

EyeGrab: A Gaze-based Game with a Purpose to Enrich Image Context Information

Tina Walber
University of Koblenz-Landau
Institute for Web Science and
Technologies
Koblenz, Germany
walber@uni-koblenz.de

Chantal Neuhaus
University of Koblenz-Landau
Institute for Web Science and
Technologies
Koblenz, Germany
cneuhaus@uni-koblenz.de

Ansgar Scherp
University of Koblenz-Landau
Institute for Web Science and
Technologies
Koblenz, Germany
scherp@unikoblenz.de

ABSTRACT

We present *EyeGrab*, a game for image classification that is controlled by the users' gaze. The players classify images according to their relevance for a given tag. Besides entertaining the players, the aim is to enrich the image context information to improve the image search in the future. During the game, information about the shown images is collected. It includes the classification concerning the tag, a rating of the given images by the user ("like" or "not like") and the eye tracking information recorded when viewing the images. In this work, we present the design of the game and compare two design variants – one with and one without visual aid – concerning the suitability of the game for image annotation. The variants of the game are evaluated in a study with 24 participants. We measured the user satisfaction, efficiency and effectiveness of the game. Overall, 83% of the users enjoyed playing the game. The results show that the visual aid is not helping the users in our application; it even increases the error rate. The best classification precision we achieve is 92% for the game variant without visual aid.

Categories and Subject Descriptors

H.5.2 [User Interfaces]: Input devices and strategies

General Terms

Human Factors

Keywords

Eye tracking, Game with purpose

1. INTRODUCTION

The search for digital images is still a challenging task. It is often performed based on context information, e.g. tags describing the image content. Tags assigned to specific image regions instead of the image can improve the search results [9]. Also the ratings of the images can be used to deliver good search results, as it is done on some image stocking pages like Photo.net.

The game *EyeGrab* has been developed as a game with a purpose (GWAP) to improve or collect these information: the description by tags, a personal rating and information about image regions.

Presented at EuroHCIR2012. Copyright © 2012 for the individual papers by the papers' authors. Copying permitted only for private and academic purposes. This volume is published and copyrighted by its editors.

The game is controlled by eye movements, which on the one hand enhances the user satisfaction and on the other hand allows for collecting gaze information that will be analyzed to gain information about the image content. The overall goal of *EyeGrab* is to enrich the images with contextual information in order to improve future search tasks.

The players look at images falling down the screen. The task is to classify the images as relevant or irrelevant to a given tag. Relevant images are rated by the participants into "I like it" or "I do not like it". We have compared two different interaction designs of the game in a study with 24 subjects and measured effectiveness, efficiency, and satisfaction. While the first variant provides visual aids in form of highlighting the interactive regions and a gaze-cursor, which visualizes the subjects' fixations on the screen, the second variant does not. Overall, we can state that the vast majority of the participants enjoyed playing the game and that the gaze-based control of the game was experienced as an improvement on the entertainment. The players which received the visual aids had the impression to be supported by it. However, the results show that the visual aids lead to significantly more incorrect and missing classifications. In fact, given a ground truth image data set, we have achieved the best classification precision with 92% for the game variant without visual aid.

The satisfaction and the precision of the results gained in our experiment are very satisfactory. Based on this outcome, we will continue with the evaluation and conduct information extraction from the gaze paths in a next step. Based on a prior experiment [8], we can use this gaze information to add region-based annotations to the images.

2. RELATED WORK

A large number of applications were introduced in the past that use eye movements as input medium, often for people with disabilities, e.g., for a drawing system [1]. Also the use of gaze information as relevance feedback in image retrieval was investigated with promising results, e.g. [6]. Walber et al. [8] showed that specific image regions can be identified, using gaze information. The development of sensor hardware like cameras in computers is continuously progressing. Already now eye tracking can be performed with a commodity web camera. San Agustin et al. [7] compare a commercial eye tracker and a webcam system. The results for the webcam system are satisfactory and comparable to the commercial system, although still with limitations concerning the comfort of use. Based on this development, one can assume that eye tracking could be performed for more users in the future and it will be possible to use the technology also in playing games.

