jet INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2024, VOL. 70, NO. 4, PP. 1063–1069
Manuscript received March 13, 2024; revised October 2024. doi: 10.24425/ijet.2024.152361

Crowdsourcing Evaluation of Video Summarization Algorithm

Avrajyoti Dutta, Dawid Juszka, and Mikołaj Leszczuk

***Abstract*—The technique of video summarizing involves selecting the most relevant and informative sections of a video to generate its shortened and faster version. Crowdsourcing is a relatively new term that has been exploited in the present study to achieve video summarization. This technique helps in dividing a task into multiple parts, and each of these parts is then evaluated using a large group of individuals to solve problems that are otherwise difficult to solve using traditional computational machines. In this study, we offer a crowdsourcing subjective experiment in which summaries of processed video sequences are evaluated. Thus, we are proposing an experiment that utilizes crowdsourcing to evaluate the efficacy of an algorithm that summarizes videos. A group of 45 individuals participated in the experiment, where each of them were asked to watch 24 videos, each of 30-second and 45-second duration. An experimental comparison was conducted with respect to presentation order and random selection methods. A content-based video segmentation was also used to represent different levels of complexities and visual richness. The findings of the assessment showed that specific characteristics of a video such as its length, complexity, and content, play a major role in improving the performance of the summarization algorithm. This study is an essential step toward the development of video summarizing systems that are both more accurate and more efficient.**

***Keywords*—Crowdsourcing Evaluation; Subjective Experiment; Video Quality Assessment; Video Summarization; Processed Video Sequences (PVS); User Generated Content (UGC); Quality of experience (QoE); Video Quality Indicators (VQI)**

I. Introduction

THE dawn of the digital age has resulted in a proliferation of video material, with millions of hours of video being posted on the internet every single day. This has led to an increase in the amount of available online viewing time. Because of the large amount of information available, it has become increasingly difficult for consumers to locate and digest content that is pertinent to their needs. Because the viewpoint of the end user has a direct impact on the Quality of Experience (QoE) of multimedia programs [1], conducting research on the end user's subjective experience is necessary to establish the degree to which the user is satisfied or unsatisfied with the program [2]. Because of its inherently subjective character and its dependency on a wide range of parameters, assessing the effectiveness of video summarizing algorithms can be a challenging and time-consuming endeavor. These criteria include, but are not limited to, the context, the purpose, and the intended viewership. The process of constructing a summary of a video clip by selecting the most relevant sections of the video to include in the summary is referred to as "video summarization" [3]. It is a challenging task that calls for a significant amount of effort and specialized knowledge.

In recent times, the idea of using crowdsourcing as a potential solution to this problem has been floated. The goal is to create summaries in a short amount of time, which requires many individuals to work on the assignment simultaneously.

The process of video summary has developed into an important part of contemporary multimedia technology. The work of summarizing short video clips can be tough and time-consuming due to the amount of information that must be covered. Because of this, there is an increasing need for the summary of material found in videos. The process of video summary is a difficult one that has historically been tackled from a variety of angles. In recent years, several different strategies for summarizing video content have been put up as potential solutions. The ever-increasing quantity of video content that can be accessed over the internet has resulted in an increase in the significance of video summarization. It has been demonstrated that crowdsourcing [4] is a beneficial method for the purpose of video summary since it enables the collection of a wide variety of opinions and points of view.

In this study, we offer an experiment in crowdsourcing for the purpose of summarizing videos according to scenarios. The purpose of the experiment is to determine whether crowdsourcing is a successful method for video summary and to examine the influence that varying circumstances have on the overall quality of the summaries. Also, we performed an experiment to determine the feasibility of this method, in which we recruited people to view and summarize video clips, and then we compared the results, based on their individual perspectives. We used a combination of 5 different algorithms, and a combination of multiple videos to see how the end user reacts to each of these videos every time they watch it. An important point that has been kept in mind is that each video comes in a multiple of two, that is one of the videos is kept same in a set of experiments. More details on how it has been performed are presented in section 3.

The experiment lasted for a total of 12 weeks and is meant to determine the feasibility of crowdsourcing methodology for producing video synopses. To build a voting poll of

Avrajyoti Dutta, Dawid Juszka and Mikolaj Leszczuk are with AGH University of Krakow, Faculty of Computer Science, Electronics and Telecommunications, Institute of Telecommunications, Poland (e-mail: avrajyoti.dutta@agh.edu.pl, dawid.juszka@agh.edu.pl, mikolaj.leszczuk@agh.edu.pl).

summaries, we invited participants to watch the video clips that were included in the dataset, which is a sample of videos. To have a fair comparison, we utilized scenario-wise algorithmic video summarizing strategies.

