Video Annotation by Cascading Microtasks

a Crowdsourcing Approach

Removed for double-blind review
Removed for double-blind review
Removed-for-double-blind-review

Removed for double-blind review
Removed for double-blind review
Removed-for-double-blind-review

Removed for double-blind review
Removed for double-blind review
Removed-for-double-blind-review

ABSTRACT

This paper presents a general approach to perform crowdsourcing video annotation without requiring trained workers nor experts. It consists of dividing complex annotation tasks into simple and small microtasks, and cascading them to generate a final result. Moreover, this approach allows to use simple annotation tools rather than complex and expensive annotation systems, and tends to avoid activities that may be tedious and exhausting for workers. The cascade microtasks strategy is included in a workflow of three steps: preparation, annotation and presentation. The preparation step describes how a complex annotation task can be divided into simple microtasks, in addition to present a workflow for the activities required before the annotation step, such as to define what should be annotated and the annotation types, as well to design the microtasks and the simple annotation tools to execute them. In the annotation step the annotation microtasks are performed by crowd workes, that are the contributors for the process. This step follows a workflow in which each microtask is followed by a specific aggregation method that generates a result, so that the output of a task feeds the next one. The presentation step displays the outcome delivered by the previous step, also at this point all partial results are available to be used in other applications. The approach introduced also allows the development of expansive video annotation systems in which it is easily possible to add new microtasks to improve the result and generate new results. To validate the presented approach was proceeded an experiment that consisted in to enrich videos using the crowd to add extra content such as images, text boxes and hyperlinks and positioning them over the video. To support the experiment was developed a toolkit that includes a set of Web-based annotation tools and aggregation methods, as well a presentation system for annotated videos. This toolkit is open source and can be downloaded and used to replicate this experiment and to construct different kinds of crowdsourcing video annotation systems.

CCS CONCEPTS

• Information systems → Crowdsourcing; **• Human-centered computing → Computer supported cooperative work**; **• Applied computing → Annotation**;

WebMedia'2017, October 2017, Gramado, RS Brazil

ACM ISBN XXX-XXXX-XX-XXX/XX/XX…$15.00
https://doi.org/XX.XXX/XXX_X

KEYWORDS

Crowdsourcing, Video Annotation, Human Computation, Microtasks, Video Enrichment

ACM Reference format:
Removed for double-blind review, Removed for double-blind review, and Removed for double-blind review. 2017. Video Annotation by Cascading Microtasks. In *Proceedings of Brazilian Symposium on Multimedia and the Web, Gramado, RS Brazil, October 2017 (WebMedia'2017),* 8 pages.
https://doi.org/XX.XXX/XXX_X

1 INTRODUCTION

Video is a very effective information container and it is a highly expressive type of media, capable of providing a large semantic load by presenting different audiovisual components coherently[25]. However, video can be considerably more useful when carrying metadata that can be used by video applications, and are often represented as video annotations.

Video annotation involves inserting tags on video objects to describe the content and the context of the videos, as well to describe characteristics of the media such as quality, coding among other propertie [37]. Annotations facilitate the manipulation of videos allowing the creation of content-based distribution applications [40], indexing [41], summarization [10], navigation [14], composition [38] among many others, both by automatic and manual means[36]. In other words, they are used to facilitate the work of users and systems that can handle annotated items.

In this paper, video annotations are differentiated into simple and complex ones, considering that simple annotations are those that can be acquired with a simple interaction of the workers in a microtask. In addition, a complex annotation is one that requires the worker execute a more exaustive and time-consuming task, in which he needs to perform multiple interactions.

For example, it is possible to annotate videos of football (soccer) matches. These annotations can describe relevant events on the game such as goals, cards and player substitutions [30]. In this paper its considering a simple annotation activity because is required a single annotation task to capture one of these events and the instant when it happened. A system that is able processing these annotations can use them to generate summary versions of the match's most important moments, to provide markers for positioning the video in each event, among other applications.

However, this example can be improved with additional annotations such as which player scores a goal or receives a card. The complexity can still increasing including annotations about how the goal was or why a player received a card. As video annotations become more complex, annotation tools need to be more elaborate and annotation tasks are more costly.

Automatic methods for video annotations present satisfactory efficiency and interesting results, however, these methods generally apply techniques that require well structured videos and extensive examples database, such as deep learning[21]. Unfortunately, many scenarios cannot provide these requirements, making it impossible to use automatic methods for video annotation [24]. In other way, manual video annotation are suitable for these scenarios because it uses human intelligence to handle the tasks. However, manual video annotation can be high-costly because of the potentially high-density of annotation points in the video, as well as the complex nature of some annotation tasks.

An alternative to achieve video annotation into a geneal scenario is to employ collaborative or cooperative approaches. In a collaborative approach the contributors work together to solve the main problem. Otherwise, in a cooperative approach each contributor attack a part of the main problem to produce a final result [22].

