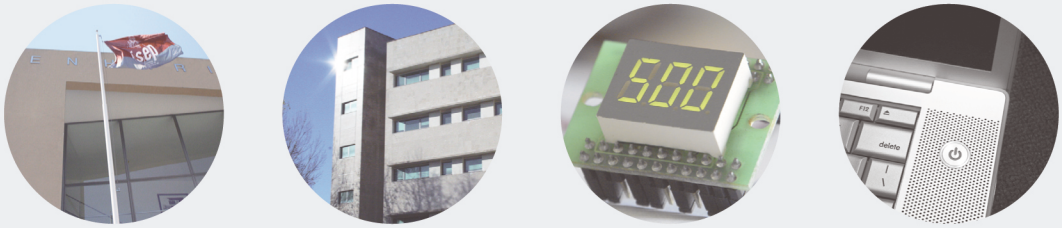
INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES



GAMIFYING IMAGE ANNOTATION: FOCUS OF ATTENTION IDENTIFICATION

VIGNESH UTHRAPATHY
Outubro de 2019

POLITÉCNICO
DO PORTO

GAMIFYING IMAGE ANNOTATION: FOCUS OF ATTENTION IDENTIFICATION

Vignesh Uthrapathy

Department of Electrical Engineering

Master in Electrical and Computer Engineering

Area of specialization in Automation and Systems

Report prepared to partially satisfy the requirements of the Thesis / Dissertation Course of the Master in Electrical and Computer Engineering

Candidate: Vignesh Uthrapathy, Nº 1161854, 1161854@isep.ipp.pt

Scientific Orientation: Professor Paula Viana, pmv@isep.ipp.pt

Organisation: INESC TEC

Supervisor: Paula Viana, paula.viana@inesctec.pt

Inês Teixeira, ines.f.teixeira@inesctec.pt

Department of Electrical Engineering

Master in Electrical and Computer Engineering

Area of specialisation in Automation and Systems

2019

ACKNOWLEDGEMENTS

I would like to thank the entire team at INESC TEC, Porto for helping me in the completion of the presented work. I would also like to thank INESC TEC for granting me the opportunity to conduct my internship at the Organisation.

I would also like to extend my gratitude to my mentor Prof. Paula Viana for guiding me and providing suggestions in improving and making this thesis.

I wish to present a special thanks to my girlfriend Spela Jeretina for all her love and moral support.

Finally, I would like to thank my family and friends for being a constant support and providing me encouragement without whom this would not have been possible

Abstract

With the growth of digital multimedia (especially images) in recent years, effective acquisition of proper metadata for images is still an open issue today. Crowd-based human computation represents a family of approaches able to provide large scale of metadata with decent quality.

This thesis aims to provide a solution for a collaborative approach of image annotation. The solution proposed by this thesis uses a collaborative approach that includes the concepts of Games with a Purpose (GWAP) to enable identifying the more relevant areas of a photo. The idea behind using GWAP here, is to gather information based on user's visual attention. The GWAP in this work depends on the ability of the user to quickly pay attention to parts of image instead of processing the whole scene of the image. This method helps the user to create tags on the image based on the important focus areas absorbed during the game. Thus, as a first objective, a game with a purpose was developed which motivates the user to identify regions of interest in the later part. The system developed includes the most important gamification elements expected to increase the user participation.

Keywords

Collaborative annotation, Annotation tool, Games with a Purpose, Visual attention.

Resumo

Ao grande quantidade de dados multimedia (especialmente imagens) produzida nos últimos anos, coloca o problema da associação de metadados que permitam garantir a sua re-utilização para diferentes fins. Sendo o processo de anotação de conteúdos moroso, diversas abordagens têm sido experimentadas, incluindo a utilização de técnicas colaborativas.

Esta tese visa fornecer uma solução para a abordagem colaborativa da anotação de imagem. A solução proposta por esta tese utiliza uma abordagem colaborativa que inclui os conceitos de Jogos com Propósito (GWAP - Games with a Purpose) com o objetivo de identificar as zonas mais relevantes de uma fotografia. Assim, como primeiro objetivo, foi desenvolvido um jogo que pretende motivar o jogador a identificar as regiões de interesse numa fase posterior.. O sistema desenvolvido possui importantes elementos de gamificação que visam aumentar a participação do utilizador.

Palavras-Chave

Abordagem colaborativa, Ferramenta de anotação, Jogos com propósito, Foco de atenção

Index

Index of Figures

Index of Tables

Acronyms

API	Application Programming Interface
AJAX	Asynchronous JavaScript and XML
CSS	Cascading Style Sheets
CTM	Telecommunication And Multimedia Centre
DB	Database
DOM	Document Object Model
ESP	Extra Sensory Perception
GWAP	Games with a Purpose
HTML	HyperText Markup Language
INESC	Instituto de Engenharia de Sistemas e Computadores do Porto
INESC TEC	INESC Tecnologia e Ciência
ISEP	Instituto Superior de Engenharia do Porto
KMS	Knowledge Management System
PHP	HyperText Prepocessor
R&D	Research And Development
SQL	Structured Query Language
URL	Uniform Resource Locator
XML	Extensible Markup Language

1. INTRODUCTION

This chapter introduces the work that aims to provide a solution for the collaborative approach of image annotation. An annotation tool (Foto in Motion) was developed at INESC TEC to help on the annotation of images by using artificial intelligence to automatically identify objects. These automatic annotations can then be manually refined though a platform that exposes the annotations and enables deleting, adjusting or adding additional information in order to enhance the description of the image.

The work developed in this thesis is intended to capture other type of information in the image and uses the concept of Games with a Purpose (GWAP) to try to gather information based on the user's visual attention, therefore providing valid annotations to the system and maintaining its main purpose, and at the same time, guarantee some interest and motivation from the users in performing this task

The idea that supports the approach is that when involved in a game in which the user has to concentrate in a set of images and to make fast decisions, he will pay attention to parts of image instead of processing the whole scene of the image. This method shall help the user in validating the tags of the image with the help of human visual memory or to identify areas relevant for the scene being described in the image. Thus, as a first objective, a game with a purpose, was developed which motivates the user to validate the existing image annotations or add a new annotation in the later part of the game. The system developed includes some of the important gamification elements that will be explained in chapter 2, and so it is expected to increase the user's participation.

1.1 CONTEXT

The work presented in this document was developed as part of the curricular unit Dissertation/Project/Internship in the Master in Electrical and Computer Engineering-Automation and Systems from Instituto Superior de Engenharia do Porto (ISEP) of the (IPP) Instituto Politécnico do Porto (IPP). This work was also developed in association with INESC TEC.

INESC TEC is a research institution dedicated to scientific research and technological development, seeking value creation and immediate social relevance. The 13 R&D Centres of INESC TEC are structured on four thematic domains: Computer Science, Industry and Innovation, Power and Energy and Network Intelligent Systems. The latter aims at creating autonomous networked intelligent hybrid systems enabled by ubiquitous sensing and information processing. This Cluster contribution is focused on developing new, smaller, smarter and adaptable sensing systems. One of the centres forming this cluster focuses on Telecommunications and Multimedia (CTM).

This work aims to contribute in the developments of the FotoInMotion and CHIC projects, being developed by the Multimedia Communications Technology group, a scientific area of the CTM centre, and ISEP. These projects aim to explore and research new approaches on Image Analysis and Multimedia Applications.

FotoInMotion is an European project with the collaboration of international institutions of technology as well as users in the areas of photojournalism, fashion and film festivals, as event organisation. FotoInMotion intends to develop mechanisms that enable transforming a single photograph into a dynamic, high quality video for storytelling and branding. The main goal of the project is to create a tool to allow both professional content producers and creative citizens to efficiently embed contextual information into a photograph, and produce videos with rich, semi-automated editing functions and dynamic effects that can easily be shared on social media, as well as on professional digital content delivery platforms. Thus, to create these animations and generate videos automatically, there is the need to decide which points on the image the software should be focusing on. To create these rules, in the past year, some work has been developed in two distinct fields: object recognition techniques, to identify specific objects and relative positions in the image, and visual attention, for detecting regions of interest that may not be associated with concrete contextual information. Nevertheless, the system produces primary results and some of the automatically generated annotations are flawed. To improve this, the annotations have to be corrected, so that the system may use the validated annotations to retrain and improve the guesses on objects and regions of interest.

CHIC exploit different applications areas that range from TV broadcasting, to Cinema or Tourism. In all these areas, multimedia content annotation is required to enable improving the efficiency of content-repurposing and access. Being able to produce metadata without having to engaged expensive professional staff on this task, is essential to guarantee that content is not deposited in the archives without these descriptions.

1.2 OBJECTIVES AND METHODOLOGY

The main objective of this work is to provide a solution for an image annotation tool with game strategies, exploring gamification techniques for player motivation. The game strategy used in this work requires human visual memory for point of interest (ROI) identification.

Humans use attention mechanisms based on goal-oriented and stimulus-driven information to define the region in the visual input where the attentional focus should be oriented. In this way, the amount of processing is limited to a certain region of the visual field and the regions to explore are prioritized in time.

To implement this work, a literature review regarding the existing methods in collaborative approaches for content annotation was done. On completion of this, attention was shifted towards the creation of basic strategy for this work. A basic structure was created after a thorough understanding of the existing methods in crowdsourcing and gamification.

The approach used has two main goals: first to implement the idea for a game and then a platform to annotate the images. From the literature review, a set of criteria were chosen and defined for the gamification techniques to support this work. The proposed solution is supported by the hypothesis that by playing the first part of the game (a classical memory game based on the images to annotate), the user will be conducted to annotate the regions that have captured his attention and enabled him identifying the card pairs.

1.3 STRUCTURE OF THE DOCUMENT

This document has the following structure:

- Introduction: The context, objective and the methodology followed in the given work is briefly described to help the reader understand the implemented work.
- Literature review: A brief synopsis and major takeaway points from various research papers are presented which acts as a guide to understand the existing methods used for collaboratives approaches in annotations.
- Proposed work: The Design, objective and the idea behind this work is described in detail to help the reader understand how this system works.
- Conclusion: This chapter presents the conclusions for the work done and add proposals for future work.