According to the findings of our research, the methodology of using crowdsourcing to summarize video clips is efficient, and it performs better than other techniques. Several criteria were applied to assess the overall quality of the summaries. The findings of the experiment demonstrated that the crowdsourcing strategy for generating high quality summaries is superior to conventional techniques in terms of both accuracy and the amount of time it takes to complete the task.

Hence, considering developing a novel metric for identifying User Generated Content (UGC) [5], and the broader expansion of this research endeavor's proposed approach. This content was frequently described as having amateurish acquisition conditions and unprofessional processing. However, with the advancement of easily accessible knowledge and affordable technology, individuals are now able to produce UGC that is nearly indistinguishable in quality from professional content [6], [7]. The evaluation will focus on the efficacy of the abstract system.

The rest of the paper is organized as follows: section 2 provides extensive details about crowdsourcing, section 3 details about the experimental design, whereas the 4th section deals with the methodology. In section 5, details about the experimental and statistical results, and finally in section 6 and 7, for the discussions and important conclusions with future directions are discussed.

II. Crowdsourcing

A method known as "crowdsourcing" is one in which a task is broken up into manageable chunks and then given to many different people to complete. The power of the crowd can be used to do work that is too challenging or impossible for machines to handle on their own [8].

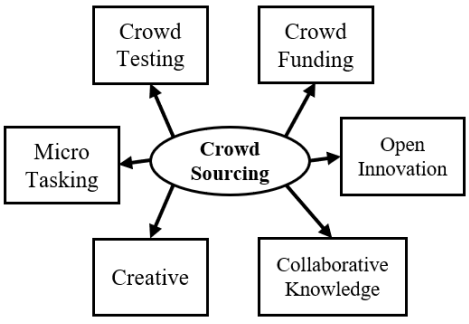

Fig. 1. Schematic representation of crowdsourcing and its different aspects

This includes delegating work to many individuals, most frequently using a digital platform. This strategy entails publishing a summary to a large audience by means of a platform based on the internet, and then collecting comments from the individuals who make up that audience. Although this strategy could provide a considerable volume of information in a relatively short amount of time, the quality of the feedback may vary depending on how it is implemented. Individuals are responsible for completing the work on their own, and the aggregated results are used to determine the outcome.

Image labeling, data annotation, and speech recognition are just a few of the applications where crowdsourcing has been employed effectively and to great effect [9]. In addition, crowdsourcing, which is depicted in Figure 1, may have wider-reaching ramifications to produce material as well as its curation. It is generally accepted that human-centered experiments on perceptual quality are the most trustworthy technique to evaluate the quality of an experience. The creation of UGC, which may be a strong tool for engagement and community building, can be accomplished through the usage of crowdsourcing. Additionally, crowdsourcing can be used to curate information for certain audiences, such as instructive videos for children or news highlights for busy professionals. This is another application of crowdsourcing. The use of crowdsourcing as an alternative to more conventional methods of summarizing is becoming increasingly common.

This strategy has several benefits, one of which is the capability to acquire summaries from a broad set of people in a short period of time, at a cost that is quite cheap, and with a limited amount of effort. As part of our strategy, we recruited people to watch video clips and vote synopses of what they saw. It is requested of the individuals that they summarize the video clips using the as per the ITU-T P.910 [10] recommendations, provided guidelines, which were handed to them along with the instructions. The generation of summaries from video clips has seen a rise in popularity in recent years thanks to the rise of crowdsourcing as an efficient method.

Nevertheless, utilizing the audience for the purpose of video summary has several obstacles to overcome. One of the difficulties is making sure that the summaries that are created by the crowd are of high quality. An additional difficulty is the possibility of bias and manipulation on the part of the crowd, both of which might lead to the production of summaries of poor quality.

Despite these obstacles, crowdsourcing for video summarizing has a huge opportunity to reap benefits in the future. Crowdsourcing can offer a complete and more varied viewpoint on video content, which can then lead to summaries that better capture the requirements and tastes of a variety of consumers. In addition, using crowdsourcing can cut down on the amount of time and money required for video summarizing, making the technique both more effective and more readily available.