Taking cooperative approaches to a higher level, crowdsourcing video annotation has emerged as a proposal to annotate videos using a large number of contributors efficiently [34]. Following the crowdsourcing principles, the tasks distributed to the workers are modeled to be done in a independent way, maximizing the parallelism [17]. Moreover, each task can be sent to many contributors, making possible to check and to aggregate the contributions as well reducing the chance of produce a biased result [13].

A frequent problem of using a crowdsourcing approach to video annotation is to balance the relationship between task complexity and cost. Simple annotation tasks, such as clicking an object on a video, can be done in a few seconds for anyone, otherwise, more complex tasks such as providing complementary content and positioning it in the right position on a video, require some expertise of contributors and are more costly to them. In a crowdsourcing context, an ubiquitous designation for simple tasks that can be performed for any contributor quickly and easily is microtask [9].

Into this scenario, the approach introduced aims to provide ways to get around some issues faced in achieving video annotation.

- By using a manual annotation, no example bases or restricted conditions are required as in automatic methods.
- By using a crowdsourcing process does not require the work of experts nor trained workers.
- By using a microtask-based crowdsourcing process, it makes the contribution process simple and quick, avoiding time-consuming and tedious tasks to workers.
- By using microtasks in which only a simple annotation is collected does not require sophisticated annotation tools.

In short, this paper aims to present a general approach to crowdsourcing video annotation without requiring trained annotators nor experts. It is an alternative approach to achieving a complex video annotation by cascaded microtasks. Instead of developing a complex annotation tool for the desired complex video annotation, a very simple annotation tool is built for each microtask. The microtasks are applied following a process workflow, in which the output provided from a microtask feeds the next one. To illustrate how this approach works was conducted an experiment, and to support it was developed a toolkit that includes a set of simple Web-based annotation tools and aggregation methods. This toolkit is open source and allows to replicate the experiment and proceed with other video annotation processes.

The rest of this paper is structured as follows. Section 2 presents related works. Section 3 presents the approach introduced by this paper. Section 3 presents the conducted experiment. Finally, section 5 concludes the paper presenting final considerations and future prospects.

2 RELATED WORKS

Crowdsourcing annotation systems are used in various applications and are used to gather information of various types, such as temporal synchronization[6, 39], events[20, 33], objects[1, 27], emotions[3, 29], actions[7, 28], quality[11, 16], geo-tagging[5, 15], social relevance[2, 18, 31] and captions[8, 19].

However, some of these works are based on complex annotation tools, demands tedious or time-consuming tasks, or require trained and skilled workers. Some relevant examples that should be regarded include works such as [1, 3, 7, 15, 18, 26, 35].

VidWiki[1] is a complex system to improve video lesson by video annotation, which provides a complex annotation tool(Figure 1) that allows the worker to edit video scenes by entering various types of notes, including LaTex equations. Another interesting paper to note was preceded in 2012 by C.Vandrick[35] in which time-consuming complex tasks were deployed in the Amazon Mechanical Turk[15] and require specialized work to perform them.

While these works often produce interesting results, this approach restricts potential workers and owners capable of developing complex tools and hiring skilled workers.

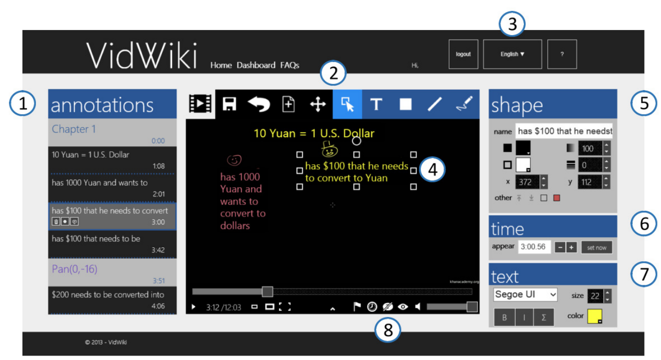

Figure 1: VidWiki annotation tool[1]

On the other hand, there are also papers on crowdsourcing video annotation that report the use of simple tools and microtasks that can be done quickly by unskilled workers. These works include [4, 6, 12, 20, 27, 28, 33, 39].

The work done by N.Gagil in 2014[12] has a very simple annotation tool(Figure 2) that allows the workers to execute an easy microtask, which consists of annotate videos with surveillance problems if any of them are found.

ReTool[4] is a work that must be mentioned because, because it presents a web-based tool for owners to create and publish annotation microtasks and workflows to execute them.

