

TAG4VD – A Game for Collaborative Video Annotation

José Pedro Pinto
INESC TEC
Rua Dr. Roberto Frias, 378
4200-465 Porto, Portugal
jppinto@inescporto.pt

Paula Viana
Polytechnic of Porto and INESC TEC
Rua Dr. Roberto Frias, 378
4200-465 Porto, Portugal
paula.viana@inescporto.pt

ABSTRACT

Creating descriptive labels for videos is an important task, with application in video retrieval, Web accessibility and computer vision. However automatic creation of such labels is difficult and, alternatively, having professionals manually describing content is too expensive. Engaging end-users in the process of describing multimedia assets may lead to good results and enables creating the sense of participation which is currently one of the key factors to attract customers to a service. The existing approaches are highly successful in terms of number of engaged players and number of collected labels, but hardly create comprehensive tag sets, contributing both with generic or too narrow meaning tags. “Games With A Purpose” are one of the approaches that have been used in an attempt to create comprehensive video descriptions by harnessing the intelligence of human players and have them contributing and collaborating towards a common goal that is recognized if successful.

This paper describes a game which implements two mechanisms for collecting data via human-based computation games. Tags introduced by registered players, in a given timecode, are validated based on a collaborative scoring mechanism that eliminates irregular annotations. Additionally, a voting mechanism that enables players to endorse or refuse existing tags, provides an extra instrument to guarantee the quality of the annotations.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous; H3.3 [Information Search and Retrieval]; I.2 [Artificial Intelligence]: Applications and Expert Systems

Keywords

Games With A Purpose, crowdsourcing, tagging, video annotation

1. INTRODUCTION

Access to large assets of video content is often difficult due to the lack of methods that enable accurate access to specific timecodes. Content-based information retrieval techniques grounded on video processing to extract crucial information are still not feasible

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ImmersiveMe'13, October 22 2013, Barcelona, Spain
Copyright 2013 ACM 978-1-4503-2402-1/13/10...\$15.00.
<http://dx.doi.org/10.1145/2512142.2512154>

given the computer resources required and the difficulty to develop a method that is universal to any type of content. Video retrieval is still mostly based on textual information that may include the specification of a category, in which the object is inserted, or a small description. Having detailed descriptions that enable random access to crucial points of the video becomes unfeasible due to the great complexity and costs associated. Furthermore, in some cases it would require professionals that continuously verify and sort the contents to be published, requiring not only a great investment and cognitive effort from them, but also highly qualified personnel. As an alternative, crowdsourcing mechanisms have been used to improve indexing and categorization of audio-visual content by adding alternative textual information (usually tags) to the one that already exists. Games With A Purpose (GWAP) may contribute to enhance this collaborative annotation process by including entertainment mechanisms (competition, prizes, etc.) that contribute to the engagement of contributors.

This paper presents a game developed to enable the introduction of tags in specific time instant of a video clip. Filtering and validation of tags is achieved through a collaborative process in which players are awarded bonus if their information is confirmed by other members of the community.

The solution developed has some advantages when compared to previous published work. Besides improving the scoring and validation mechanisms, introducing new algorithms and features that contribute for a better collection of metadata linked to timecodes, all the media players are HTML5 based making portability to other devices (mobile or even smart TVs) seamless.

2. RELATED WORK

Engaging player in a game requires the implementation of mechanisms that motivate its users. Some aspects that can contribute to increase the effectiveness of the game include [11]:

- *Time limit*: this acts as a challenge to the player. In some cases the player is rewarded with bonus points according to his performance.
- *Score*: one of the simplest motivation methods. One usual approach is to award pairs of players that label the same content with the same statement.
- *Levels of skills and/or difficulty*: challenges increase depending on the points accumulated. New players or players with low scores are requested simple tasks. Achieving certain goals is required to advance to the next level, where the difficulty is higher, but the likelihood of increasing the score is also higher.

- *Ranking*: a list of players which obtained the best scores is displayed. This type of information can be listed as daily, monthly or ever. Alternatively the listing can distinguish the best players at different levels.
- *Randomness*: in a multi-player game each time a player logs in, he is challenged with another one which is unaware of his identity. The choice of partners is performed randomly, to avoid cheating. This approach also allows matching players with different difficulty levels, making the game more attractive and interesting.