2. LITERATURE REVIEW

An annotation is a metadata, i.e., data attached to a data, which in this case, are images. Creating these annotations is a tedious and time-consuming process. It needs highly experienced and professional workspace to produce large volumes of annotated data. Searching and browsing large collection of image depends on the description of the content. However, having professionals manually describing content can be expensive. Engaging the end users in the process of describing the contents of the image assets may provide better results. However, a great part of user-generated annotations has no quality control. To ensure the effectiveness of the data collected, crowdsourcing has proved to be an effective method. Several approaches have been proposed in the literature to enhance the accuracy of the metadata. This thesis will exploit the concept of GWAP as a methodology to engage the user in the annotation task and to contribute to enhance the quality of the annotations.

This chapter presents the literature review required to develop the intended project work. In the following sections, the implementation of the collaborative based mechanisms to collect the metadata in media and techniques to extract the relevant information from the media will be explained. Motivation mechanisms usually used in games will also be identified.

2.1 COLLABORATIVE APPROACHES FOR CONTENT ANNOTATIONS

Annotation of audio-visual files is a difficult work. Automated approaches on the audio-visual files have limits in terms of types and level of detail of the metadata they provide. Despite continuous effort to improve automated annotations, some approaches focus on increasing human motivation to manually contribute with annotations. In order to enhance human interaction with the data, game-based approaches coined as Games with a Purpose emerged. GWAPs transform human intelligence tasks into appealing games by aligning the winning conditions of a game with task solving. GWAPs serve as an alternative approach

for achieving large-scale data processed by humans in cases where the tasks are hard to perform by the machines.

2.1.1 GAMES WITH A PURPOSE (GWAP) FOR CONTENT ANNOTATION

GWAPs are the class of games that uses gamification and human computation to collect data from the interaction with human users. The idea of using games with multimedia annotation was introduced by Luis von Ahn with the ESP game [1]. This multiplayer game randomly linked two players and made them to match their guesses on the same image within a limited period time. The objective is for the two players to enter the same word or phrase, which earns them points and becomes a label to describe the image. Figure 1 shows the screenshot of the ESP game interface where the players try to match the words. The ESP Game has served as a prototype for several later successors. The idea to implement games to collect metadata have been used in audio, video and images archives.

Figure 1 ESP Game Demo

2.1.2 GWAPS FOR AUDIO

There are two proposed methods for audio annotations. The first one is a real time two player game called Herd It where tasks are given to classify the music. The classification to be performed depends on spirit of music, genre of the song, name of the artists and name of the song. Figure 2 shows the interface of the game where the user answers some question based the song, he/she is listening. The points are rewarded based on the answers, the closer guess to the answer, the higher the points the players will receive [1][2].

Figure 2: Herd it Demo

The second method is TagATune which is also a two-player game where the players tag music clips. Players listen to the music clip and try to describe it in the best way they can, while the other player will decide whether they are listening to the same music clip or not. Scoring is based on the right guess [3][4]. Figure 3 shows the interface of the game where the users provide some description on the tune they listen to.

Figure 3 :TagATune Game Demo

2.1.3 GWAPS FOR VIDEO

There are few approaches for video annotations, implemented as a game. VideoTag is a one player game, where the player is presented with some Youtube videos which must be annotated within a specific time limit. Tags can be freely introduced by the player and scoring are based on the tag. The player can also ask for suggestions to tag the clip but when the player uses this option, he will be penalized for that. The scoring is based on three type of tag: Golden tag, Top Tag and Simply tag. The Golden Tag has the highest point and the Simple Tag has the lowest point [4].

Waisda? (which translates to "What's that?") is a video labeling game where the players are invited to tag what they hear and see. The points are awarded for a tag if it matches with the opponent player. Like in other games, limit time is imposed for each round [5][6].

Guess What? [7] is a Facebook app game where the players are invited to watch random video clip and then to answer the questions about it. The main objective is to find what the other account players guessed and the points are awarded based on the guesses.

The "Yahoo! Video Tag Game" is a multi-player game based on temporary tagging agreement between two players. Scoring ranges from 2 to 20 based on the temporal distance

between tags [8]. This mechanism doesn't aggregate the tags but just looks for the closest tag introduced by any player. The user must wait for another player to join. Figure 4 shows the screenshot from the game.

Figure 4: VideoTag Demo

Along with these games, other well-known application like Youtube [16] and Viddler[17] also rely on user generated tags to annotate the video content. Unfortunately, these applications have restrictions on the users who can annotate content.

TAG4VD (Tag for Video) is a single player game, where the user is asked to label some videos within a short time interval [9][10]. TAG4VD has two main mechanisms: Votes and Tags. Tags are introduced by the players and validated based on the collaborative mechanism that eliminates irregular annotations. Votes are used as an additional mechanism to guarantee the quality of the annotations by allowing the users to classify the good or bad quality of the provided tags.

There are two type of players in this game: Registered user and Guest. The Registered players can watch the clip, tag and vote. But the guest players are limited to just watch the clip and vote. The scoring mechanism is based on three concepts: Tag, Cluster of Tag and Number of times the tag appears in the Cluster of Tag. Points are provided for each action and additional information like metadata. The players can give opinion on the tags involved while watching the video clip. Figure 5 is the screenshot from the game where a user tries to provide tags on what they see in the video clip.

Figure 5: TAG4VD Demo

They can like or dislike the tags present in the tag cluster. Additionally, besides providing metadata, the registered players can also provide information on a specific timecode that they consider relevant. These timecodes associated with a "like" can be highlighted as the most impressive moments of the video content by pressing the high bars. Figure 6 shows the like and dislike feature from the game.

Figure 6 :TAG4VD Tagcloud

2.1.4 GWAPS FOR IMAGE

Since the ESP game was developed, other games have been developed with similar objective. Brooklyn Museum's online game **Tag! You're it!** is a crowdsourcing game based on ESP platform [11]. The game was a project for Brooklyn Museum. The project was developed to maintain the online collection of images of the museum's artifacts. But this game has been removed from the museum's website.

Peekaboom[12] is developed by the same creators as the ESP Game, is a web-based game that helps computers to locate the objects in the image. The game has two main components: "Peek" and "Boom". Two random players from the web participate by taking different roles in the game; when one is Peek, the other one is Boom. Peek starts with a blank screen and Boom starts with an image. The goal of the game is for the Boom to reveal the parts of the image to the Peek, so the Peek can make guess of the associated word. When Boom is revealing the part of image the Peek can make his guesses of Boom's word and Boom can see Peek's guesses. When Peek guesses the correct word, the roles gets switched and points

are allocated. If both players decide to change the image-word pair, they can select “pass” or opt out. Figure 7 shows the screenshot from the game. The game also has other option [12] like Hints and Bonus Rounds. When compared with the ESP Game, Peekaboom had improved the data collection from the ESP Game and for each object in the image, it outputs precise location information and other useful information for training computer vision algorithms.

Figure 7: PeekaBoom interface

Tag4Fun is a single player game developed using the 3D graphic library, OpenGL. The main objective of the game is to annotate images according to their content [13]. The game uses a technique called “Simple Image Block Technique”- which means different annotation layers are possible, distinguishing, for an example, the background from the object. When the game starts, the player is showed an image and is allowed to comment on contents of the image. Tag4Fun has a similar game structure like “Tetris”. The character drops from top to bottom and the player selects them using the keyboard. Random magic characters also appear that can be used in place of any other character. The player needs to form a keyword related to the content of the image. Points are provided based on the keyword formed which is matched with pre-defined classifiers. The game has three databases: fully annotated, non-annotated and partially annotated. When an image passes through Tag4Fun it contains a number of possible labels for it. If an image describes using the same label five times that keyword will be associated as a taboo word for the image and won’t allow players to use the taboo words for further labelling. If an image got eight taboo words the image will fully annotate and be discarded from the database. All other information captured will be saved for future references. The player is fed with random image from any of three databases. If the player

forms a non-related word for the fully annotated image, the game identifies the word as irrelevant, and blocks it from labelling. Figure 8 shows the screenshot from the game. The performance of Tag4Fun wasn't measured because it has to be played by a large number of users and since the project was not commercialized, it was difficult to measure the performance of the game.

Figure 8: Tag4Fun Game Framework

ClueMeIn,[15] is a two-player game. Unlike the ESP Game, which examines a single image, ClueMeIn, presents the pair of players with three to five similar images. These images can be selected from ESP Game or other sources. The game is designed to come up with labels that distinguish similar images from one another by providing additional informative labels without penalties associated. There are two roles in the game: Guesser and Cluegiver. The players are presented with a set of images in a random order. The game indicates one image to the cluegiver to describe it to the guesser. The players switch the roles after five rounds on different set of images. Scores are provided based on the how the guesser and cluegiver play the game.

- Cluegivers are given points based on whether it gives the guesser the correct choice and on how unique their clues(words or phrases) are – they examine the label frequency and once a clue gets mentioned three times(across multiple games), it is added to the "taboo" words. By dividing the number of terms supplied for the image overall by the number of instances the word has appeared previously, the score gets computed.
- Guessers are given points based on how few guesses they use to identify the correct image. They can only make a single guess after a clue has been provided by the

cluegiver. The number of guesses is counted. Therefore, guessers are given more points for guessing the first image correctly out of five images than guessing the second time correctly out of the remaining four. Figure 9 shows the screenshot from the game indicating the view of guesser(top) and cluegiver(bottom).

- To prevent the cluegiver from supplying intentionally useless labels or the guesser from making intentionally poor guesses, there is a common set of points that are assigned to both players based on their mutual performance.

The image labels found in the earlier session of the game became a "Taboo word" (i.e., no score is allocated for the known words by the system) in the later session for other players.

Figure 9: Cluemeln- Screenshots from the game indicating the view of guesser (top) and cluegiver (bottom). Players take turns in each role and have different incentives to facilitate meaningful clues

Apart from the Typical Two player games, KissKissBan [14] uses an innovative competitive collaborative model with three players to annotate images. The game has two different roles: Blocker and couples. One of the players is the "blocker" and the other two players are the "couples". All the three players are presented with the same image, the couples try to match the keywords with each other(kiss) and the blocker tries to block the couples from making the match (ban). When the game starts,

- The blocker has seven seconds to make a list of blocked words. He/she is able to see every word the couples are typing during the game. Monitoring the game provides the blocker an opportunity to stop the couples from achieving some unified strategy. Figure 10 shows the interface for the blocker.

- After that, the couples have thirty seconds to match the guesses. If the couples make a word from the blocked list, the game time decreases by five seconds, also it is not considered a match. If the couple manages to find the match apart from the list of blocked words, they win the round and a new tag is added to the system.

