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 an alternative approach to collaborative video annotation, which consists of to cascade very small and simple annotation microtasks[1] , rather than defining a single large task for contributors. This approach also grants smaller granularity in verifying and aggregation of contributions and allows workers[2] to accomplish their tasks in a few seconds with extremely low effort. Efficiency is increased by using crowdsourcing[3] techniques to quickly acquire a large number of contributions as well as to validate and converge resulting in satisfactory outcomes. To demonstrate that it is possible to achieve complex results by cascading simple microtasks, was proceeded an video enrichment experiment, adding extra content[4] on videos. All content insertion point identifications, content providing and positioning of extra content was done entirely by the contributors, generating interactive multimedia videos without the need for specialized professionals or heavy tasks. A WEB based system was developed to provide computational support for the experiment. This system allows workers to perform all microtasks, as well allows the owner[5] to apply the convergence algorithms for each task. This system is open source and allows to replicate the experiment and proceed with other video annotation experiences.

CCS CONCEPTS

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

KEYWORDS

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

[1]Really small and simple tasks that can be executed for ordinary people.
[2]Contributors are also appointed as workers because they perform tasks.
[3]Crowdsourcing is a cooperative approach that can handle massive scale contributions.
[4]Extra content such as images, texts, other videos and hyperlinks for websites.
[5]An owner is who initiate a crowdsourcing project.

WebMedia'2017, October 2017, Gramado, RS Brazil

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

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),* 7 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[20]. However, video can be considerably more useful when carrying metadata that can be used by video applications. This metadata is often represented as video annotations that can be generated manually or automatically [16, 21, 23]. These annotations are object-related metadata that must be described in relation to certain aspects, providing understanding about artifacts. Annotations are the basis for semantics web, as well as for search engines and information retrieval systems [2, 3, 6, 13].

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 [15, 18, 26]. Annotations facilitate the manipulation of videos allowing the creation of content-based distribution applications [29], indexing [7, 30], summarization [8, 12], navigation [10], composition [27, 28] among many others, both by automatic and manual means[25]. In short, they are used to facilitate the work of users and systems that can handle annotated items.

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. Unfortunately, many scenarios cannot provide these requirements, making it impossible to use automatic methods for video annotation. 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 [17].

Taking cooperative approaches to the next level, crowdsourcing video annotation has emerged as a proposal to annotate videos using a huge number of contributors efficiently [1, 24]. Following the

crowdsourcing principles, the tasks distributed to the contributors are modeled to be done in a independent way, maximizing the parallelism [11]. 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 [9].

A frequent problem of using a crowdsourcing approach to video annotation is to balance the relationship between 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 [5].

This paper presents an alternative approach for achieve complex video annotation by cascading microtasks. Instead design a complex annotation tool for the desired complex video annotation, are designed a very simple annotation tool for each microtask. The microtasks are applied following a process workflow, in which the output provided from a microtask is the taken as input for the next one. To demonstrate how this approach works was conducted an experiement, and developed a WEB based system to support it. This system is open source and allows to replicate the experiment and proceed with other video annotation experiences.

The rest of this paper is structured as follows. Section 2 presents the approach proposed in this paper. Section 3 presents the conducted experiment. Section 4 presents related works. Finally, section 5 concludes the paper presenting final considerations and future prospects.

2 CASCADING THE MICROTASKS

The approach proposed in this paper consists in a three-steps workflow: Preparation, Annotation, and Presentation. These steps contains specific activities, and are executed sequentially how can be seen in Figure 1.

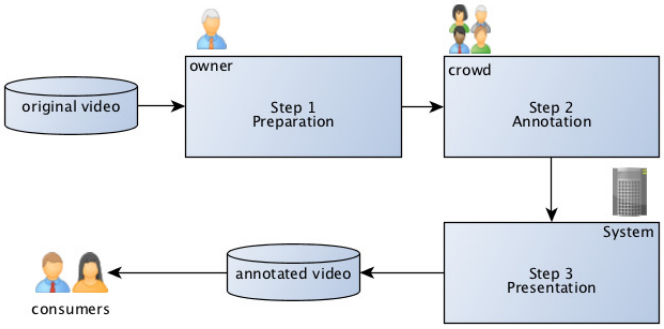

Figure 1: Process workflow

Preparation: all activities executed on this step are performed by the owner. At this step is determined the content and characteristics of the video annotation desired as an outcome. So, it is identified both annotated objects and different attributes annotated about each object. For each pair object-attribute is defined an annotation microtask to provide only this specific annotation. Each of these microtasks are supported by a very simple annotation system that receive several contributions from the crowd. As well, each microtask is related to an aggregation method, that process its contributions database, and deliver a result according the characteristics of the annotations collected by each task.