III. Experiment Design

For summarizing the video clips, we have created a total of 5 unique algorithms, all of which are implemented in Python. Each method takes as input a video file, along with parameters specifying the desired length, and provides a summary of the video clips.

Using the pySceneDetect package [11] the video is segmented into scenes before being processed by each of the methods that are described below. Using the agh-vqis package

[12], a calculation is made to determine the average levels of Spatial Activity (SA) and Temporal Activity (TA) for each scene [13]. To represent the SA and TA of a single scene with a single value (which will be referred to in the following paragraphs as a "Coefficient"), a multiplication of these parameters is calculated for each scene. In addition, a detector of UGC is used for each scene to automatically determine whether the scene in question is made by users or by professionals. There is additional information regarding this UGC detector that may be found in [14].

The initial algorithm, denoted by the letter "A," is a baseline algorithm. More information regarding this method may be found in [15] . The only factors considered are TA and SA. The sequences in the video are sorted in a descending order based on the coefficient, which promoted the scenes that had a higher level of activity overall. Scenes from this list are picked to be included in the summary, working their way down from the top of the list until the desired length of the summary is reached. In the end, the chosen shots are arranged in the order that they appeared in the initial video's chronology.

The second method, denoted by the letter "B," made use of the UGC detector to high-light UGC scenarios in the summary. The remaining scenes are quantified into two distinct groups: the UGC scenes and the non-UGC (professional) scenes. Based on the descending order of the coefficients, each of these scenes are sorted. The scenes at the top of the list of UGC are selected first, and this process continues until the desired length of the summary is reached. If there is an inadequate number of UGC scenes, scenes from the top of the list of non-UGC scenes are chosen and added until the intended length of the summaries are reached. Following the same procedure as before, the chosen videos are organized according to their respective in the original video. We tested the hypothesis to explore the premise that UGC communicates more vital information than other types of material.

The UGC detector is employed by the third algorithm ("C") as well. To begin, the initial shot of the video that is used as input is included in the summary. This is because, in most cases, the initial shot communicates crucial information regarding the substance of the video. In this method, the shots are categorized into UGC and non-UGC lists before being ranked based on the coefficient, as is done previously. On the other hand, the shots are chosen in an alternating manner from the UGC and non-UGC lists until the appropriate length of time is achieved. After everything is said and done, the chosen videos are arranged in the correct chronological order. In this section, we investigated the hypothesis that the summary ought to include both UGC and non-UGC material in equal measure.

The fourth method, denoted by the letter "D," is precisely the same as the second one, denoted by the letter "B," with the sole distinction being that the first shot of the input video is always included in the summary. We have chosen to include this scenario because the "C" algorithm produced some encouraging results, which may be attributable to the fact that we included the first shot.

The fifth algorithm, "F," is quite comparable to the fourth and second situations, "D" and "B," respectively. The first shot is always a part of it. After that, this algorithm pushed shots that are not UGC in a manner that is analogous to the way the "B" scenario promoted UGC. In this experiment, we investigated the hypothesis that the UGC shots communicate less relevant information than the non-UGC shots (this is the opposite of the hypothesis that is tested in "D"), and the results are as expected.

After the scenes are selected for each algorithm, the resultant video is then automatically edited to concatenate the selected clips (in the order in which they occurred in time) and to incorporate the appropriate audio track. This is accomplished with the help of the well-known ffmpeg application [16].

The primary goal of industrial research is the development of new visual indicators as well as improvements for existing summarizing algorithms. The application utilizes several different visual cues as its foundation for its way of summarizing video sequences [17], [18]. This is the method by which the additional benefit of utilizing the UGC indicator are being measured.

IV. Methodology

A. Content and Encoding

The investigation is carried out with a group consisting of over 45 different people. The participants were given short video clips ranging from 30 to 45 seconds in length and asked to evaluate condensed versions of the material using a voting poll. To assess how well the proposed method works, we carried out an experiment with a set of 24 videos. Both the length and the subject matter of the video clips played a role in the selection process as shown in the flow chart in Figure 2. We compared our strategy to others, including strategies based on presentation order and random selection. A strategy known as content-based segmentation is utilized to cut the videos into more manageable chunks. The clips were chosen to provide a variety of examples of varying degrees of intricacy and visual depth. We also carried out a subjective review by requesting that human assessors describe the summaries based on the usefulness, coherence, and completeness of the information they contained.