The roles of the players change every five rounds in a fifteen rounds game. However, KissKissBan suffers from some of the same design issues of the ESP game.

Figure 10: KissKissBan – interface for blocker.

The main drawbacks of these proposals are either related to the mechanisms implemented to guarantee the quality of tags which are too generalist, not correctly describing the content or to the computation cost related to the multimedia data. In a collaborative environment, motivation mechanisms are key factor to increase the number of contributions. Some of these approaches have motivation techniques, but still there are more important techniques which are described in the following topic.

2.2 MOTIVATION MECHANISMS

The motivation of playful experience is often used to engage users into solving the human intelligence tasks in the field called gamification [20] which is related to GWAPs. The difference between GWAPs and gamification-based approaches is that: In gamification-based approach, game elements like leaderboards and achievements are involved along with the existing process. The GWAPs on the other hand, create a new working process as a game:

they transform the problem solving into game rules that force players to disclose their knowledge or solve an instance of a problem. The GWAPs have in fact turned into the brilliant side of crowdsourcing as they offer a more controlled condition for information acquisition than other regular and uncontrolled crowdsourcing.

Motivating a player to engage in something requires a stable mechanism. The gamification technique is getting popular to be implemented in learning and training applications. Gamification [18] refers to the use of game elements in a non-game context to increase the human – computer interaction [19]. Some of the Game elements that can be used in gamification as motivation mechanisms are:

- Game design patterns and Mechanics
- Points
- Badges
- Challenges and Contests
- Awards and Prizes
- Rankings and Leaderboards

Some elements construct the game mechanics (points and badges that provide feedback on participant's progress), while others build up the replication mechanics (leader boards that make participant's performance level public) and social mechanics (peer ratings and comments that help to build relationships within the gamified application)

2.2.1 POINTS

The fundamental part of any gamified system is the ability for the player to earn points. Players are rewarded with points for performing various tasks, such as providing a comment to a picture, tagging, with upvote or downvotes, correcting a tag or posting a picture. Moreover, users are motivated by social pressure and competition to accumulate more points than other users and may also be motivated by the mere satisfaction of having more points than before. When designing a gamified system, it is important to set up the point-awarding mechanism transparently, so the users know what activities are rewarded with what amounts of points.

2.2.2 BADGES

Badges are ornaments or symbols players can earn for pre-defined favorable behaviors. These behaviors may include reaching thresholds in point accrual, completing certain tasks, winning challenges, performing certain task for several consecutive days and so on. When the user earns a badge, he can display it in his company profile or attach it to his avatar for everyone to see. Public display of badges allows for comparison with others, delivering satisfaction to those with many hard-to-earn badges, and motivating those who are below the average. Just like in children's scouting, badges are awarded for reaching a pre-set goal. However, in a gamified system, badges are virtual only and are awarded for reaching different goals. Badges represent goals and performance in three ways:

- With goals, people anchor their expectations higher, this in turn increases their performance;
- Self-efficacy is enhanced with assigned goals; and
- Upon goal completion users' satisfaction increases, leading to increased performance in the same activity in the future.

Badges themselves are a valid motivational tool and mere proximity of reaching a badge threshold may result in a peak in performance just to receive the badge earlier.

2.2.3 CHALLENGES AND CONTESTS

Challenges and Contests can also be included in gamified Knowledge Management System (KMS). These tasks are typically limited in time: users have a pre-set timeframe to fulfill them – two hours, a day or a week. Challenges are tasks that need to be fulfilled in the time given in order to receive a reward, a badge or a pre-defined amount of points. In contests, users compete for the best performance within a given timeframe. Turning a task into a time-constrained challenge renders it to be of higher priority. Users who decide to pursue the challenge allocate more resources to task-relevant activities and avoid distractions more. Setting up challenges is not optimal for all kinds of tasks though. Time pressure increases performance, but at the same diminishes precision and output quality. Users under time pressure rely more heavily on various heuristics and their focus scope is narrowed. Thus

time-limited challenges and contests are not suitable for creative tasks or those that require great precision and analytical approach.

2.2.4 RANKING AND LEADERBOARD

Rankings and Leaderboards represent the cumulative results of points collected, from completing challenges and badges earned. These are well known in the world of video games and provide additional motivation for extended effort and engagement. Being the best, or among the best in a certain area and having others see the user's success in the leaderboard motivates them to earn more points and badges.

2.3 ARTIFACT VALIDATION

All GWAPs must address one design paradox: they have to give players immediate feedback during or shortly after the game to maintain player attention. However, if the feedback (i.e., score) is not based on the purpose of the game, players might try to win the game in some other way next time when they play the game. But how can the game validate the results the player has produced, if these are the defined products of a human intelligence task, which cannot be performed by a machine? Some GWAPs use multiple players to validate each other's outputs at the same time [1], which introduces a serious cold start problem since few players play the game at the beginning. This means, it is impossible to match them effectively, especially if the condition of the game is, that the players do not know each other. Others acquire the knowledge from players by validating their behavior with already annotated resources or known facts.

For GWAPs, the resolution of cold start issues is relevant for several reasons:

- Many GWAPs do not survive the initial deployment phase when few players play them – their design is prone to cold start which prevents that. Only once a GWAP has enough players, does game design preventing the cold start problem become irrelevant. While the attraction of more players is a matter of other design aspects, if GWAP design

mitigates this problem, it is able to function with fewer players and has better chances to establish itself in the long term.

- If the game is used in a specific domain circle, the pool of players shrinks, rendering the need for solving the cold start problem more severe.
- Cold start design may also be impractical for researchers and practitioners who want to experiment with their design on smaller scale.

3. ARCHITECTURE AND IMPLEMENTATION

The work presented in this document proposes a solution for an image annotation tool using a collaborative approach that includes the concepts of crowdsourcing and gamification to collect metadata. Information is linked with different contents of the image contributing to enhanced search and access to image content. Previously proposed solutions [12][13] based on human computation games tried to bridge the semantic gap between descriptions by using dictionaries or similarity lists. But in some cases, the inputs requested by the player are still too complex [14][15]. Our approach goes beyond these solutions and uses common crowdsourcing methods to validate the metadata along with the crowdsourced tag-abased dictionaries. The idea behind the GWAP used in this work, is based on user's visual attention. Moreover, the advancement in web-based and HTML technologies has opened paths for the community of contributors to participate easily from different devices, providing better performance.

3.1 DESIGN OVERVIEW

Foto In Motion Game, whose purpose is to validate the annotations for input images, is a computer adaptation of the popular Concentration game. Concentration is a turn-based board game for two or more players, who compete to collect most of the image card pairs, which are mixed and placed on a table facing down (a standard game comprises a board of 8 X 8 cards).This game relies on the ability of the user to quickly pay attentions to parts of image instead of processing the whole scene of the image. This process helps the user to validate the images, later in the session, with more attention given to the important parts of the image. In each turn of the game, the player may flip two arbitrary cards to see whether they form a pair. If yes, the player keeps the pair (and receives points for it), otherwise the cards are flipped back, and the next player continues. The key to success is to

memorize the positions of the cards that have been flipped during unsuccessful attempts (one's own or those of other players) and retrieve them during one's turn.

We devised Foto In Motion Game as a computer modification of the Concentration game with a key difference: it is a single player (to address the cold start problem often encountered in other GWAPs). The scoring function has been therefore redesigned and is based on:

- The number of cards flips needed to find all pairs,
- The total number of card pairs present in the game,
- The elapsed game time.

Foto In Motion Game presents a challenge of card positions memorizing since a simple "blind" scanning of many images consumes some time and that time negatively affects the score. In addition to this, the player has the chance to annotate the focus regions of the images based on the visual memory the player gathered from the game.

3.2 HOW THE GAME WORKS?

The Foto In Motion Game is implemented as a web-based application, as it is suitable for mass public deployment on the web which is the usual choice for most GWAPs. A typical session can be summarized as follows:

1. The game starts by initializing the game board with hidden cards facing down and starting the game timer.
2. The face down cards have the images obtained by the annotation tool which are already annotated with an object recognition algorithm.
3. When the first card is flipped, the countdown at the timer begins.
4. The player continuously makes turns, flipping two cards at each round
5. When two cards are flipped, the game registers a move and increments the move counter value and checks for a match.
6. If the cards are not matched, the cards are flipped (hides)back and the player continues to check for a match. After every two cards are flipped, move is incremented by one.

7. If an identical pair is flipped, points are awarded, and the pair remains visible on the board permanently.
8. The game ends once all pairs are discovered. The player can view the final score of the game and the rank in the leaderboard.

The scores are provided based on the time, number of card-flips and total number of cards i.e., totalscore = (score for matches+ score for the time + score for the moves). The game has three modes of difficulty: easy, medium and hard. Each of these modes, have different points system. When the difficulty mode is higher, more points are rewarded i.e., for instance, the player plays one hard mode game and receives 30 points, the same points can be achieved by the player by playing three easy mode games.

For each pair found, the user gets 5 points in easy mode,10 points in medium mode and 20 points in hard mode. For each second elapsed in the countdown timer, the user loses 2 points in easy mode, 3 points in medium mode and loses 5 points in hard mode i.e., the user should find all the pairs in a short time to get more points. And for the number of moves, a maximum limit for the moves is set (30 moves), so for every move, the count is subtracted from the maximum limit. So, after 30 moves, points will not be provided to the user for the moves. The points are distributed as shown below

For example: If the user takes 20 moves to find all pairs with 30 seconds left in the timer, then

(Mode: Easy)

- Movescore = (30-20)*2 = 20;
- Timescore = 30 * 2 = 60; (total time = 90)
- Pairsmade = 6* 5= 30; (a perfect game has 8 pairs)
- The final score will be 20+60+30= 110.

(Mode: Medium)

- Movescore = (30-20)*3 = 30;
- Timescore = 30 * 3 = 60; (total time = 90)
- Pairsmade = 8* 10= 80; (a perfect game has 8 pairs)
- The final score will be 30+90+80= 200.

(Mode: Hard)

- Movescore = (30-20)*5 = 50;
- Timescore = 30 * 5 = 150; (total time = 90)
- Pairsmade = 8* 20= 160; (a perfect game has 8 pairs)
- The final score will be 50+160+150= 360.

After the all the pairs are found, the score for the game is displayed. If the user is satisfied with the score, he/she can submit it and proceed to the next part of the game or the user can choose to play again to achieve a better score.

After the game is completed, the user has the option to validate the annotation of the images which were presented to him in the memory game.