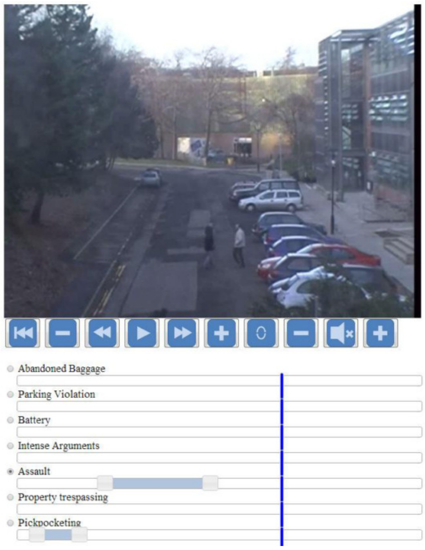

Figure 2: Simple surveillance annotation tool[12]

ToolScape[20] is a work that deserves prominence, as it is strongly related to this. In this work, they provided simple annotation tools (Figure 3) to workers execute a sequence of three micro-tasks, that was used to extract the step-by-step structure of the instruction videos. As well in presented a design pattern to define the workflow of these tasks. Among the works analyzed, TollScape is the closest to it.

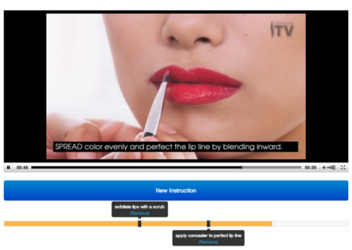

Figure 3: ToolScape annotation tool[20]

The study of the related works was enlightening, mainly to understand the characteristics of the micro-tasks, as well as the simple annotation tools used to execute them.

3 THE APPROACH INTRODUCED

The crowdsourcing video annotation approach presented in this paper follows three steps: Preparation, Annotation, and Presentation. These steps contains specific activities, and are executed sequentially how can be seen in Figure 4.

Preparation: all activities involved in this step are performed by the owner, who started the video annotation process. At this step is determined what must be annotated, as well how they should be annotated. In this way the owner must determine:

(1) What kinds of point of interest should be annotated. Ex: events, objects, subjects, issues.

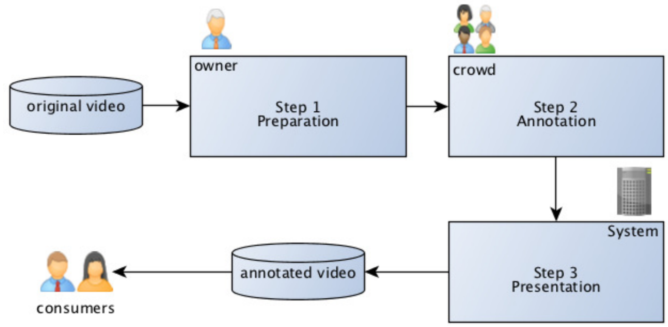

Figure 4: Process workflow

(2) What annotation type will be used for each of these kinds. Ex: free write, item select, button click, image upload.
(3) What data type will be collected for each annotation type. Ex: plain text, location, image, video.

To illustrate this, the example of the football(soccer) match annotation will be recalled. In a football(soccer) match video the kinds of point of interest correspond to events such as goals, cards and substitutions of players. For each point of interest observed it should be collected its kind as well as the instant each event happened. The annotation type to be used on the annotation tool can be a set of icons related to each event. Finally, the data type collected in this case may be plain text that contains the kind of event identified and the instant it happened [30].

Also it is important to provide explanations or guidelines that can instruct the workers about how execute the microtasks. An additional activity on the Preparation step is determine what section of the video should be send by each worker, this division can be made by duration (ex: to send a 5 seconds segment to each worker), or user contextual criteria such as to send to each user a segment that contains a single dialog. The activities sequence for this step can be observed on Figure 5.

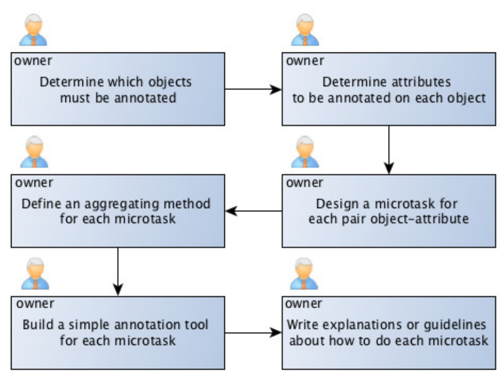

Figure 5: Preparation step

Annotation: an essential aspect for this step is to determine the microtasks' workflow, so the output from a task is taken as input by the next one, generating a final outcome at the end of the last microtask. This cascade workflow is illustrated in Figure 6.

It is important to notice that each task cell in composed by two activities, the microtask in self, and the aggregation method that generates the output from the obtained contributions. In this way, the output from the last task cell is the final outcome provided by the system.