B. Experimental Setup

The following procedures are included in the planned scientific investigation:

Step 1: The first step is the selection of video clips, and we choose these clips from a wide variety of sources, including the news, sports, entertainment, and other areas. The video is split into smaller chunks using a variety of criteria, such as user-defined segmentation or content-based segmentation, which are both examples of segmentation. The video clips should be between 30 and 45 seconds in length and incorporate a straightforward narrative structure. In order to create the video clips, we use different algorithms. Two distinct scenarios were outlined for each video clip, and participants were given a random assignment to one of the scenarios. The experiment has a total of 5 different scenarios.

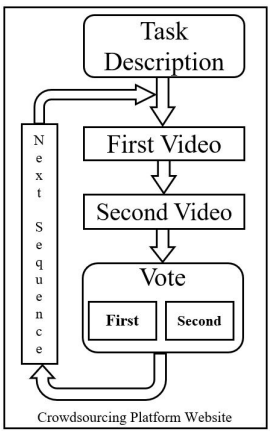

Fig. 2. Flow chart of the experiment

C. Participant Experiment

Step 2: Crowdsourcing Platform: In this step, we will recruit participants for the experiment using an online crowdsourcing platform, such as our own website (apart from sites such as Amazon Mechanical Turk and MicroWorkers). It is needed of the participants that they have a strong command of the Polish language and have a reliable internet connection to use with their laptop or tablet. In our experiment, we have designed a web-site for the assessment of the subjective video quality. We designed this platform in PHP, HTML5 and Java Script for video assessment, where a user watches a random set of videos and after watching two videos, provides a suitable vote about the quality of the summarized video. In order to eliminate the potential interference caused by the network connection, the browser downloads a video clip onto its local storage on each of the sequence. The videos are displayed in original resolution mode, and the participants are required to view the entire video before submitting their votes. It should be noted that no user-based modifications can be done while watching the videos. More details regarding this platform have been added as an open-source software on GitHub [19].

Step 3: Task Description: We give the participants a task description (as per the ITU-T P.910 recommendations) [10] that explains the goal of the experiment and provides clear directions on how to provide a summary of the video clip by offering a voting poll. In this step, we ask the participants to vote on which statement best summarizes the video clip.

Step 4: This step is a video summary, in which the participants are shown a brief video clip, instructed to view the entire video, and then asked to provide a summary of their evaluation in the video as their vote. Using a crowdsourcing platform, the selected votes are then provided to human annotators for review.

Step 5: The fifth step is to perform an analysis of the data collected. To do this, we collect the summaries that the participants have provided and examine them using histograms, pie charts and bar graphs. After that, we evaluate each of the summaries and then write an abridged version of the video's summary.

Step 6: In this step, we evaluate the quality of the summaries that were provided by the participants and compare them to summaries that were generated by state-of-the-art summarization algorithms. In order to evaluate how successful the crowdsourcing strategy is, we first compiled a summary of the voting results and then examined those results. For the purpose of assessing the statistical analysis, we make use of measures such as statistical significance formulas.

V. Results

The findings demonstrated that the individuals who took part in the discussion were able to appropriately evaluate various video clip summaries. We analyzed the outcomes of our methodology in comparison to the outcomes of existing video summarizing methods. We judged the effectiveness of the summaries based on how thoroughly and pertinently they covered the material. For the purpose of determining the overall state of the summary, the statistical analysis of scores based on subjective opinions is carried out. It is determined that the proposed method had a statistical significance level of 53.4%, which indicated that the summaries had a quality that is acceptable. Table I summarized with the all scenario wise values in percentage for which the experiments were performed.

TABLE I
Comparison between of each scenario

Different Scenarios	First Video	Second Video	Maximum Vote
Scenario A-B	58%	42%	First
Scenario A-C	51%	49%	First
Scenario A-D	39%	61%	Second
Scenario A-F	55%	45%	First
Scenario A-D-45	48%	52%	Second

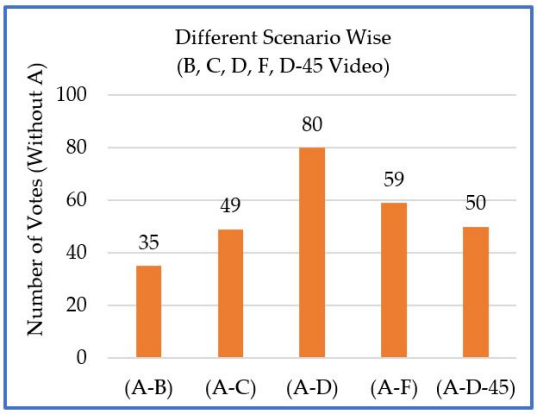

Fig. 3. The histogram of number of votes for different scenarios mentioned in Table I. It is visible that A-D scenario has outperformed in second video selection.