In the second part of the game, the user is presented with gallery of the match images. When the image is clicked, the user is taken to a page which contains the image and the region to insert tags. The user can make changes to the tag list by adding or deleting the tag elements which he/she considers as important while playing the memory game. Once the user is satisfied with the tags of the image, he can submit the changes and move on to the next image from the gallery. Once all the images are validated, the user is redirected to the Profile page, where he can check his score for the game and the rank from the leaderboard.

The game is developed in two versions: In the first version, the memory game has to be completed in order to access the tag validation page. In the second version, tag validation can be done as soon as a match is found in the memory game. Both versions of the game are explained later in this chapter.

3.3 SYSTEM ARCHITECTURE AND FUNCTIONALITIES

The game described in this dissertation was developed as a web application and includes several components that allow the implementation of the necessary functionalities. To this end, various technologies and programming languages were used, including:

- Apache Server xampp v7

- Server Database MySQL
- MySQL Workbench 8.0
- Programmin Languages: Javascript, Hypertext Preprocessor (PHP), Hypertext Markup Language (HTML) e HTML5, Cascading Style Sheets (CSS), Asynchronous Javascript and XML (AJAX).

The application architecture developed is essentially divided into two parts, namely, the client and the server. The client side application corresponds to the Web interface, developed in HTML, with which the user interacts, and through which information is collected from the system. These user interactions with the system are collected through a set of Javascript scripts, with AJAX integrated, and PHP. Thus, the gathering of information becomes totally imperceptible to the player as there is no need to constantly reload the Web page each time information is sent to the server. The information gathered, both implicitly and explicitly, is later processed by a set of functions developed in PHP, and which interact with the MySQL database server. Figure 11 shows the representation of the system architecture.

The server is made up of an Apache Xampp web server, responsible for Web pages, and a MySQL database server, which has the function of storing information relating to the data recording of users, and the entire annotation process. This also includes storage of scores gained by the users.

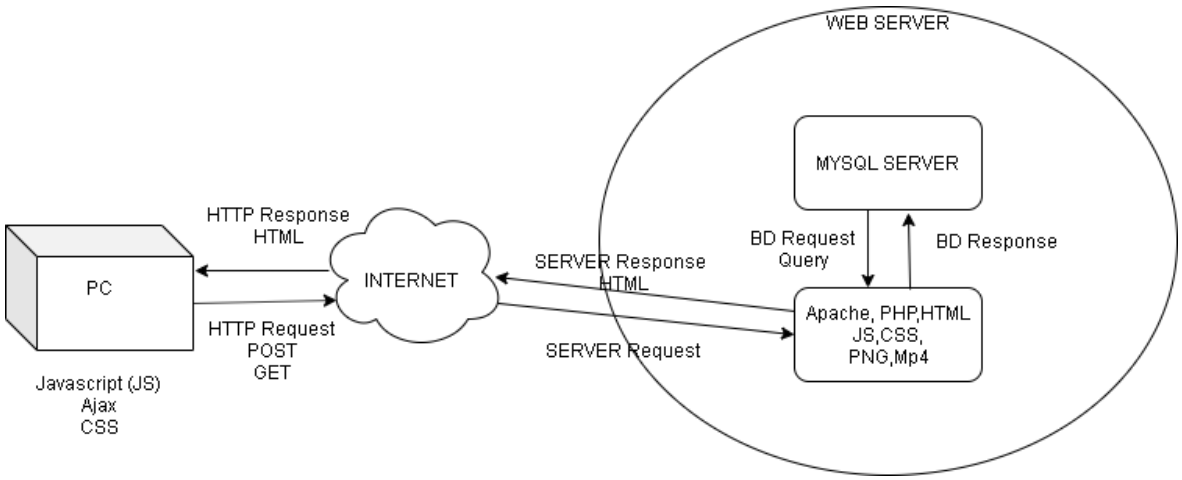


Figure 11 System Architecture

The architecture can be divided into three modules: Login/registration module, memory game and the validation module. These modules are explained in detail in the following topics.

3.3.1 DATABASE

The application support database has one table named users. The users table allows you to store all user data that is currently registered including:

- The username used at registration
- The password created, which will be encrypted the moment it is entered in the database
- Email address

Along with this, the score obtained by the user from the game is also stored in the table with respect to the username of the user. When the user submits the score, the score gets updated in the `score` column and also the cumulative score (sum of all scores) is saved in a separate column, `Sum_score`.

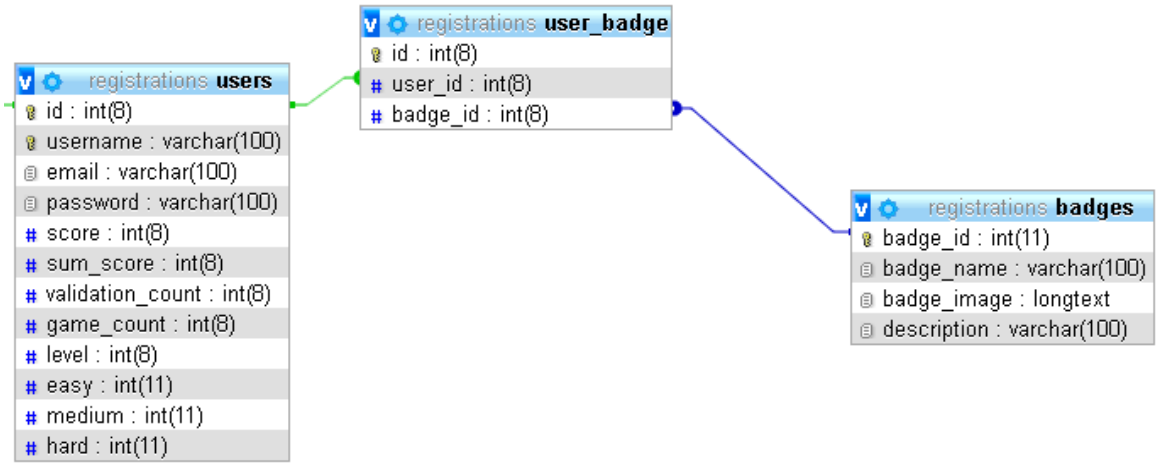


Figure 12: Database Structure

The badges table stores the badge names and the images. The `badge_id` from badges table and the `id` from `users` table are linked with the `user_id` and `badge_id` present in the `user_badge` table. The `user_badge` table holds the information about the users and their respective badges they received.

3.3.2 REGISTRATION AND LOGIN SYSTEM

In order to identify each of the players, a fairly comprehensive registration system is available which allows users to access the game and thus identify each one individually, controlling their scores, so that they enjoy greater entertainment.

Each new player must create an account by filling in the following data: username, email, and password. Once submitted, the player is referred to the page as shown in the Figure 13, where the user can log in. In case of login, the player will be taken to his private area as shown in Figure 14, where the user can enter the game and check the rank from the scoreboard. Also there are options displayed throughout the game as a menu form, allowing you return to the homepage at any time about the game.

The game as a guest was even considered for implementation but was quickly discarded as this type of login does not encouraged the player to play again in the future as their score would not be stored. In Figure 15 we can see the registration page created.

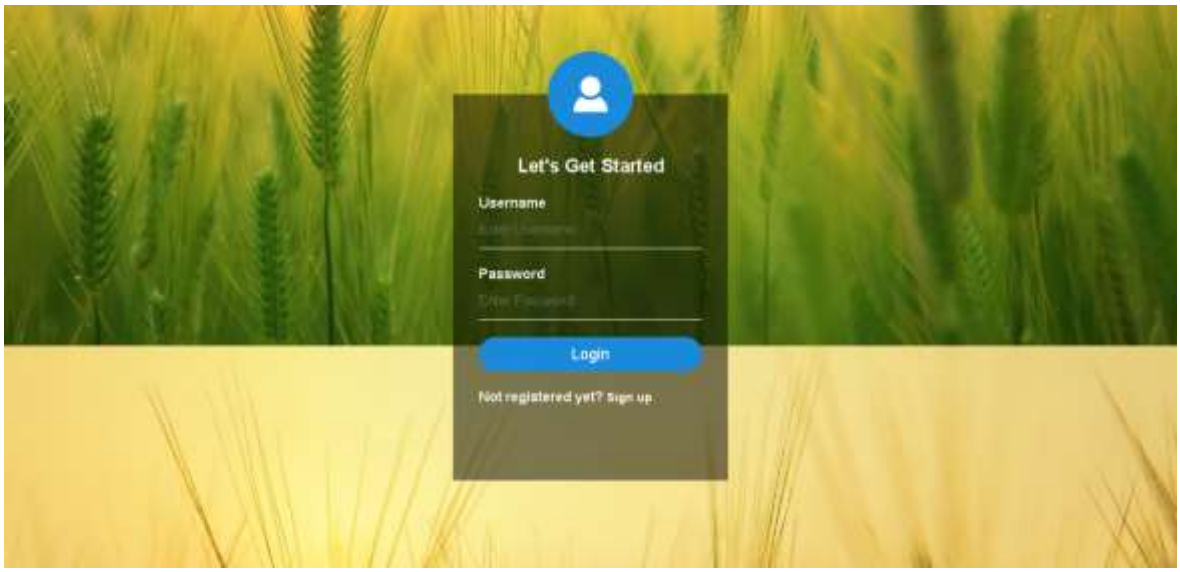


Figure 13 Screenshot of login form

Once the player has logged in, in addition to providing information about the game and the corresponding button that will start the game, there is also another option: Scoreboard (see Figure 14) that will display the leader board of all the users and their ranks.