Data obtained from eye tracking is less accurate than, e.g., from a computer mouse, due to natural movements of the eyes. It can be

difficult for the users to focus the gaze on a specific region to select a button. One possibility for supporting the users in controlling an application by gaze is to visualize the gaze as a cursor. Some related work indicates the problem of distraction from this kind of visualization [2], others see the chance of such a natural “pointer” [5].

Some years ago, a new class of applications appeared, the so-called games with a purpose (GWAPs) [3, 4]. The goal of GWAPs is to gain information from humans in an entertaining way. One example is the game Peekaboom [4] where two users play together for labeling image regions. Another is the ESP-Game [3] with two randomly assigned players, each tagging one image and trying to provide the same tags as the team mate. Tobii recently introduced the game EyeAsteroids and claims to be the first purely eye-controlled arcade game. It is entertaining, but does not have the goal to benefit from the users' activities. Eye tracking fascinates users as an unusual kind of input device. One can benefit from this curiosity by offering entertaining applications that also gain some information from the users. Despite the variety of eye tracking applications and games, *EyeGrab* combines – to the best of our knowledge – for the first time both, the aspects of leveraging from user activities like in GWAPs and controlling the application by the use of eye movements.

3. DESCRIPTION OF THE GAME

EyeGrab is a single-player game that takes place in a galactic dimension. The task is to clean up the aliens' universe by categorizing and rating images. Before the game starts, the user is asked to enter his or her nickname using the keyboard (Figure 1). The rest of the game is then played exclusively by the use of the eye tracker. Every gaze-based selection takes place after a dwell time of 450 milliseconds to avoid random choices. For example, the selection of the gender is done by focusing a male or female character as shown in Figure 1. Subsequently, the player is shown a small introduction to the game's rules (no screenshot).

The game has three rounds, with three categories (“car”, “house”, and “mouse”). First, the category is shown (see Figure 2), next the round starts. 30 images fall down the screen as depicted in Figure 3. The player categorizes the falling images into one of three categories. He or she can select an image by fixating it for more than 450 milliseconds. When an image is selected, it is highlighted by a thin, red frame. Next the image is classified into “like it”, “don't like it”, or “not relevant”, where the first two imply that the image is described by the named tag and the third specifies that the image does not belong to that tag. To classify an image, the user looks at the area of the intended classification on the screen as shown in Figure 3 (same dwell time as above). The player receives points for each correctly categorized image, negative points for each false one and no points for images that fell off the screen without classifying them. To further challenge the user, the speed is increased with higher levels. A high score list is presented to the user at the end of the game.



Figure 1. Start screen with gaze-selected “male”.



Figure 2. Presentation of the category.

For each category, the images were chosen from the 100 most relevant Flickr-pictures. 20 of them were randomly selected and combined with 10 pictures of a different category. An inter-rater agreement with 3 neutral persons was used to confirm the categorization and to create the ground truth.

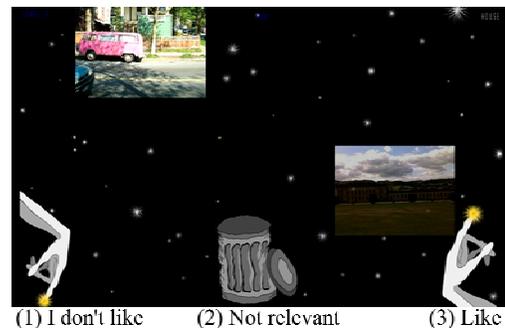


Figure 3. Gaze-based image classification.

Two versions of the game have been implemented, one offering visual aids (see Figure 4a) to the user and the other one without such help (see Figure 4b). The visual aids include a highlighting of the “action areas”, i.e., areas which perform an action when being fixated, and the visualization of the gaze point on the screen (gaze-cursor). Examples are the classification buttons as shown in Figure 3 (details in Figures 4a and 4b).