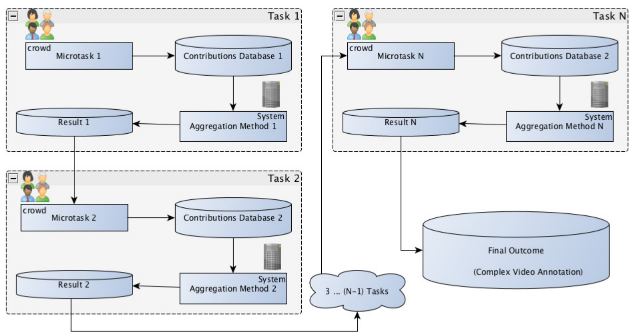

Figure 6: Annotation step for N microtasks

Presentation: at this step is generated a annotated video including the original video and the final outcome from the previous step. Other activities that can be proceeded at this step is to generate, or to render, media items selected from the crowd annotations, as well aggregate these items over the videos to compose a multimedia presentation.

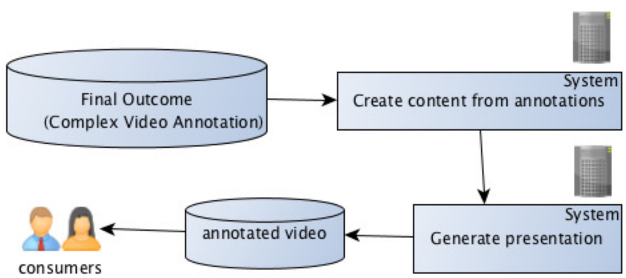

Figure 7: Presentation step

4 EXPERIMENT

In order to demonstrate how the proposed approach can be applied in a crowdsourcing video annotation process, an experiment was conducted in which four different microtasks were cascaded to obtain video enrichment by adding extra content such as images, text, Hyperlinks and other videos. To provide computational support to the experiment was developed a toolkit that includes all annotation tools and aggregation methods required to perform the microtasks, as well the presentation system capable to play the generated annotated videos.

The presentation system can handle the annotated video to play coherently the video and the extra content produced and positioned by the crowd, according to the annotations provided by them.

All the tasks performed by the workers were designed to be simple, easy and small, and through them everything was done by the crowd: interesting points identification, content providing, content ranking, even to determine the position at which each content should be displayed on video scenes.

The crowd involved in this experiment was recruited by an open call that consisted of a broadcast message posted on a social network, asking volunteers to access an annotation tool and use it to contribute to an academic work. Of course, the reach of the open call would be much greater, as well as the number of contributions, if a commercial system such as Amazon Mechanical Turk or CrowdFlower was used, but it would require financial resources that were not available [9]. However, the open call worked and the number of contributions received was sufficient to generate the desired outcome.

Each microtask was active for 24 hours, then it was closed and the specific aggregation method was executed. After the aggregation method produced the result for a microtask, the next one became active, taking that result as input. To take advantage of the first open calls, the strategy was to use the same URL for all microtasks, redirecting the contributor to the microtasking that was currently active.

The decision to leave each microtesk active for 24 hours had two reasons: to share the time available for the experiment among all microtasks in a balanced way and to measure how many contributions can be acquired by a subtle open call in a single day.

In order to reduce duplicate contributions without the need to register contributors, it was collected a fingerprint using the IP address and Web browser ID for each contribution [23, 32]. This technique provided a good indication of who was contributing, so it was possible to discard multiple contributions from the same brand for the same item to be annotated. As tasks 1 and 2 aimed to collect as many items as possible from the crowd, this technique was applied only in tasks 3 and 4, in which workers contributed based on the items already registered in the system. However, even in tasks 3 and 4 the system accepted annotations with the same fingerprint for different items.

In this experiment, two annotated videos with approximately 1 minute duration were produced using the cascade process. It is next described how the proposed approach was applied in this experiment, as well as the annotation tools and the aggregation methods used.

4.1 Step 1: Preparation

Before beginning this step, it is important to determine which annotations should be generated at the end of the process. In this case, the annotated video must be related to items displayed at certain positions in specific scenes. This way, the following activities must be specified:

- **Determine which objects must be annotated:** The items to be observed in these experiments are called Points of Interest. Each point of interest is related to one of these types of occurrences.
  (1) Expression that requires synonyms or definitions;
  (2) Information that requires additional explanation.
- **Determine attributes to be annotated on each object:** Each point of interest must be annotated in two aspects.
  (1) Which content should be displayed over it;
  (2) In what position the content should be displayed.