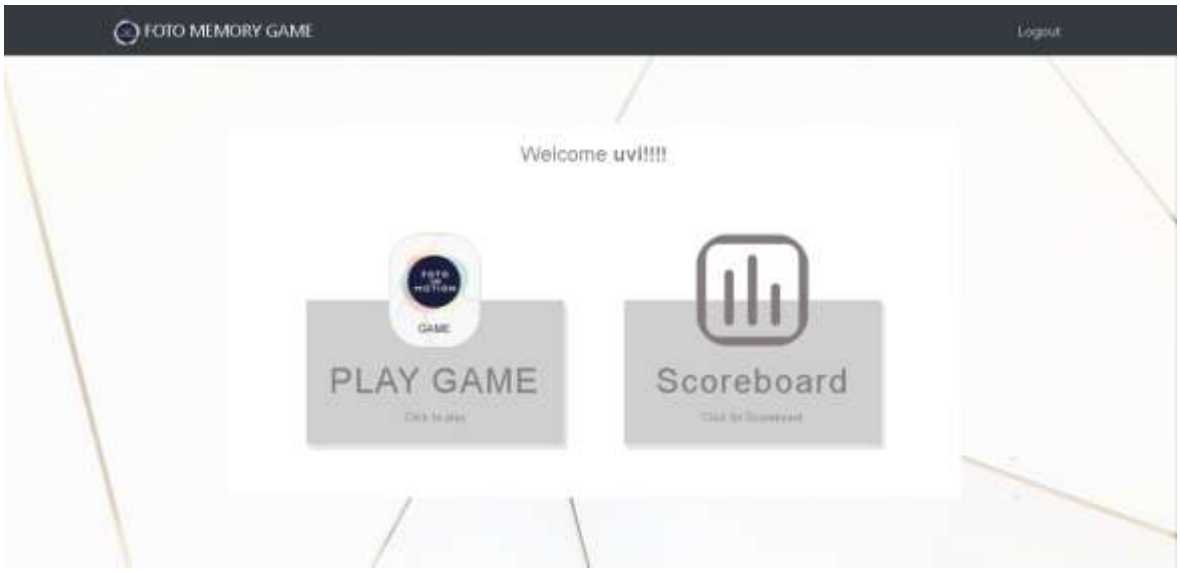


Figure 14 Screenshot of user profile page

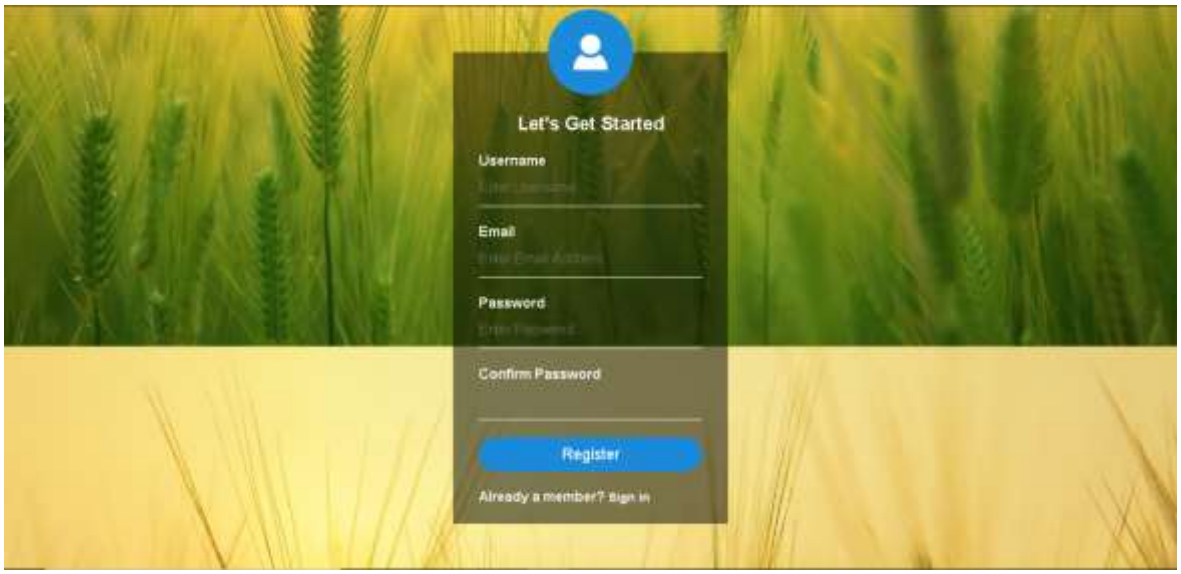


Figure 15 Registration Page

At the beginning of the game some information about the game is presented. This information is intended to make the player aware of how the game is handled. The page also provides information about the objectives as well as some of the instructions and advice to follow in order to achieve better results. Once the player is ready, he/she can start the game.

The login and registration page were developed with PHP and HTML. In the registration page, the form was developed with the help of HTML. PHP is used to make the connection with the database and store the details entered in the registration form to the database table. At login page, the PHP code checks the database table for username and password with the values entered in the login form. If found the values in the database table, the code is directed to profile page. If the values are not found in the table, the program provides a message 'invalid username/password'.

3.3.3 GAME MODULES

The game modules are created to achieve better game presentation and make player interaction more fun.

Throughout the game, a menu will follow the player on all pages. This was created to allow easy navigation to main pages of the application. The menu has the following options:

- Home: allows you to return to the profile page and ensures the possibility of start a new game or view the information about the leader board.
- My Profile: gives the information about the user achievements and the badges received through various activities in the platform
- Logout: Allows the player to log out of the game, and to be re-routed back to home page.

There are three modes of difficulty that can be chosen by the user before playing the game.

- Easy
- Medium
- Difficult

In easy mode, the game is with 12 cards and the objective is to find 6 pairs in 60 seconds. In medium mode, the number of cards is increased to 16 cards and the objective is to find 8 pairs in 60 seconds. In the difficult mode, there are 16 cards in total and the objective is to find 8 pairs in 40 seconds.

The game has two versions: First version follows a stage by stage process i.e., memory game has to be completed to access the tag validation page. The second version follows a parallel approach, where the tag validation can be done as soon as a match is made in the memory game.

As regards to the main game page in the first version, it can be divided into two parts, which are distinguished based on the interactions made by the player and the way they are presented the data. In Figure 16 we can see the final aspect of the game page developed.

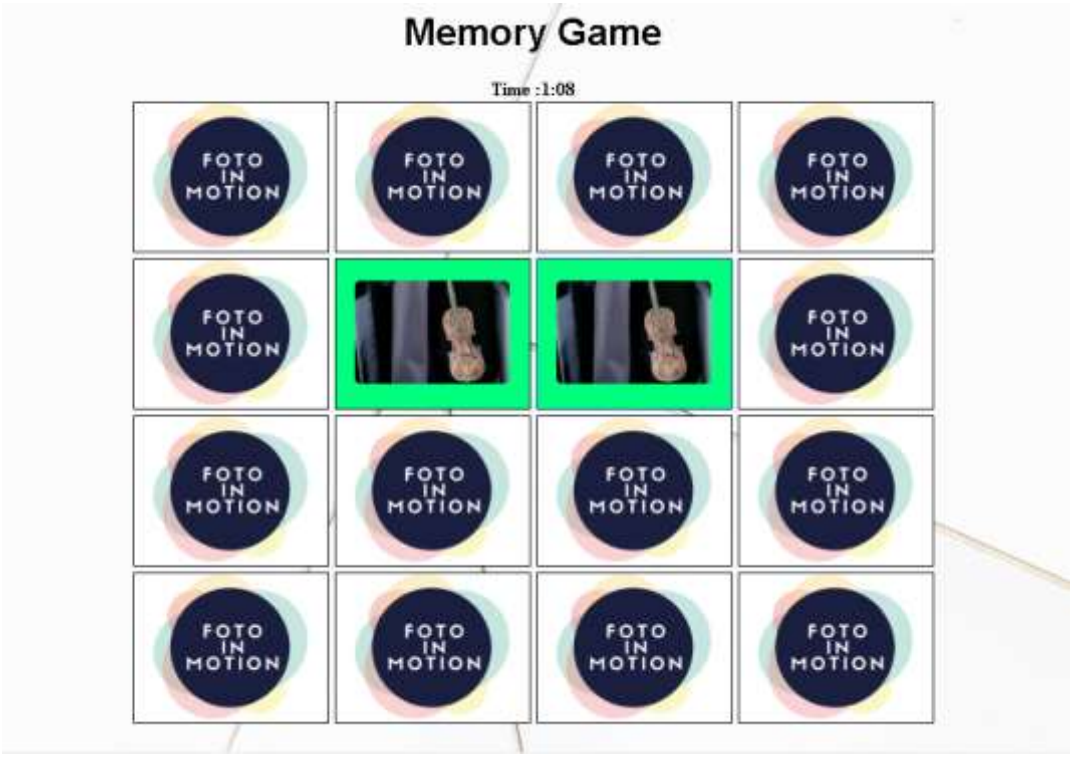


Figure 16 Game Page -version one

- At top center of the page, the time still available for game (initially set at one minute and 30 seconds and decreasing until it runs out and the game is over).
- In the central part, is the memory game which contains the images retrieved from the image directory which contains the dataset produced by the annotation tool (Foto in Motion). So, each time the player enters the game, new set of images are retrieved from the image directory and placed under the cards of the game.

As soon as the game ends, the scores are displayed to the user as shown in Figure 17. The score generated is based on the number of moves made and time taken to find all the pairs. If the user is satisfied with the score, he/she can continue to next part of the game by submitting the score. This score is added to the user's database.

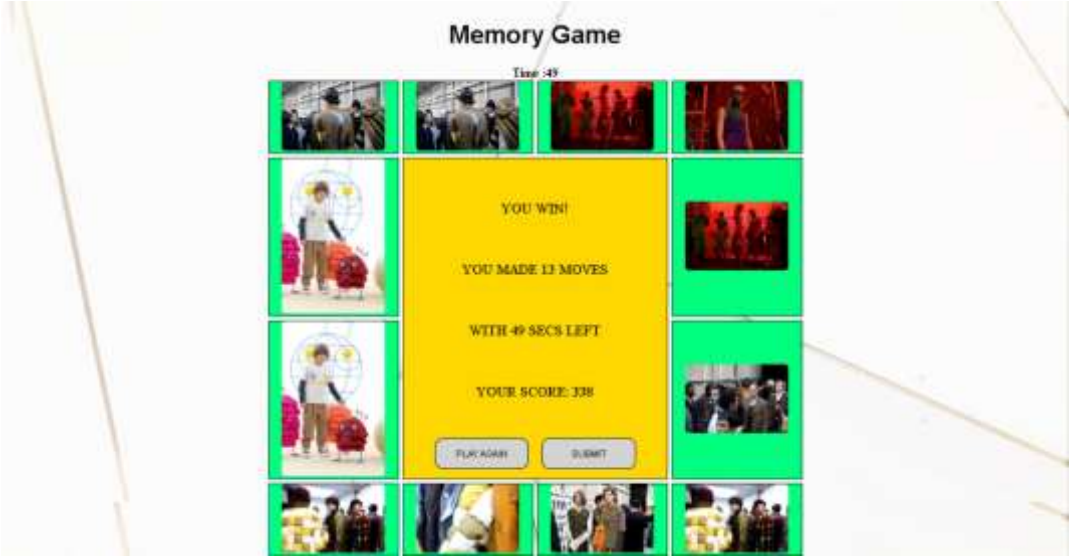


Figure 17 Score page of the game

Once the score is submitted, the user is directed to another page, where he can see the images used in the game.