Moreover, the owner must determine for each task what interesting points should be annotated, and what data types can be provided as annotation (ex: text, image, video, hyperlink, location). As well it is important to provide explanations or guidelines that can instruct the workers about how execute the microtask.

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 ectivities sequence for this step can be observed on Figure 2.

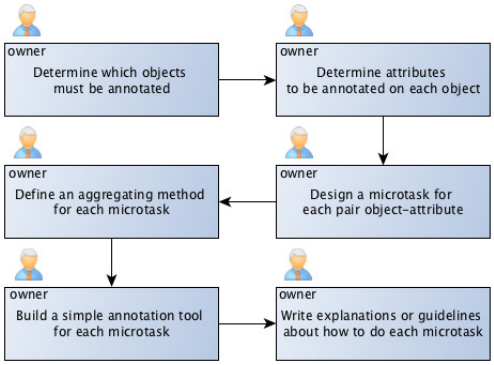

Figure 2: Preparation step

Annotation: an essential aspect for this approach is to determine the correct sequence for the microtasks, 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 cascading is illustrated in Figure 3. 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.

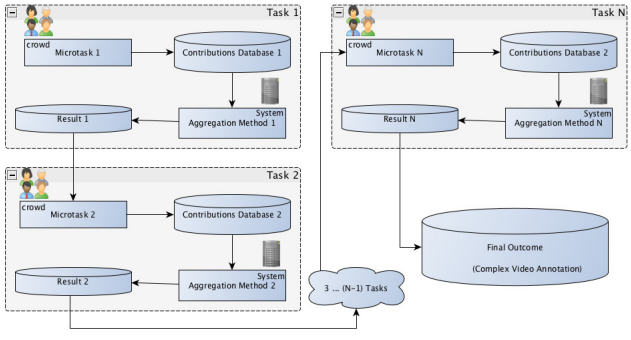

Figure 3: Annotation step for N microtasks

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 from the annotations, as well aggregate these items over the videos to compose a multimedia presentation.

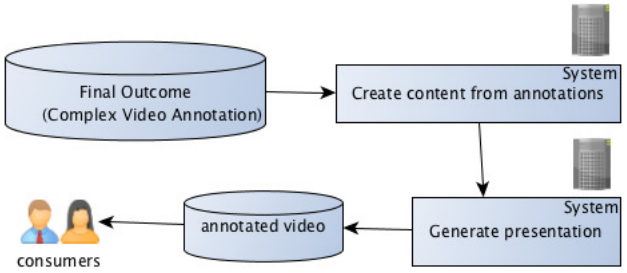

Figure 4: Presentation step

3 EXPERIMENT

To demonstrate how the proposed approach can be applied to the creation of video annotation systems, a crowdsourcing experiment was conducted in which four different micro-tasks were cascaded to obtain video enrichment by adding extra content such as images, texts, hyperlinks and other videos.

In this way, a WEB-based system was developed that provides a simple annotation tool for each microtask, as well as a specific aggregation method to process the result of each one. This system takes regular videos as input and delivers multimedia presentations as results. These presentations include the original videos and the extra content produced and positioned by the crowd, according to the notes provided by them.

All the tasks performed by the employees 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 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 [5]. However, the open call worked and the number of contributions received was sufficient to generate the desired outcome.

Each of the microtasks 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. 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. However, the system accepted notes with the same fingerprint for different items.

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

3.1 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.

- **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.
- **Design a microtask for each pair object-attribute:** To reach 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 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 provided contents.** Contributors receive a random suggested content for a point of interest and vote on the most appropriate.
  - **Task 4: Positioning 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 5) 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.