- **Describe a microtask for each pair object-attribute:** To achieve the desired annotated video, a microtask is needed to identify the points of interest, two microtasks to collect their attributes and a microtask to ranking the suggested content to be associated with each point of interest.
  - **Task 1: Identify the points of interest.** Contributors should watch video segments and mark a point of interest if found, as well identify its subject using plain text. This text can be a word or a phrase according the point of interest.
  - **Task 2: Suggesting contributions content.** Contributors receive a random point of interest and suggest some content to cover it. The provided content may be an image, a hyperlink to a website, or plain text.
  - **Task 3: Ranking the provided contents.** Contributors receive a random suggested content for a point of interest and vote on the most appropriate.
  - **Task 4: Positioning the items over a scene.** Contributors receive a a random scene and should position the item in the position over the scene that reduces occlusion problems. In this task, each position is captured as a coordinate pair [X,Y] considering the superior-left corner of the video as the coordinate [0,0].
- **Define an aggregating method for each microtask:** According to the approach followed, it is necessary to define an appropriate aggregation method for each microtask.
  - **Task 1: Temporal grouping.** Identified points of interest are grouped by time, with a tolerance of 0.5 seconds. For each group, a content analysis is performed to merge equivalent contributions. Finally, the predominant input is selected and marked as the point of interest at that time in relation to the timeline.
  - **Task 2: Grouping by point of interest.** The content provided by contributors in task 2 is grouped by point of interest. Therefore, a content analysis is done to bring together equivalent contributions.
  - **Task 3: Ranking by voting.** For each point of interest, the most popular suggestion is selected based on contributions to task 3.
  - **Task 4: Average coordinates.** The contributions are grouped by point of interest and, for each point, the average coordinate is determined.
- **Build a simple annotation tool for each microtask:** Still following the presented approach, a simple annotation tool must be built for each microtask. Each of these tools is designed to collect a specific annotation.
  - **Task 1:** The first annotation tool (Figure 8) consists of a video player, with a navigation bar that allows the collaborator to watch the video and pause it at the moment the point of interest is found. Thus, the worker can write the subject for this and send the contribution.
  - **Task 2:** the second annotation tool (Figure 9) presents the collaborator a point of interest and positions the video at the moment it occurs, for use of context and reference. If the interesting point is a word or expression, the worker can write a definition, a synonym or upload an image that illustrates it. If the point of interest is a fact or information that needs to be explained, the contributor can write a textual explanation, upload an image that explains it or even provide a link to a website with information about it.

Figure 8: Annotation tool for task 1

Figure 9: Annotation tool for task 2

  - **Task 3:** The third annotation tool (Figure 10) presents to the collaborator with a point of interest, as well as the different contents suggested to cover it. The contributor should navigate through the suggested content using the button bar on the contribution form and choose which one is most appropriate.
  - **Task 4:** The fourth annotation tool (Figure 11) shows the contributor the content chosen in task 3 to cover a point of interest, and asks to choose the best position in the scene to display by clicking on the desired position.
- **Write explanations about the microtask:** The last activity of the preparation step in to write explanations about how to execute each annotation microtask. These explanations are very important because, in a

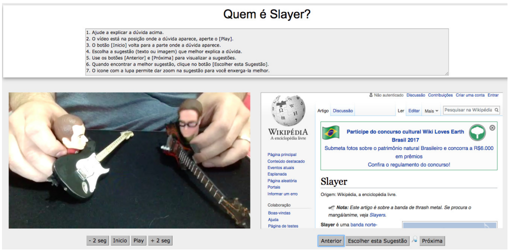

Figure 10: Annotation tool for task 3

Figure 11: Annotation tool for task 4

crowdsourcing scenario, usually is not possible to instruct personally the workers about how to contribute.

For each annotation tool developed to this project, was written a sequence of instructions that explain how to use it. In all these tools, the instructions are presented on the top of the page, as can be seen on Figures 8, 9, 10 and 11.

4.2 Step 2: Annotation

- **Task 1 - Identify Points of Interest:** the first annotation microtask was supported by the tool represented in Figure 8, collecting identification for points of interest. In this task the contributor receive a segment of video that should be watched, and if was found any word, expression or information that require additional definitions or explanations, it should be marked as a point of interest.

  The segment of video sent to each user is chosen randomly according the criteria defined by the user. Two strategies was available to determine the segments: time length and semantic blocks. For this experiment was used semantic blocks, it means that for each video was previously defined start and stop times for segments of the video that should be presented as a semantic cell. This strategy demands an initial effort from the owner, but deliver to worker contextualized video segments.

  Was collected 68 contributions and, after close the task and execute the aggregation method, they was merged into 37 different points of interest.
- **Task 2 - Suggesting Contribution Content:** the second task taken as input the aggregated result from the task 1, with 37 points of interest. In order to take advantage of the open call made for the task 1, was used the same URL redirecting the workers to the second annotation tool.

  The open call was reinforced by sharing it on a social network, and it resulted in 308 contributions in 24 hours. After execute the second aggregation method, the 308 contributions was merged into 239 suggestions of content to cover the points of interest. These suggestions included plain text, images and hyperlinks.
- **Task 3 - Ranking Suggestions:** the third microtask aimed ranking the 239 suggested contents that resulted from the task 2. Was repeated the strategy of use the same URL for the new task, and reinforce the same open call.

  In this task was noticed a issue involving the suggestions associated to hyperlinks. Some of contributors related problems to visualize these suggestions such as "PAGE NOT FOUND" or "BLOCKED WEBSITE". Maybe because of this, most hyperlinks voted as most appropriated content point to Wikipedia or Youtube.