The big brain of the memory game was developed in Javascript and is responsible for handling much of the information that is presented. The main modules that characterize the system are:

- Display and randomness of images
- Game time, session cookies and GameId
- Get score
- Presentation of result

PHP is used in the first module to fetch random images from the dataset produced by the annotation tool and sends to the Javscript where the images are duplicated, shuffled and placed below the cards. In the second module, PHP checks for the session and holds the data related to the session. HTML is used for the setting up the game platform. HTML5 events frameworks are used to get the event elements when the cards are clicked. Javascript holds the logic for the game to check the matches in the third module and the score is obtained in the fourth module. The flowchart of the memory game is shown in Figure 18.

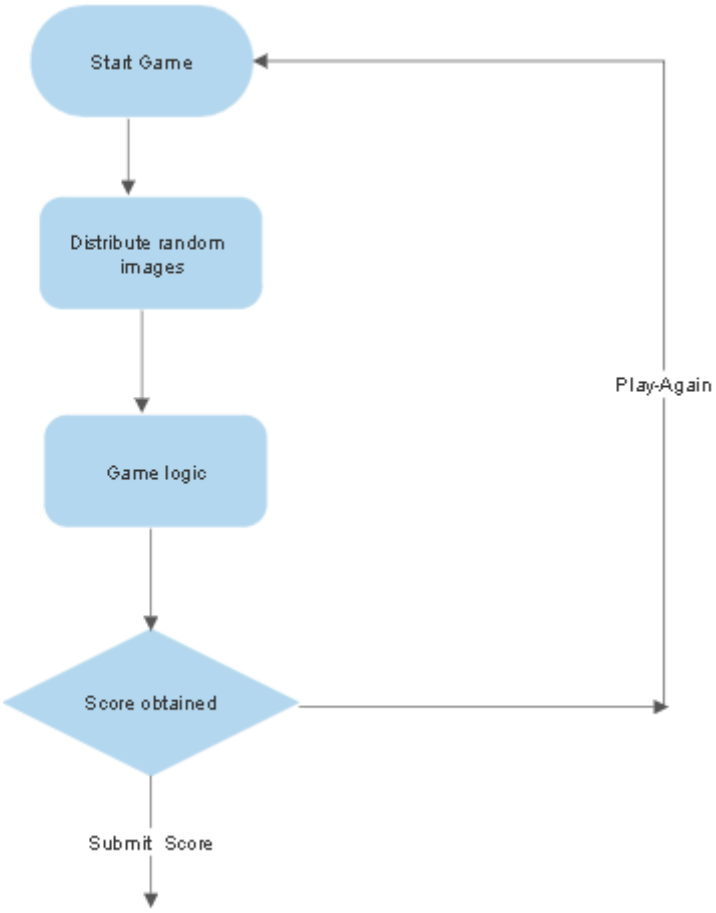


Figure 18 : Flowchart for memory game

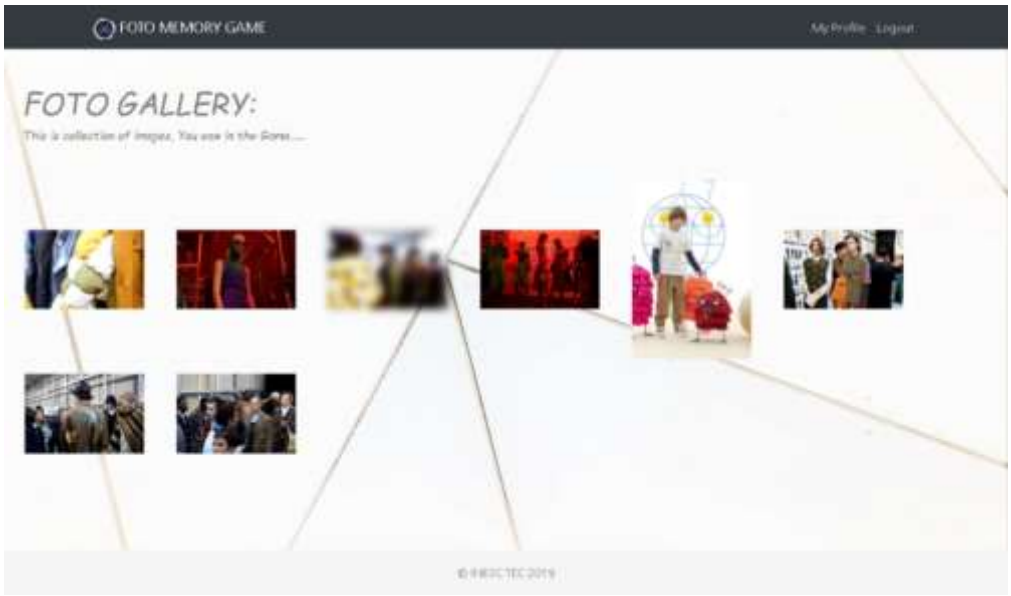


Figure 19 Page that shows the gallery of images used in the game.

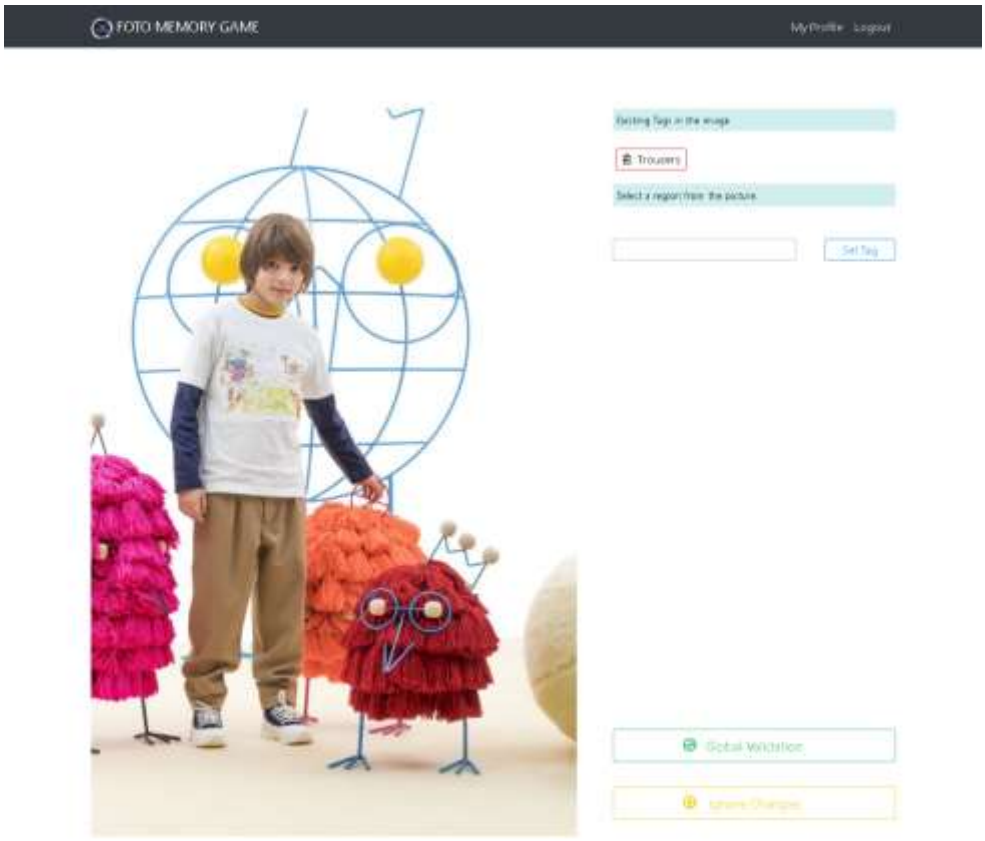


Figure 20: Tag validation page.

In the tag validation page, the user can see the existing tags from the image in a separate section. The tags already present in the image are obtained from the xml file related to the image present in the dataset and provided by previous annotation tools. The user can delete the existing tags present in the image and add new tags. In Figure 20, the user is on the page with the selected image and an option to add a new tag. The user will be able to tag the parts of the image, which he remembers from his visual memory while playing the game. In Figure 21, the user finds the important part of the image from his visual memory and makes a new tag.

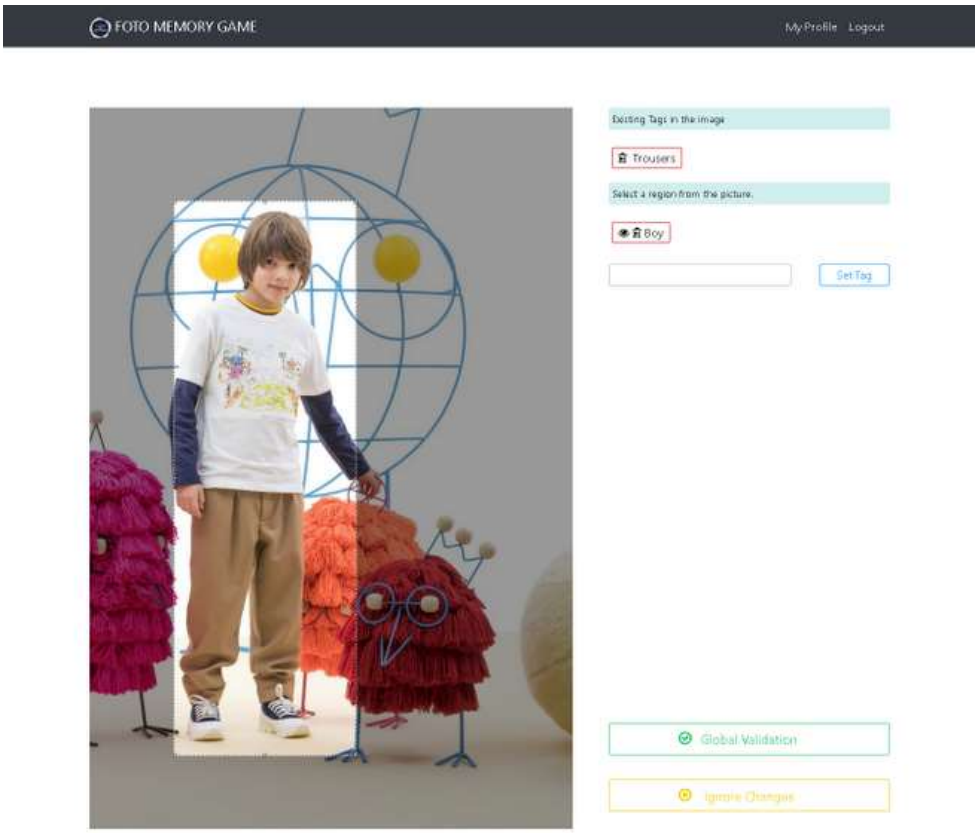


Figure 21: Inserting a new tag

Once the user is satisfied with the tags of the image, he can confirm it by clicking the global validation button. When the button 'Global Validation' is clicked, the tags and the position of respective bounding boxes are updated to the xml file of the image present in the dataset. Also it sets a flag value in the xml file of the image, to show that it is validated. Once the image is validated, it cannot be validated again i.e., the image is blocked of validation from other users. After the global validation of the image, the user is redirected to the gallery page, where he can find the image blurred indicating that the image is validated and blocked from future validation. Figure 22 shows the gallery page after the image is validated. The "ignore changes" button present in the tag validation page, just ignores all the action performed on the page and redirects the user to the gallery page.