The idea of using games in the context of multimedia annotation was first introduced by Luis von Ahn with the ESP Game [1]. This multiplayer game randomly linked two players and made them match their guesses (tags) on the same image within a limited time period [3].

Matchin [3] main objective is to classify the quality of the images in a data set according to user inputs. For that, two different pictures are presented to the players who are asked to choose the one that they think their partner would prefer. A score is given if they agree with each other. An additional bonus is awarded if a certain number of matches is achieved within the time limit imposed.

Herd It, and TagATune are proposals for audio annotation. The first one is a real time two player game where different tasks related to music classification are to be performed: state of spirit of the music, identification of the artist, the name of the song or even the genre of music. The more consensual the answers are, the higher the score players will receive [1][3]. TagATune is also a two player game used to tag music clips. Players listen to a music clip and try to describe it the best way they can, so that the other peer can decide if they are listening the same music clip or not. A right guess enables scoring [3][4].

The task of video annotation was also implemented as a game in a set of experiences. VideoTag is a one player game which has the objective of annotating Youtube videos presented to the player within a specific period of time. Tags can be freely introduced by the player, ensuring a better score, or he can ask for a suggestion to help him tag the clip. When using this facility, he will be penalized for that. The game has five different difficulty levels that include having more videos to classify within the same time interval [5]. However, not much detail is provided concerning the scoring and tag validation mechanism used. Waisda? (which translates to "What's that?") is a video labeling game that invites users to tag what they see and hear. Points are awarded for a tag if it matches one introduced by the opponent player. As in the other games, a limit of time for each round is imposed and score can be improved if some considered crucial information is provided: geographical location of the scene, identification of people or match the tags introduced by other players within a time frame of 10 seconds [8][9]. By using just a two players approach to validate tags, scoring may not be fair as global information is missing and depending on the adversary. In TubeLink, another one player game, proposals of tags are presented together with the video. The player is given the opportunity to match tags with scenes and scoring requires a minimum number of accurate tags [7][10]. OntoTube tries to annotate random Youtube videos. The same video is presented to two players together with some pre-defined questions, such as "Is this video fiction or non-fiction", "is this black and white or color", etc. Players may always answer the questions or skip them. If they agree on an answer they earn

points and they may proceed to the next question. Validation of the input is based on consensus and majority of answers [3][11].

The "Yahoo! Video Tag Game" is a multi-player game based on temporal tagging agreement between players. Scoring can range from 2 to 20 points according to the temporal distance between tags [13]. This mechanism doesn't however aggregate the opinion of a set of players but just looks for the closest tag introduced by any individual player. Another limitation is related to the fact that an user has to wait for another player to join. This may result in the waiver of potential contributors.

Besides the games used to create video metadata, some other well-known applications such as Youtube [14] and Viddler [15] rely also on user generated tags to annotate content. However, some limitations can be pointed: restrictions on the users who are allowed to annotate content; not always associating tags to timecodes; not effectively using them by searching engines and no validation mechanism for the metadata inserted.

Tag4VD proposes a solution that is able to improve some of the main drawbacks found by implementing a complex scoring and validation mechanism that includes the functionality of collaboratively rejecting previously introduced tags and using aggregated information from groups of players. These mechanisms contribute to enhance the quality of the metadata accepted by the system. Moreover, by using HTML5 to implement all the video related functionalities, instead of proprietary players requiring the installation of plugins like the ones used in Waisda? and Yahoo! Video Tag Game, portability is considerably improved.

3. TAG4VD – A Video Tagging Game

TAG4VD (Tag for Video) is a one player game, in which players are asked to label some videos that are presented within a time interval.

Tags introduced are associated to specific time instants contributing to enhance the search mechanisms and to direct the results of a query or browse of the dataset to the right instant of the video clip. For the player to be compensated for his contribution, the system has to validate the tags. This is done taking into account the number of times a given tag is used in the same or similar clip point. To turn the game fair, the system keeps track of all the tags inserted and scoring can be assigned to a player both while a game is active or offline, if additional information confirms previously introduced tags.

3.1 Scoring Mechanism

The intelligence of the game resides in the scoring mechanism that awards players based on how close in time a reasonable number of players introduce the same tag to describe a scene. Tags are then associated to a timecode, allowing a greater accuracy in the description of the video clip.