  The number of contributions collected in 24 hours was 255, and them were enough to determine the most appropriated content to all 37 points of interest.
- **Task 4 - Positioning Items:** by foreign affairs the fourth annotation microtask was performed about one week after the previous one. Because of this was opted for recruit contributors by a new open call. However, this open call was similar to the first one.

  This microtask was the simplest task, and could be done in a few seconds. In 24 hours were collected 541 contributions that consisted in suggestions about in which position each item should be positioned over a video scene. These contributions were enough to determine the average position for all 37 points of interesting.

4.3 Step 3: Presentation

With all annotations generated, the presentation system was able to use them to compose a presentation using both the original video and annotations.

- **Generating Outcome:** The annotations that represent the content that must be displayed to cover the interesting points have been manipulated in different ways according to their media types. The images were scaled according to the presentation area, the hyperlink was used to load previews for web pages, and the texts were formatted to be displayed correctly.

  In addition, for each item, a second view was generated to be used in the zoom box, activated by the user to enlarge the items for a better visualization, as can be seen in the Figure 12.
- **Presentation:** The presentation system can be seen on the Figure 14. This system is capable of reproducing the

Figure 12: Extra content into the zoom box

original video synchronized with the extra content. When the user clicks on some extra content displayed in the video, the presentation is paused and a larger preview for the selected content is displayed in the zoom box as shown in the Figure 12.

This systems features navigation by extra-content instead the traditional timeline navigation, making available a button-bar with buttons to navigate among the extra contents.

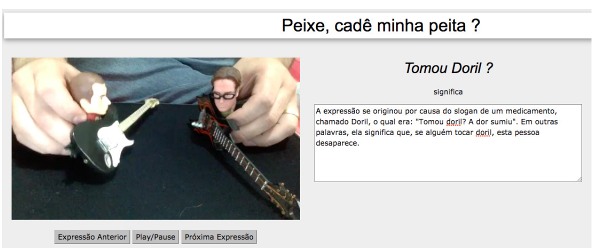

Figure 13: Previous presentation system

The first version of the presentation system (Figure 13) presented the extra content in a delimited area on the right side of the page. The reason for this is that Task 4 was not applied at the beginning, so the system did not have information on where to display the items on the video.

Fortunately, it was possible to perform the fourth microtask later and it was possible to create the new version of the presentation system (Figure 14). In addition, this issue demonstrates that it is possible to reuse or improve an annotation system using this approach simply by adding new microtasks.

5 CONCLUSION

This paper introduced a crowdsourcing approach to annotate videos without requires experts, trained workers nor time-consuming tasks. Moreover, was conducted an experiment to validating it by generating interesting annotated videos that could be used to create interactive multimedia presentations. To support this experiment

Figure 14: Presentation system

was developed a toolkit that includes the presentation system, and a set of video annotation tools and aggregation methods.

During the annotation stage, it was noticed that the faster microtasks received more contributions, because the workers contributed more times, annotating more items. One conclusion about this is that volunteers use to dedicate a set time to perform tasks, so they were willing to execute any number of microtasks during that interval.

Another observation about the approach is that the cascade of tasks results in the generation of partial results that can be used for other purposes. For example, content suggestions that have been collected to annotate the video can be used to populate an online dictionary or encyclopedia.

Moreover, the individual aggregation of the result of each microtask allows more adequate processing for each annotation, as well as specific validations for them.

Perhaps one of the most interesting results was to see if this approach is capable of generating systems that can be reused and expanded. This can be observed when the first presentation system was generated and later a new task was added in the process, allowing the construction of an improved presentation system.

In addition to the approach presented, which was able to guide crowdsourcing annotation processes with a certain degree of complexity, a system was also generated that demonstrates how this approach can be applied. This system is available for use and can be used both to replicate this experience and to perform other works.

5.1 Next steps

An immediate improvement in the system includes changes in the aggregation methods of tasks 1 and 2. Currently, the similarity comparison uses simple syntactic techniques for content analysis, however, a method is being developed that performs these comparisons through morphosyntactic analysis.

The owner module will also be developed, which will allow this system to be used even outside the academic environment.

Currently, the system counts only as microtask execution module, which was necessary to perform the experiment.

This work also served as a starting point for a series of projects that will be developed in the near future. In particular, the approach presented will be refined to become a complete method.

ACKNOWLEDGMENTS

Removed for double-blind review.

REFERENCES

[1] 2014. *VidWiki: Enabling the Crowd to Improve the Legibility of Online Educational Videos.* ACM Conference on Computer Supported Cooperative Work. https://www.microsoft.com/en-us/research/publication/vidwiki-enabling-the-crowd-to-improve-the-legibility-of-online-educational-videos/

[2] Marco Bertini, Alberto Del Bimbo, Andrea Ferracani, Francesco Gelli, Daniele Maddaluno, and Daniele Pezzatini. 2013. Socially-aware video recommendation using users' profiles and crowdsourced annotations. In *Proceedings of the 2nd international workshop on Socially-aware multimedia.* ACM, 13–18.