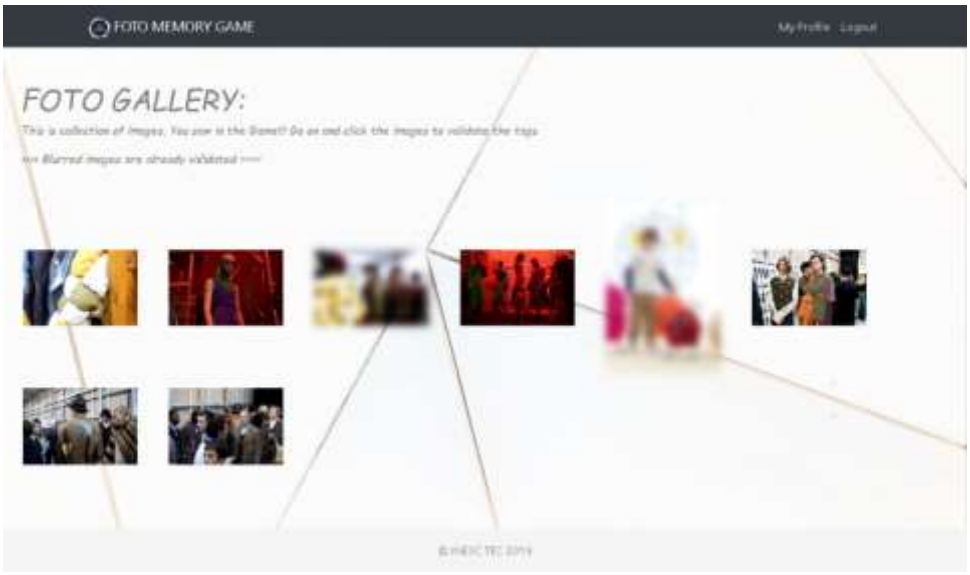


Figure 22 Gallery after tag validation

It is not compulsory that the user has to validate all the images present in the gallery. He can go back to the profile page to check the scoreboard or play the game again or just quit the session anytime with the help of the options present in the menu.

The tag validation module was developed with combination of HTML5 and Javascript languages to provide an interactive experience. Along with this, PHP and Ajax are used to retrieve the information from the xml files of the image present in the dataset. All the tags present in the image are saved as objects in the xml file. When the page is loaded, the image is displayed, and the respective xml file is fetched with help of Ajax scripts and PHP. The PHP code first reads the filename to check if it is the right xml for the image. The xml files present in the dataset have the same filename as that of their respective image filename. i.e., for an image (image1.jpg) its xml will be (image1.xml). These xml files are auto generated from the annotation tool. Then the PHP code reads each and every element present in the xml file and retrieves the objects (image tags) and displays them in the top right section of the page.

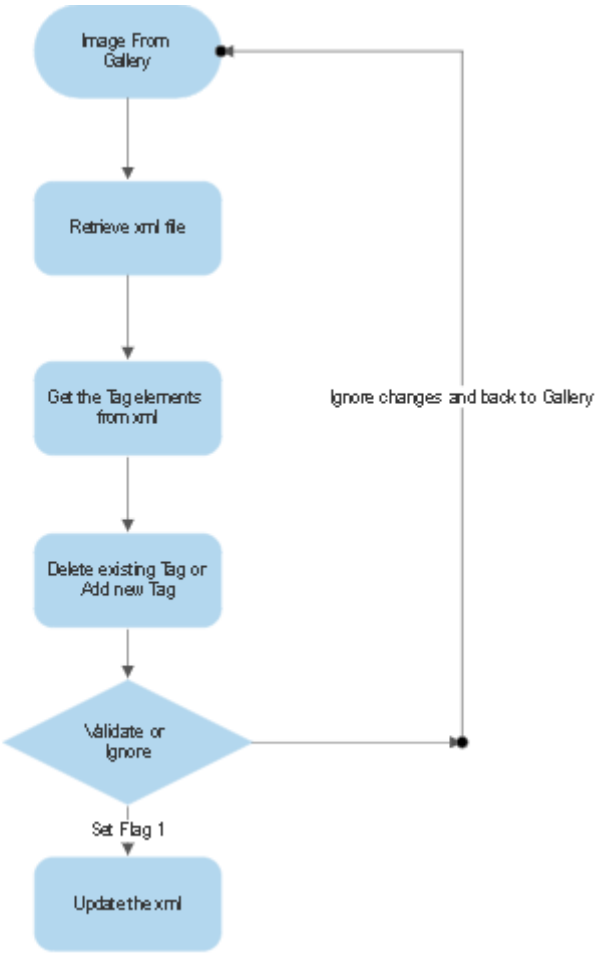


Figure 23: Flowchart for the tag validation module

When a tag is to be set, a bounding box shall be drawn over the image first. The tag cannot be set without the bounding boxes and a message will highlight this requirement.: 'Select a region from the picture'. Once the tag is set, it is added as a new object to the xml file of the image. The new object is named under the type 'Focus Area', so it can be differentiated from the existing tags. When the 'Global Validation' button is clicked, the PHP will set a flag 'approve' to '1' and saves the xml. Next time, when the program reads the xml files to display the images in the gallery and if it finds the flag value set to '1', it will block the image from entering the validation process again. Figure 23 shows the flowchart of the tag validation module.

As regards to the main game page in the second version, it can be divided into two parts, the memory game and the section for displaying the matched image from the game. Annotation of the image can be done as soon as the match is found. Figure 24 shows the game page of the second version.

Figure 24 Game page – second version

This version of the game does not have time constraints, which allows the player to find the matches with less pressure, thus providing the user more freedom to focus on the visual attention of the image. Once a match is found, the user can annotate the image at the same page. For every match made, the image is updated at the side, and it is open for annotation. If the matched image is already annotated, the image will be blocked from any further annotations.

The two versions of the game were developed with same objective: to find the focus of attention on the image. The user can try both versions of the game and annotation results can be compared to understand which version of the game is more preferred by the user for finding the focus of attention on the image.

The user can keep track of his scores and achievements in the profile page. The user unlocks badges on completion of certain tasks like number of games played, number of images annotated and also badges on level achieved. The badges and the descriptions are shown in Table 1. In the database, when each condition is satisfied, the badges are linked with the user `id` in the `user_badge` table. These badges are showcased in the profile page. The details about the badges can be seen if clicked. Figure 25 shows the profile page of the user, where the milestones and badges earned are displayed. As the user levels up, the complexity of the images that are used in the game also changes i.e., when user is in the newbie level, he is presented with simple images which has one main focus object like an image of a person

standing. But as the user levels up, say he is in Expert level, he is presented with images that can have more than one focus area like an image with group of people or images from events, here annotating the focus areas will be quiet challenging. Thus, making the game more interesting and challenging for the user.

Table 1: Badges and description

Badges	Description
First Game	Play first game
First Validation	Make one validation
15-Easy	Play 15 easy mode games
7-Medium	Play 7 medium mode games
5-Hard	Play 5 hard mode games
25 Games	Play a total of 25 games
50 Validation	Make 50 validations
Newbie, Rookie, Regular_User, Expert, Professional	Level up rewards