The number of votes cast for the second video is higher than those cast for the first video; specifically, 58% of the votes were cast for the second video clip that is shown. One strategy for achieving this improvement is by performing an analysis on the output that these algorithms create.

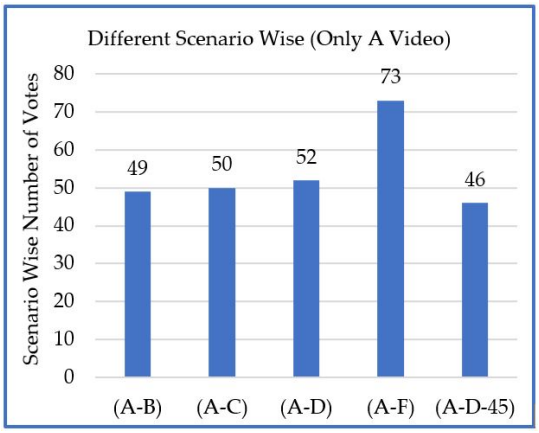

Fig. 4. The histogram of number of votes for different scenarios mentioned in Table I. It is visible that A-F scenario has outperformed in first video selection.

The ”A” scenario is assessed in comparison to 5 separate scenarios, namely B, C, D, F, and D-45, along the X axis. According to our research, the ”D” scenario garnered a total of 80 votes in support of the second video on the Y axis in Figure 3. Conversely, the scenario denoted as ”F” produced contrasting outcomes in Figure 4. The results of this study indicate that various attributes of the video, including its content, duration, and intricacy, have a substantial influence on the effectiveness of video summarizing algorithms.

It is possible to uncover observed patterns and trends through the examination of algorithmic output, which can then serve as indicators of prospective areas where algorithms should be improved. The generated data can be utilized to modify the parameters of the algorithm or to design fresh approaches for the processing of data, with the intention of increasing the precision of the algorithm.

The evaluation of algorithmic outputs can not only assist us in identifying potential areas for improvement, but it can also facilitate our grasp of the functionality and the elements that affect the performance of the algorithms. The findings of the subjective evaluation demonstrated that the summaries produced by the proposed technique are judged to be more helpful, cohesive, and comprehensive than the summaries produced by the other methods. This data has the ability to either assist in the development of novel algorithms with improved efficacy or to optimize the production processes of algorithms that are already in existence. The findings of the experiment demonstrated that using crowdsourcing as a method for evaluating video summary can be an efficient and useful technique.

In general, the evaluation of the results produced by algorithms is a powerful tool that may be used to improve the accuracy and efficiency of the algorithms being used. It is possible to get substantial insights on the functionality of the algorithms and the variables that affect their behavior via the analysis of the output of the algorithms. With the aforementioned information at our disposal, we can proceed with the development of new algorithms or the improvement of ones that already exist. This can help in providing solutions that are more accurate and efficient for difficult problems. In this contribution, we gave best practices for quality of experience crowd testing by providing a comprehensive analysis of the primary challenges that come up during crowd testing for quality of experience and the solutions that are connected to those challenges using the example of quality of experience evaluation for video. We focused primarily on the design, execution, and reliability evaluation of the crowd testing campaigns for effective Quality of Experience (QoE). We have compiled a set of best practices for crowdsourcing quality of experience testing, which can be found here. The quality of the user experience is receiving an increasing amount of attention from service providers as a direct result of the growing significance of the level of customer satisfaction about internet applications and services [20]. The process of dividing video clips into discrete clips that display varying degrees of complexity and visual richness is accomplished by the employment of criteria and an examination of content features, which is referred to as the approach of content-based segmentation. The method intends to improve the video’s manageability so that it may be processed or analyzed more easily in the future.

