*, 319–326. ACM Press.