[3] Joan-Isaac Biel and Daniel Gatica-Perez. 2013. The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *Multimedia, IEEE Transactions on* 15, 1 (2013), 41–55.

[4] Chen Chen, Xiaojun Meng, Shengdong Zhao, and Morten Fjeld. 2017. ReTool: Interactive Microtask and Workflow Design Through Demonstration. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17).* ACM, New York, NY, USA, 3551–3556. https://doi.org/10.1145/3025453.3025969

[5] Si Chen, Muyuan Li, Kui Ren, and Chunming Qiao. 2015. Crowd map: Accurate reconstruction of indoor floor plans from crowdsourced sensor-rich videos. In *Distributed Computing Systems (ICDCS), 2015 IEEE 35th International Conference on.* IEEE, 1–10.

[6] Marcello N de Amorim, Ricardo MC Segundo, and Celso AS Santos. 2016. LiveSync: a Tool for Real Time Video Streaming Synchronization from Independent Sources. (2016), 112–115.

[7] Travis Desell, Kyle Goehner, Alicia Andes, Rebecca Eckroad, and Susan Ellis-Felege. 2015. On the effectiveness of crowd sourcing avian nesting video analysis at Wildlife@ Home. *Procedia Computer Science* 51 (2015), 384–393.

[8] Rucha Deshpande, Tayfun Tuna, Jaspal Subhlok, and Lecia Barker. 2014. A crowdsourcing caption editor for educational videos. In *Frontiers in Education Conference (FIE), 2014 IEEE.* IEEE, 1–8.

[9] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G. Ipeirotis, and Philippe Cudré-Mauroux. 2015. The Dynamics of Micro-Task Crowdsourcing: The Case of Amazon MTurk. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15).* ACM, New York, NY, USA, 238–247. https://doi.org/10.1145/2736277.2741685

[10] Guilherme Fião, Teresa Romão, Nuno Correia, Pedro Centieiro, and A. Eduardo Dias. 2016. Automatic Generation of Sport Video Highlights Based on Fan's Emotions and Content. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology (ACE2016).* ACM, New York, NY, USA, Article 29, 6 pages. https://doi.org/10.1145/3001773.3001802

[11] Bauke Freiburg, Jaap Kamps, and Cees GM Snoek. 2011. Crowdsourcing visual detectors for video search. In *Proceedings of the 19th ACM international conference on Multimedia.* ACM, 913–916.

[12] Neeraj J Gadgil, Khalid Tahboub, David Kirsh, and Edward J Delp. 2014. A web-based video annotation system for crowdsourcing surveillance videos. In *IS&T/SPIE Electronic Imaging.* International Society for Optics and Photonics, 90270A–90270A.

[13] FRANCIS GALTON. 1907. Vox Populi (The Wisdom of Crowds). *Nature* 75, 1949 (1907), 450–451. https://doi.org/10.1038/075509f0

[14] Dan B. Goldman, Chris Gonterman, Brian Curless, David Salesin, and Steven M. Seitz. 2008. Video Object Annotation, Navigation, and Composition. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology (UIST '08).* ACM, New York, NY, USA, 3–12. https://doi.org/10.1145/1449715.1449719

[15] Luke Gottlieb, Jaeyoung Choi, Pascal Kelm, Thomas Sikora, and Gerald Friedland. 2012. Pushing the limits of mechanical turk: qualifying the crowd for video geo-location. In *Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia.* ACM, 23–28.

[16] Chul-Hee Han and Jong-Seok Lee. 2014. Quality assessment of on-line videos using metadata. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on.* IEEE, 1385–1388.

[17] Jeff Howe. 2006. The Rise of Crowdsourcing. *Wired Magazine* 14, 6 (06 2006). http://www.wired.com/wired/archive/14.06/crowds.html

[18] Samuel Huron, Petra Isenberg, and Jean Daniel Fekete. 2013. PolemicTweet: Video Annotation and Analysis through Tagged Tweets. In *Human-Computer Interaction–INTERACT 2013.* Springer, 135–152.

[19] Hernisa Kacorri, Kaoru Shinkawa, and Shin Saito. 2014. Introducing game elements in crowdsourced video captioning by non-experts. In *Proceedings of the 11th Web for All Conference.* ACM, 29.

[20] Gunhee Kim, Leonid Sigal, and Eric P. Xing. 2014. Joint Summarization of Large-Scale Collections of Web Images and Videos for Storyline Reconstruction. In *Proceedings of the 2014 IEEE CCVPR (CVPR '14).* IEEE Computer Society, Washington, DC, USA, 4225–4232. https://doi.org/10.1109/CVPR.2014.538

[21] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.

[22] Melanie Misanchuk and Tiffany Anderson. 2001. Building Community in an Online Learning Environment: Communication, Cooperation and Collaboration. (2001).