VI. Discussions

We compared the ’A’ scenario to the ’B’, ’C’, ’D’, ’F’, and ’D-45’ situations when carrying out our analysis. According to the results of our investigation, the ”D” scenario obtained 80 votes cast in support of the second video. On the other hand, the outcomes of the ”F” scenario were completely different. Here the first video was viewed the most, in comparison to the second video. These findings imply that aspects of the video, such as its content, length, and complexity, can have a major impact on how well video summarizing systems work. It is essential for these aspects to be considered in the investigation and development of video summarizing algorithms. In addition, the findings of this study show the importance of conducting an exhaustive and methodical evaluation of video summarizing algorithms by making use of a variety of datasets and scenarios, as opposed to depending on just one evaluation metric or scenario. Doing so will allow us to obtain a better knowledge of the capabilities and constraints of various algorithms, which will influence the development of such algorithms for a wide variety of applications.

Amongst other things, the video quality indicators developed by the team at the University of Texas at Austin’s Laboratory for Image and Video Engineering (LIVE), such as the FRIQUEE indicator and OG-IQA, the HOSA indicator, the WaDIQaM indicator, or the CORNIA indicator [21], have been used for a comparison with respect to our proposed methodology.

The proposed research [22] is justifiable in view of the aims of the experiment since it provides more quality indicators to the collection of tools that support the summarizing system. Additional indicators may lead to a system that creates summaries of a higher quality (particularly in terms of selecting visually appealing situations for the material recipient), which may be the result of a summarizing system that could lead to the generation of summaries.

The findings of our experiments indicate that the crowdsourcing strategy that has been offered is superior to other methods in terms of both its relevance and its level of detail. Our research contributes to the closing of this knowledge gap and investigates the influence that the layout of the study's interface had on the participants' impressions of the video's overall quality. The objective and subjective evaluation of the quality of video summarizing algorithms may show both the former's merits and the latter's faults. By soliciting the feedback of human evaluators, we may be able to identify areas in which further development is warranted and construct more robust algorithms that are able to successfully extract the most important information from the source video and present it in a condensed and easily digestible way.

VII. Conclusions

In this research work, we proposed a novel method for the summarizing of videos that makes use of crowdsourcing. We devised an experiment in which people were paid to view and summarize video segments, and then we watched the results. We demonstrated the usefulness of the method that is suggested by contrasting it with many alternative approaches to video summarizing that are already in use. The findings of our experiment demonstrated that the strategy that is proposed offers a considerable improvement over existing methods and is a potential way for summarizing the content of video clips. The proposed method offers several benefits, including the fact that it is adaptable, scalable, and economical. The evaluation of the algorithms that are used to summarize videos is very important, and the subjective experiment of the video's quality plays a vital role in this. This provides useful insights on how the user perceives the summary and identify potential areas where the summary can be improved. The selection of the method for evaluating subjective quality is dependent on several different elements and requires careful thought in order to guarantee the accuracy and consistency of the results.

The findings of the experiment indicate that crowdsourcing enables the compilation of a wide variety of viewpoints and has the potential to be an effective method for the summarizing video content. It excels in performance compared to other methods, with the quality of the summaries varying according to the circumstances. In the end, the utilization of crowdsourcing for the purpose of video summary has the potential to change the way we produce and consume video information, ultimately rendering it more exhaustive, diversified, and accessible to everyone.

The experiment is not without its flaws, such as the restricted number of possible outcomes. The usage of additional scenarios and the evaluation of the summaries by end-users to assess the usability and efficacy of the findings could be investigated in further research in the future. Apart from that, other approaches of segmentation and selection will be investigated, and the proposed method will be tested using more extensive datasets.

Acknowledgment

This work was funded by Ministry of Science and Higher Education as part of a subsidy for the Faculty of Computer Science, Electronics and Telecommunications, AGH University of Krakow. The authors would like to thank experts for their appropriate and constructive suggestions to improve this article.