Figure 5: Task 1 - Identify points of interest

  - **Task 2:** the second annotation tool (Figure 6) 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 6: Task 2 - Suggesting contributions content

  - **Task 3:** The third annotation tool (Figure 7) presents 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. In addition, the collaborator can use the zoom button to better visualize the contents.

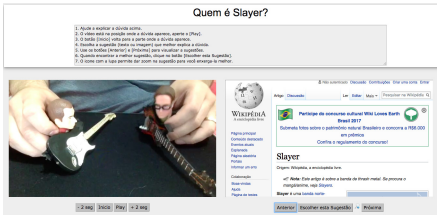

Figure 7: Task 3 - Ranking provided contents

(Figure 8)

  - **Task 4:** The fourth annotation tool 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 it. The worker can choose the position simply by clicking the video in the desired position.

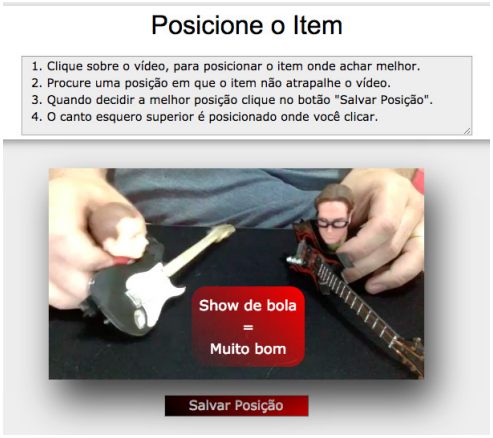

Figure 8: Task 4 - Positioning Items over a scene

- **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 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 5, 6, 7 and 8.

3.2 Annotation

- **Task 1 - Identify Points of Interest:** the first annotation microtask was supported by the tool represented in Figure 5, 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.

3.3 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 9.

Figure 9: Zoom box on the presentation system

- **Presentation:** The presentation system can be seen on the Figure 10. This system is capable of reproducing the 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 9.

  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 10: Presentation system

The first version of the presentation system (Figure 11) 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 10). In addition, this issue demonstrates that it is possible to reuse or improve an annotation system using this approach simply by adding new microtasks.

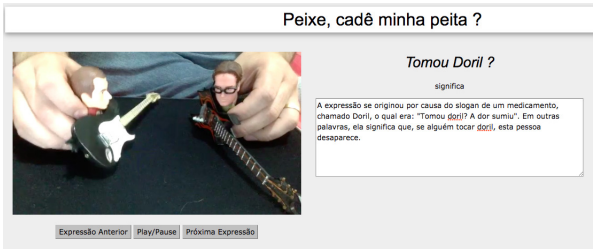

Figure 11: Previous presentation system

4 RELATED WORKS

In this section is presented a set of works that adopted crowdsourcing video annotation to achieve their results. For each of them is described how could be applied the proposed approach to abstain similar results.

[4] [14] [28] [22] [19]

TO DO

5 CONCLUSION

This paper presented a crowdsourcing approach to achieving a complex video annotation by cascading annotation microtasks. Moreover, an experiment was driven by this approach, validating it by generating interesting and annotated videos that could be used to create interactive multimedia presentations.

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 to the system allowing the construction of an improved presentation system.

In addition to the approach presented, which was able to guide the construction of a crowdsourcing annotation system 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 will be the change in the aggregation function 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 the morphosyntactic analysis.

The owner's 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 carry out 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 method.

ACKNOWLEDGMENTS

Removed for double-blind review.

REFERENCES

[1] Luis von Ahn. 2011. Massive-scale online collaboration. (Apr 2011). http://www.ted.com/talks/luis_von_ahn_massive_scale_online_collaboration

[2] Philipp Cimiano, Siegfried Handschuh, and Steffen Staab. 2004. Towards the Self-annotating Web. In *Proceedings of the 13th International Conference on World Wide Web (WWW '04)*. ACM, New York, NY, USA, 462–471. https://doi.org/10.1145/988672.988735

[3] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. 2008. Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Comput. Surv.* 40, 2, Article 5 (May 2008), 60 pages. https://doi.org/10.1145/1348246.1348248

[4] R. Di Salvo, D. Giordano, and I. Kavasidis. 2013. A Crowdsourcing Approach to Support Video Annotation. In *Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications (VIGTA '13)*. ACM, New York, NY, USA, Article 8, 6 pages. https://doi.org/10.1145/2501105.2501113