[23] Martin Mulazzani, Philipp Reschl, Markus Huber, Manuel Leithner, Sebastian Schrittwieser, Edgar Weippl, and FC Wien. 2013. Fast and reliable browser identification with javascript engine fingerprinting. In *Web 2.0 Workshop on Security and Privacy (W2SP)*, Vol. 5.

[24] Venkatesh N Murthy, Subhransu Maji, and R Manmatha. 2015. Automatic image annotation using deep learning representations. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval.* ACM, 603–606.

[25] Marcello Novaes, Celso Alberto Saibel Santos, and Orivaldo Tavares. 2016. Ex-CAM - Uma metodologia Crowsourcing para a autoria de conteudo extra para videos. In *WebMedia 2016 WTD ().* Teresina - PI, Brazil.

[26] Sunghyun Park, Philippa Shoemark, and Louis-Philippe Morency. 2014. Toward crowdsourcing micro-level behavior annotations: the challenges of interface, training, and generalization. In *Proceedings of the 19th international conference on Intelligent User Interfaces.* ACM, 37–46.

[27] José Pedro Pinto and Paula Viana. 2013. TAG4VD: a game for collaborative video annotation. In *Proceedings of the 2013 ACM international workshop on Immersive media experiences.* ACM, 25–28.

[28] Laurel D Riek, Maria F OâĂŹconnor, and Peter Robinson. 2011. Guess what? a game for affective annotation of video using crowd sourcing. In *Affective computing and intelligent interaction.* Springer, 277–285.

[29] Dairazalia Sanchez-Cortes, Shiro Kumano, Kazuhiro Otsuka, and Daniel Gatica-Perez. 2015. In the Mood for Vlog: Multimodal Inference in Conversational Social Video. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 5, 2 (2015), 9.

[30] CAS Santos, Alexandre SANTOS, and TA Tavares. 2007. Uma estratégia para a construção de ambientes para a descrição semântica de vídeos. (2007).

[31] Elizeu Santos-Neto, Tatiana Pontes, Jussara M Almeida, and Matei Ripeanu. 2014. Towards Boosting Video Popularity via Tag Selection.. In *SoMuS@ ICMR.* Citeseer.

[32] Ricardo Segundo, Marcello N de Amorim, and Celso AS Santos. 2016. Crowdsourcing & Multimedia: Enhancing Multimedia Activities with the Power of Crowds. In *Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web.* ACM, 11–12.

[33] Fabio Sulser, Ivan Giangreco, and Heiko Schuldt. 2014. Crowd-based semantic event detection and video annotation for sports videos. In *Proceedings of the 2014 International ACM Workshop on Crowdsourcing for Multimedia.* ACM, 63–68.

[34] Luis Von Ahn. 2005. *Human Computation.* Ph.D. Dissertation. Pittsburgh, PA, USA. Advisor(s) Blum, Manuel. AAI3205378.

[35] Carl Vondrick, Donald Patterson, and Deva Ramanan. 2013. Efficiently Scaling Up Crowdsourced Video Annotation. *Int. J. Comput. Vision* 101, 1 (Jan. 2013), 184–204. https://doi.org/10.1007/s11263-012-0564-1

[36] Meng Wang and Xian-Sheng Hua. 2011. Active Learning in Multimedia Annotation and Retrieval: A Survey. *ACM Trans. Intell. Syst. Technol.* 2, 2, Article 10 (Feb. 2011), 21 pages. https://doi.org/10.1145/1899412.1899414

[37] Meng Wang, Xian-Sheng Hua, Jinhui Tang, and Richang Hong. 2009. Beyond Distance Measurement: Constructing Neighborhood Similarity for Video Annotation. *Trans. Multi.* 11, 3 (April 2009), 465–476. https://doi.org/10.1109/TMM.2009.2012919

[38] Stefan Wilk, Stephan Kopf, and Wolfgang Effelsberg. 2015. Video Composition by the Crowd: A System to Compose User-generated Videos in Near Real-time. In *Proceedings of the 6th ACM MSC (MMSys '15).* ACM, New York, NY, USA, 13–24. https://doi.org/10.1145/2713168.2713178

[39] Bin Wu, Erheng Zhong, Ben Tan, Andrew Horner, and Qiang Yang. 2014. Crowdsourced time-sync video tagging using temporal and personalized topic modeling. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 721–730.

[40] Jun Zhang, Xiaoming Fan, Jianyong Wang, and Lizhu Zhou. 2012. Keyword-propagation-based Information Enriching and Noise Removal for Web News Videos. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12).* ACM, New York, NY, USA, 561–569. https://doi.org/10.1145/2339530.2339620

[41] Yifan Zhang, Xiaoyu Zhang, Changsheng Xu, and Hanqing Lu. 2007. Personalized Retrieval of Sports Video. In *Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval (MIR '07).* ACM, New York, NY, USA, 313–322. https://doi.org/10.1145/1290082.1290126