Validation of a tag for the attribution of points is based on three concepts: the tag itself, the clusters of existing tags and the number of times that tag appears in the respective cluster. Figure 1 depicts an example where these concepts are represented.

Clusters of tags are the main concept behind the scoring system and are used to associate tags that, although not introduced in the same exact moment, are somehow near each other. The same tag can co-occur throughout the video clip but be classified as describing a different happening. For that, the system keeps track

of all the tags introduced and of the respective timecode and defines clusters described by its centroid (the mean value of all the timecodes of the tags belonging to that cluster).

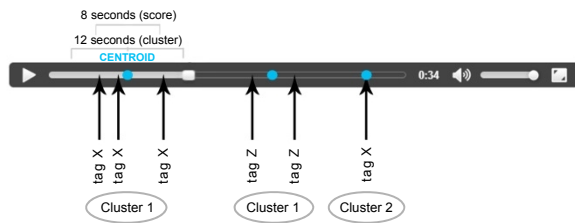


Figure 1 - Cluster and score attribution

Whenever a tag is introduced, the system verifies if that tag can be associated to one of the existing clusters or if a new cluster is to be defined (even if the tag already exists in the database). Assigning a tag to a cluster requires that its' timecode is within a pre-defined distance to the centroid of that cluster. That distance is established from the centroid point and will include all the tags within a distance of 6 seconds. To keep the information updated, centroids are re-calculated whenever a new tag is allocated to one cluster.

The action of introducing a tag can result or not on the attribution of a score, depending on the tag position, the centroid of that cluster and the predefined interval of more or less 4 seconds. Information is considered valid if it can be associated to a cluster having a dimension of more than three elements. In such case, a player will be awarded one of following score levels (Figure 2):

- 100 points for each tag inserted within a distance of 2 seconds to the cluster's centroid;
- 50 points for tags within a range of 3 seconds of the centroid;
- 10 points when the distance to the centroid is up to 4 seconds.

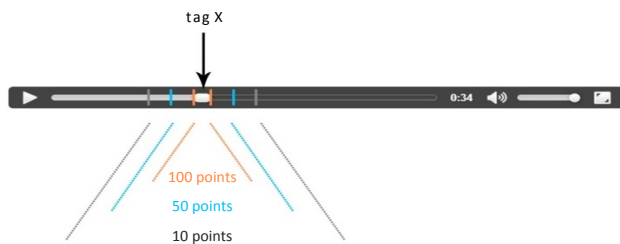


Figure 2 - Score assignment

In order to avoid a player to be penalized for being the first to introduce a specific tag which is later validated by several players (the phenomenon of "cold start"), an offline awarding system was implemented. This means that a player may be assigned zero points for a tag, during a game, because scoring conditions are not met, but be compensated beyond the game, when offline, if the necessary conditions for score attribution are reached. In short, when any player introduces a valid tag, the system checks all the other registered players that previously were assigned zero points for that precise cluster and upgrades his/her total scoring. Furthermore, an additional bonus is given for having been the first player to use that tag (instead of the usual 100 points, 200 points are given).

Due to the dynamics of the game, the centroid of a cluster may also change and this fact will also awaken the tracking mechanism responsible for updating players' scoring.

3.2 Playing TAG4VD

TAG4VD allows two usage modes: a limited functionality access to content, available to any users (guest), or a full mode usage requiring users to be registered.

Registered users are allowed to enter the contest by introducing tags that describe the content, as well as just watch available videos and use a voting mechanism that enables them either giving their opinion on the quality of the tags or on their preference to some instant of the video. Guest users will only have the two last functionalities available and are not directly engaged in the game. However, they may also provide useful information through the voting systems.

Navigating in the media archive is assisted by several approaches:

- a tag cloud (Figure 4) enables a first insight on the frequency of the tags while also providing a mechanism to direct the user to specific points of interest within the video;
- a search mechanism based on the metadata available in the system and that includes, besides the tags, structured info that has to be inserted when uploading a video. This includes a category and information on the creator;
- a "like" bar chart that is presented by sliding the mouse pointer over the video. These bars provide information on the time instants that users of the system have identified as the most interesting moments (Figure 4). By clicking one of the bars, users are routed automatically to that part of the video.

When entering the competition, users will be presented a randomly chosen set of video clips that they are required to annotate. The remaining time to complete this task, together with the list of tags that are being introduced and the score obtained by each of the actions are presented (Figure 3). Metadata introduced by the system manager or the creator is also shown as it may provide useful information.