[5] 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

[6] Pavel A. Dmitriev, Nadav Eiron, Marcus Fontoura, and Eugene Shekita. 2006. Using Annotations in Enterprise Search. In *Proceedings of the 15th International Conference on World Wide Web (WWW '06)*. ACM, New York, NY, USA, 811–817. https://doi.org/10.1145/1135777.1135900

[7] Emna Fendri, Siwar Jallouli, and Hanene Ben-Abdallah. 2010. SVE: An Integrated System for Soccer Video Edition. In *Proceedings of the International Conference on Multimedia Information Retrieval (MIR '10)*. ACM, New York, NY, USA, 381–384. https://doi.org/10.1145/1743384.1743446

[8] 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

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

[10] 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

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

[12] Katsunori Ishino, Ryohei Orihara, Hiroyuki Nakagawa, Yasuyuki Tahara, and Akihiko Ohsuga. 2012. Support for Video Hosting Service Users Using Folksonomy and Social Annotation. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01 (WI-IAT '12)*. IEEE Computer Society, Washington, DC, USA, 472–479. http://dl.acm.org/citation.cfm?id=2457524.2457633

[13] João Rocha Junior, Ícaro Cavalcante Dourado, and Osvaldo Matos Júnior. 2008. MuSE: Multimedia Search Engine. In *Companion Proceedings of the XIV Brazilian Symposium on Multimedia and the Web (WebMedia '08)*. ACM, New York, NY, USA, 113–116. https://doi.org/10.1145/1809980.1810011

[14] 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

[15] J. Li and J. Z. Wang. 2008. Real-Time Computerized Annotation of Pictures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 6 (June 2008), 985–1002. https://doi.org/10.1109/TPAMI.2007.70847

[16] Rada Mihalcea and Andras Csomai. 2007. Wikify!: Linking Documents to Encyclopedic Knowledge. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management (CIKM '07)*. ACM, New York, NY, USA, 233–242. https://doi.org/10.1145/1321440.1321475

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

[18] Milind R. Naphade and John R. Smith. 2004. On the Detection of Semantic Concepts at TRECVID. In *Proceedings of the 12th Annual ACM International Conference on Multimedia (MULTIMEDIA '04)*. ACM, New York, NY, USA, 660–667. https://doi.org/10.1145/1027527.1027680

[19] Michael Nebeling, Maximilian Speicher, and Moira C. Norrie. 2013. CrowdStudy: General Toolkit for Crowdsourced Evaluation of Web Interfaces. In *Proceedings of the 5th ACM SIGCHI SEICS (EICS '13)*. ACM, New York, NY, USA, 255–264. https://doi.org/10.1145/2494603.2480303

[20] 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.

[21] Eyal Oren, Knud Möller, Simon Scerri, Siegfried Handschuh, and Michael Sintek. 2006. What are semantic annotations. *Relatório técnico. DERI Galway* 9 (2006), 62.

[22] Judith Alice Redi, Tobias Hossfeld, Pavel Korshunov, Filippo Mazza, Isabel Povoa, and Christian Keimel. 2013. Crowdsourcing-based Multimedia Subjective Evaluations: A Case Study on Image Recognizability and Aesthetic Appeal. In *Proceedings of the 2Nd ACM International Workshop on Crowdsourcing for Multimedia (CrowdMM '13)*. ACM, New York, NY, USA, 29–34. https://doi.org/10.1145/2506364.2506368

[23] Ayush Singhal and Jaideep Srivastava. 2014. Generating Semantic Annotations For Research Datasets. In *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14) (WIMS '14)*. ACM, New York, NY, USA, Article 30, 11 pages. https://doi.org/10.1145/2611040.2611056

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

[25] 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

[26] 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

[27] Stefan Wilk and Wolfgang Effelsberg. 2014. Mobile Video Broadcasting Services: Combining Video Composition and Network Efficient Transmission. In *Proceedings of the 22Nd ACM International Conference on Multimedia (MM '14)*. ACM, New York, NY, USA, 643–646. https://doi.org/10.1145/2647868.2654857

[28] 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

[29] 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

[30] 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