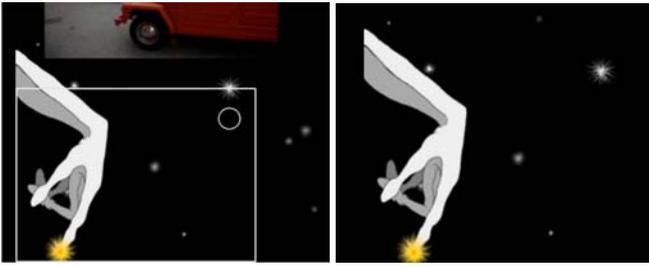


Figure 4a. Visual aids
(rectangle: action area, here: “not like”, circle: gaze cursor).

Figure 4b. No visual aids.

4. EVALUATION DESIGN

In order to evaluate *EyeGrab*, 24 subjects (7 female) played the game. The subjects’ age was between 15 and 32 years (mean: 24, SD: 3.9). 19 subjects were students, 2 research assistant, one pupil and 2 had other professions. Most of the players had experience in gaming (mean: 3.5, SD: 1.31). Only a few were familiar with eye tracking, this is indicated by 19 subjects rating the question concerning their eye tracking experience with one (mean: 1.63, SD: 1.38). The subjects were randomly divided into two groups A and B. Group A had no visual aids during the game, whereas group B did. 8 users were wearers of glasses or contact lenses (4 in each group). There were no problems using the eye tracker for those subjects.

To avoid distractions, the game was played in our eye-tracking lab providing a chair, a desk with an eye tracker, and a standard monitor. The first step was a calibration of the eye tracker. After this was done, the game was started without further instructions and was played with 30 images in each of the three rounds. The data from the first round is not used in the later analysis, because it has only served for getting the subjects acquainted with the usage of the eye tracker as input device. At the end of the experiment, every user filled out a questionnaire, including personal information and questions about the performance of the game. The answers were given on a 5-point Likert scale.

5. EVALUATION RESULTS

5.1 Satisfaction

The questionnaires show that the subjects enjoyed playing the game. On average, the statement “It was fun playing the game.” is rated 3.46 (SD: 0.93) considering all 24 subjects. 20 of the 24 users agreed to this statement. One of the following questions was if the participants felt like the interaction with the eye tracker increases the fun of the game. 14 subjects agreed or strongly agreed to this statement (mean: 3.5, SD: 1.25). Also, most of the subjects did not feel disturbed by the eye tracker (mean: 2.25, SD: 1.5).

5.2 Effectiveness and Efficiency

One round of the game comprises 30 images and it takes about two minutes including the introduction and the input form. Each level has a different pace at which the images fall down the screen. Thus, the classification per image takes between 2.6 and 4 seconds.

In total, 1440 pictures were shown to the subjects within the game (30 pictures per category “house” and “mouse” times 24 subjects). Only in 42 cases, the image passed without classification, resulting in a total of 1398 classified images. 1162 images were correctly classified (83%). Thus, only 236 images were incorrectly classified. Overall we had 897 true-positive classifications, 128 false-negative and 108 false-positive classification, which leads to a precision of 89% and a recall of 88% over all users. For the group with the better results (the group without visual aid, see next section) we obtain a precision of 92%.

5.3 Visual aid

The subjective perception of the users in group B (the group that was provided with visual aids) was that the visual aids supported them in the classification tasks. The question regarding the visual highlighting of the active areas was rated as very helpful with an average of 4.67 (SD: 0.49). The subjects also answered that displaying the gaze point was very helpful and scored this question on average with 4.5 (SD: 0.67). However, to our surprise, the following statistical analysis of the data shows that group B with the visual aids misclassified significantly more images than group A did.