Figure 3 - Game page

Once the timer runs out, or the player decides to end the game, the final game score, the total player's score and his ranking is presented together with a summary of his annotations within this game slot.

3.3 Additional Tag Validation Mechanism

Besides the tag validation mechanism described in the previous section, an additional feature was implemented. Registered or unregistered users are allowed to give their explicit opinion on the quality of a tag attached to a given timecode and contribute to enhance the quality of the annotation by enabling the system to discard “bad” tags introduced in the game mode.

While watching a video clip available in the asset, users can “like” or “dislike” a tag that is superimposed to the player within a balloon (Figure 4). Tags presented are chosen randomly among the existing clusters. The centroid is used as the presentation timestamp.

Besides voting on the quality of the tags, users are also allowed to provide information on his favorite moments within a video. This “like” information is used to increment the navigation facilities as system users may be automatically routed to the most impressive moments of the content by pressing the higher bars.



Figure 4 - Vote, preview, watch videos page

4. CONCLUSION

Several applications have been developed over the last years with the aim of enabling labeling of contents as videos, music, pictures, etc., turning access to these materials more efficient.

Tag4VD relies on a collaborative process that engages users in a game having as main objective to implement a tagging process. Labels may be freely introduced and players are compensated if their contributions are considered valid. The scoring mechanism takes into account all the past information and considers all the players and tags introduced, making the game fair and annotations more reliable.

Future work includes introducing tag recommendation mechanisms based on user profiles and the use of dictionaries that enable semantic classification/grouping of tags. Validation of the approach used is foreseen in two aspects: user satisfaction concerning functionalities and awarding mechanisms as well as tag validation. This last step requires the implementation of a large field test using already annotated content. Improvements in the scoring mechanism are also foreseen. This may include the implementation of a pitfall mechanism that penalizes those users that introduce too obvious tags just to win points. Evaluation of more complex tag-clustering membership approaches, taking into account the video semantics is also foreseen. In a first attempt to

improve this, the genre of the video clip and the scenes length will be used to adjust the required parameters.

5. ACKNOWLEDGMENTS

The work presented in this paper was partially supported by: Portuguese Foundation for Science and Technology within project FCT/UTA-Est/MAI/0010/2009; the North Portugal Regional Operational Programme (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF) and by QREN-AdI-SI-ID&T-13720.

6. REFERENCES

- [1] Ahn, L., Dabbish, L. Designing games with a purpose. *Communications of ACM*, Vol. 51, Issue 8. pp. 58-67, 2008
- [2] Bartholomaeus, Steinmayr, et al Karido: A GWAP for Telling Artworks Apart, 16th International Conference on Computer Games, pp. 193-200, 2011
- [3] Thaler, Stefan, et al. A survey on games for knowledge acquisition. STI technical report, University of Innsbruck 2011
- [4] 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
- [5] VideoTag, available at <http://videotag.co.uk> accessed on July 2012
- [6] GWAP, available at <http://www.gwap.com> accessed on June 2012
- [7] Games for semantic content creation, available at <http://ontogame.sti2.at/games/> accessed on July 2012
- [8] 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
- [9] Hildebrand, Michiel, Van Ossenbruggen, Jacco. Linking user-generated video annotations to the web of data. 18th International conference on Advances in Multimedia Modeling pp. 693-704, 2012
- [10] Siorpaes, Katharina, Simperl, Elena. Incentive, Motivation, Participation Games: Human Computation for Linked-Data. Proc. of the Workshop on Linked Data in the Future Internet at the Future Internet Assembly, Volume 700, Ghent, Belgium, 2010
- [11] Siorpaes, Katharina, Hepp, Martin. Games with a purpose for the semantic web, IEEE Intelligent Systems. Volume 23, Issue 3, pp 50-60, May 2008
- [12] Serious Game Initiative, available at <http://www.seriousgames.org> accessed on June 2012
- [13] Van Zwol, Roelof, Garcia, Lluís, Ramirez, Georgina, Sigurbjörnsson, Börkur, Labad, Marcos. Video Tag Game. In WWW 2008, April 2008.
- [14] Youtube, available at <http://www.youtube.com> accessed on June 2012
- [15] Viddler - Video Hosting and Sharing, available at www.viddler.com accessed on June 2012