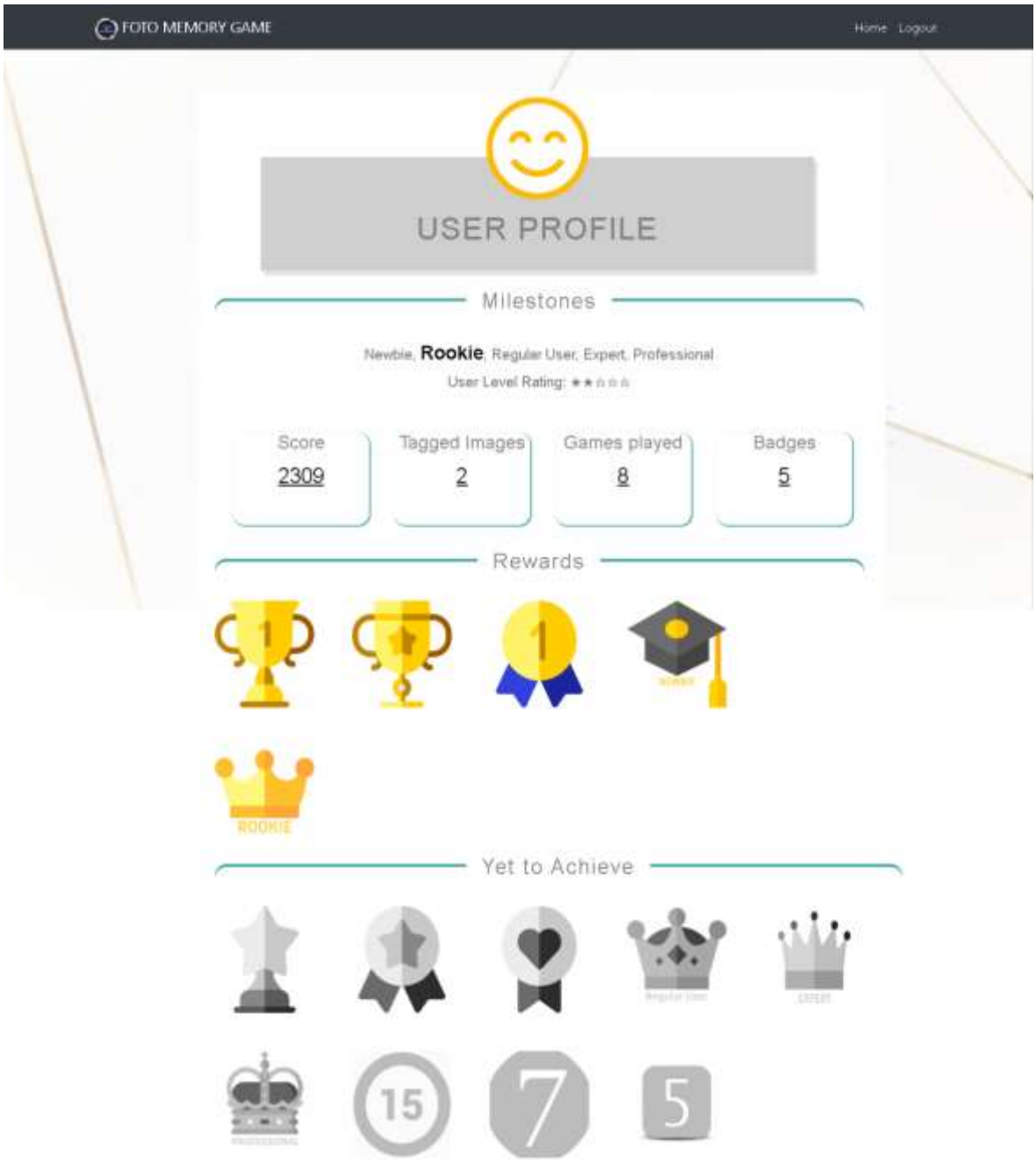


Figure 25 Screenshot from user achievements

The user can look at the points he gathered by various activities in the platform by clicking on the scoreboard in the user home page shown in the Figure 14. The scoreboard showcases the scores of all the users of the system and their rankings. The score is stored in accumulative way. The user can also view the score of his last activity in the scoreboard. Figure 26 shows the scoreboard with all users and their ranks in the overall table.

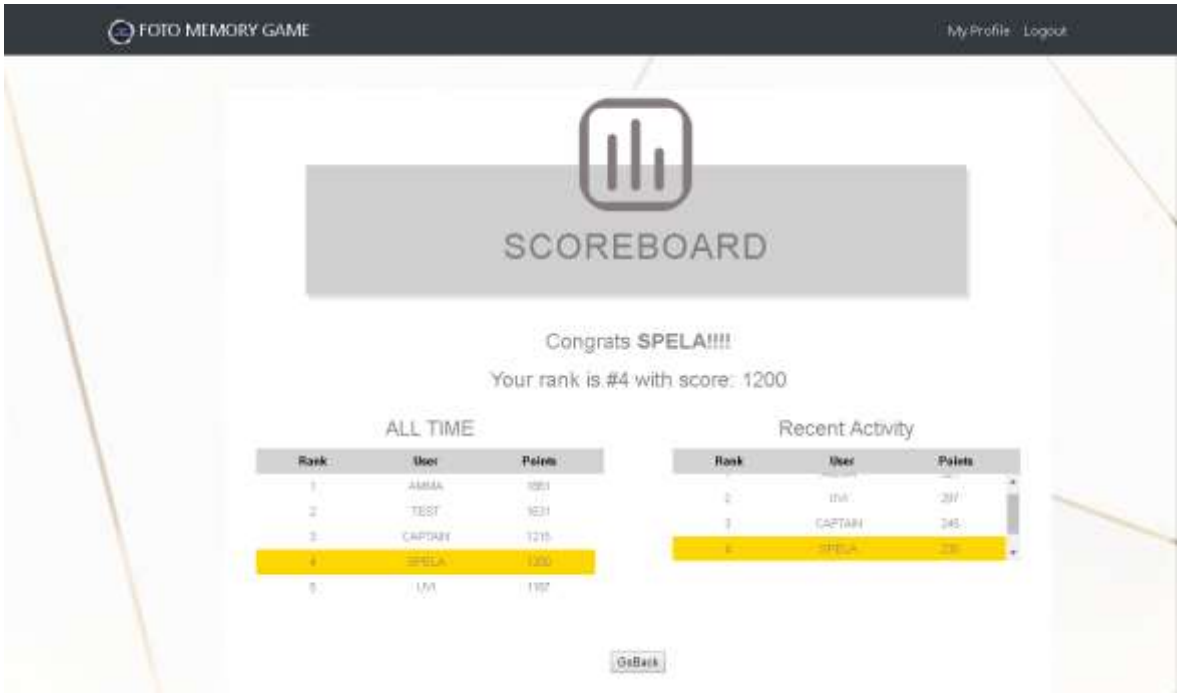


Figure 26 : Scoreboard Page

4. CONCLUSION

Several applications have been developed over the last few years for content annotation using collaborative approaches. The work described in this dissertation sought to study the area of collaborative content annotation systems and proposed a solution based on the concept of game with the purpose to validate a set of images obtained from the annotation tool from Foto in motion. The game uses Human visual memory and process that information and help in tag validation of the images. Each player's personal information and scores are saved, so that each player has an individual personal space, which enhances the user participation. The system presented is based on a distinct collaborative annotation technique: it has a traditional memory game followed by a tag validation.

Compared to the other systems that have been referenced throughout this dissertation, the developed system presents a new GWAP that is not present in any of these systems. This is reflected in the fact that the system created to validate the tags, which in the future will facilitate access to many information, as it is made more accurate with the help of the information gathered from human visual memory. The developed application has satisfied the objectives initially made.

However, certain set of features can be considered for the future development:

- Tag recommendation system: A tag recommendation system could also be implemented, so as to assist the user in adding new tags to the image.
- Multiple validation: The application can be upgraded to next stage by introducing multiple validation. The present system allows only one validation per image. This can be changed to multiple validation- where an image after first validation will store its value in a table , so that when it is validated again by some other user, both the values can be checked and save only the values which are common. By this way, user's interaction with the application is made more competitive, also it can increase the accuracy of the validation.

- Improved Game levels: With the existing game levels, certain additional feature could be added like, when a user reaches certain level, the tags made by the user can be checked with that of the annotation tool, to check if the tags made on focus regions are same in both files. This can provide much better annotation on the image.
- Testing the game versions: The two versions of the game could be studied by providing a testbed for a set of users and the results could be compared to find the best approach that will be preferred by the users.

References

[1] Ahn, L., Dabbish, L. Designing games with a purpose. Communications of ACM, Vol. 51, Issue 8. pp. 58-67,2008

[2] Thaler, Stefan, et al. A survey on games for knowledge acquisition. STI technical report, University of Innsbruck 2011

[3] Law, Edith, Von Ahn, Luis. Input-Agreement: A new mechanism for collecting data using human computation games, Conference on Human Factors in Computing Systems, CHI'09, pp. 1197-1206, USA, April 2009

[4] VideoTag available at http://www.videotag.co.uk/ accessed on May 2019

[5] Oomen, Johan, et al. Emerging Practices in the Cultural Heritage Domain - Social Tagging of Audiovisual Heritage, WebSci10: Extending the Frontiers of Society On-Line, April 2010

[6] HildeBrand, Michiel, Van Ossenbruggen, Jacco. Linking user-generated video annotations to the web of data. 18th International conference on Advances in Multimedia Modelling pp. 693-704, 2012

[7] Riek LD, O'Connor MF, Robinson P, Guess what? a game for affective annotation of video using crowd sourcing. In Proceedings of the 4th international conference on affective computing and intelligent interaction-volume part I. Springer, Berlin, pp 277–285, 2011.

[8] Van Zwol, Roelof, Garcia, Lluis, Ramirez, Georgina, Sigurbjornsson, Börkur, Labad, Marcos. Video Tag Game, WWW 2008, April 2008.

[9] Viana Paula, Pedro Pinto Jose, TAG4VD - A Game for Collaborative Video Annotation, Conference: ImmersiveMe 2013 - ACM International Workshop on Immersive Media Experiences, Barcelona, October 2013.

[10] Viana Paula, Pedro Pinto Jose, A collaborative approach for semantic time-based video annotation using gamification, Journal Human-centric Computing and Information Sciences, Springer-Verlag London, UK, Volume 7 Issue 1, December 2017

[11] Brooklyn Museum- Tag! You're it!

https://www.brooklynmuseum.org/community/blogosphere/2008/08/01/tag-youre-it/

[12] Louis Von Ahn, Ruoran Liu and Manuel Blu. Peekaboom: a game for locating objects in images, SIGCHI conference on Human Factors in computing systems (CHI '06) 55-64. ACM, 2006

[13] Seneviratne Lasanth, Izquierdo Ebroul, Image Annotation Through Gaming, 16th International Conference on Digital Signal Processing, July 2009.

[14] Chien-Ju Ho, Tao-Hsuan Chang, Jong-Chuan Lee, Jane Yung-jen Hsu, and Kuan-Ta Chen, KissKissBan: a competitive human computation game for image annotation, ACM SIGKDD Explorations Newsletter, 12,1: 21-24, 2010.

[15] Christopher G.Harris, ClueMeIn: Enhancing the ESP Game to Obtain More Specific Image Labels, CHI PLAY'18 Extended Abstracts, Melbourne, Australia , Oct. 28–31, 2018

[16] Youtube, available at http://www.youtube.com accessed on May 2019

[17] Viddler - Video Hosting and Sharing, available at www.viddler.com accessed on May 2019

[18] Firas layth Khaleel, Noraidah sahari, Tengku Siti Meriam Tengku Wook, Amirah Ismail- "Gamification Elements for Learning Applications, International Journal on Advanced Science, Engineering and Information Technology, Vol. 6, No. 6, 2016, p. 868-874, 2016.

[19] Vijayabaskar.P.S, Shanmuganathan AravindaSiva, Vignesh, "Knowledge management using Gamification", international conference on innovations in information Embedded and Communication Systems (ICIIECS'16), 2016.

[20] Jakub Šimko, Michal Tvarožek, Mária Bieliková – “Human computation: Image metadata acquisition based on a single-player annotation game”, International Journal of Human-Computer Studies, 933–945, October 2013