Group A correctly classified 296 images for category “house” whereas group B correctly assigned 264 images for this category. Regarding the category “mouse”, in total 317 correct assignments were made by group A whereas group B correctly assigned 287 images. Regarding the misclassified images, group A misclassified 59 images for category “house”, whereas group B wrongly assigned the image category in 81 cases. For the category “mouse”, the number of incorrect assignments is 37 for group A and 57 for group B. We compared the values for correct and incorrect assignments for group A and B in a 2x2 Chi-square test for both categories. The differences are significant regarding a significance level of $\alpha = 0.05$ with $\chi^2(1, N = 700) = 5.14, p = 0.023$ for the category “house” and $\chi^2(1, N = 698) = 5.6, p = 0.018$ for “mouse”. In group B 31, images passed without classification, in group A only 11 images were not classified.

These results indicate that the visual support is not improving the classification. Despite the good impression of the visual support that group B expressed, the following question might be an indicator that this group felt less comfortable with the eye tracker-based interaction than group A did: we asked the subjects to state if they preferred a mouse-based interaction instead of the eye tracker-based one. On average, subjects of group A scored this question with 2.17 (SD: 1.47) and group B with 3.25 (SD: 1.48). Using a two-tailed Mann-Whitney U-Test, a weakly significant difference was determined stating a preference of group B over group A to use the mouse to play *EyeGrab* ($U = 43, z = -1.719, n_1 = n_2 = 12, p = .085$).

6. FUTURE WORK

For the current version of our *EyeGrab* game, we have used pre-classified images in order to verify the classification performance of the subjects. We plan to use images without annotations in future extensions of the game.

Also the detailed analysis of the gaze information will be performed in a next step. In a small sample of 5 images classified by one user, we received 231 gaze points on the images. An example of a gaze path visualization is shown in Figure 5. We

expect a sufficient number of fixations and correct classification to allow a detailed analysis.

We received 897 ratings for the shown images. 556 of them were positive. The quality of these ratings has to be investigated in a future experiment, e.g., by repeating the ratings in another context with the same users or by using a ground truth set with images, often liked by a big number of other users. However it has to be clear, that a subjective rating can never be “correct” or not. These investigations can only provide an indication of the worth of the rating. Overall, this detailed analysis will allow us to identify the regions that correspond to the category given in the EyeGrab game. Such region-based annotations will allow for a better retrieval of the images in the future.



Figure 5. Visualization of fixations on a classified image.

7. SUMMARY

We have introduced the gaze-based game with a purpose *EyeGrab* to classify images using an eye tracker. We have shown that the game has the potential to entertain the players and that the classification results are good enough to advance beyond the gaze analysis. This analysis is the first step in the direction of extending image context information with information gained in an eye tracking game. The next step will be the analysis and

evaluation of the gained information and to use it for improving image search tasks.

8. REFERENCES

- [1] Hornof, A.J. and Cavender, A. 2005. EyeDraw: enabling children with severe motor impairments to draw with their eyes. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 170.
- [2] Jacob, R.J.K. 1993. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction*, 151–190.
- [3] Von Ahn, L. and Dabbish, L. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 319–326.
- [4] Von Ahn, L., Liu, R. and Blum, M. 2006. Peekaboom : A Game for Locating Objects in Images. *SIGCHI conference on Human Factors in computing systems*, 55–64.
- [5] Zhang, X. and Ren, X., and Zha, H. 2008. Improving eye cursor’s stability for eye pointing tasks. *SIGCHI conference on Human factors in computing systems*, 525–534.
- [6] Kozma, L., Klami, A. and Kaski, S. 2009. GaZIR: gaze-based zooming interface for image retrieval. In *Proceedings of the 2009 international conference on Multimodal interfaces*.
- [7] J. San Agustin, H. Skovsgaard, J.P. Hansen, and D.W. Hansen. Low-cost gaze interaction: ready to deliver the promises. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 4453–4458. ACM, 2009.
- [8] Walber, T. and Scherp, A. and Staab, S. 2012. Identifying Objects in Images from Analyzing the Users’ Gaze Movements for Provided Tags. *Advances in Multimedia Modeling*, 138-148.
- [9] Carson, C., Thomas, M., Belongie, S., Hellerstein, J. and Malik, J. 1999. *Blobworld: A system for region-based image indexing and retrieval*. Visual Information and Information Systems.