References

[1] M. Grega, L. Janowski, M. Leszczuk, P. Romaniak, and Z. Papir, "Quality of experience evaluation for multimedia services," *Przeglad Telekomunikacyjny I Wiadomości Telekomunikacyjne*, vol. 81, pp. 142–153, 01 2008. [Online]. Available: https://sigma-not.pl/publikacja-34775-quality-of-experience-evaluation-for-multimedia-services-przeglad-telekomunikacyjny-2008-4.html

[2] T. Hoßfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, "Best practices for qoe crowdtesting: Qoe assessment with crowdsourcing," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 541–558, 2014. [Online]. Available: https://doi.org/10.1109/TMM.2013.2291663

[3] M. Leszczuk, M. Grega, A. Koźbiał, J. Gliwski, K. Wasieczko, and K. Smaïli, "Video summarization framework for newscasts and reports – work in progress," in *Multimedia Communications, Services and Security*, A. Dziech and A. Czyżewski, Eds. Cham: Springer International Publishing, 2017, pp. 86–97. [Online]. Available: https://doi.org/10.1007/978-3-319-69911-0_7

[4] S.-Y. Wu, R. Thawonmas, and K.-T. Chen, "Video summarization via crowdsourcing," in *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 1531–1536. [Online]. Available: https://doi.org/10.1145/1979742.1979803

[5] M. Leszczuk, L. Janowski, J. Nawała, and M. Grega, "User-generated content (ugc)/in-the-wild video content recognition," in *Intelligent Information and Database Systems*, N. T. Nguyen, T. K. Tran, U. Tukayev, T.-P. Hong, B. Trawiński, and E. Szczerbicki, Eds. Cham: Springer Nature Switzerland, 2022, pp. 356–368. [Online]. Available: https://doi.org/10.1007/978-3-031-21967-2_29

[6] M. Leszczuk and M. Duplaga, "Algorithm for video summarization of bronchoscopy procedures," *Biomedical engineering online*, vol. 10, p. 110, 12 2011. [Online]. Available: https://doi.org/10.1186/1475-925X-10-110

[7] P. Romaniak, M. Muy, A. Mauthe, S. D'Antonio, and M. Leszczuk, "Framework for the integrated video quality assessment," *Multimedia Tools and Applications*, vol. 61, 12 2011. [Online]. Available: https://doi.org/10.1007/s11042-011-0946-3

[8] W.-T. Tsai, L. Zhang, S. Hu, Z. Fan, and Q. Wang, "Crowdtesting practices and models: An empirical approach," *Information and Software Technology*, vol. 154, p. 107103, 2023. [Online]. Available: https://doi.org/10.1016/j.infsof.2022.107103

[9] M. Shahid, J. Søgaard, J. Pokhrel, K. Brunnström, K. Wang, S. Tavakoli, and N. Gracia, "Crowdsourcing based subjective quality assessment of adaptive video streaming," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014, pp. 53–54. [Online]. Available: https://doi.org/10.1109/QoMEX.2014.6982289

[10] International Telecommunication Union (ITU). [Online]. Available: https://www.itu.int/rec/T-REC-P.910-202207-I/en

[11] "Pyscenedetect." [Online]. Available: https://www.scenedetect.com

[12] AGH Video Quality of Experience (QoE). [Online]. Available: https://qoe.agh.edu.pl/indicators/#indicators

[13] P. Romaniak, L. Janowski, M. Leszczuk, and Z. Papir, "Perceptual quality assessment for h.264/avc compression," in *2012 IEEE Consumer Communications and Networking Conference (CCNC)*, 2012, pp. 597–602. [Online]. Available: https://doi.org/10.1109/CCNC.2012.6181021

[14] M. Leszczuk, M. Kobosko, J. Nawała, F. Korus, and M. Grega, "In the wild" video content as a special case of user generated content and a system for its recognition," *Sensors*, vol. 23, no. 4, 2023. [Online]. Available: https://doi.org/10.3390/s23041769

[15] A. Badiola, A. M. Zorrilla, B. Garcia-Zapirain Soto, M. Grega, M. Leszczuk, and K. Smaïli, "Evaluation of improved components of amis project for speech recognition, machine translation and video/audio/text summarization," in *Multimedia Communications, Services and Security*, A. Dziech, W. Mees, and A. Czyżewski, Eds. Cham: Springer International Publishing, 2020, pp. 320–331. [Online]. Available: https://doi.org/10.1007/978-3-030-59000-0_24

[16] ffmpeg Application. [Online]. Available: https://www.ffmpeg.org

[17] N. Cieplińska, L. Janowski, K. De Moor, and M. Wierzchoń, "Long-term video qoe assessment studies: A systematic review," *IEEE Access*, vol. 10, pp. 133 883–133 897, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3231747

[18] P. Pérez, L. Janowski, N. García, and M. Pinson, "Subjective assessment experiments that recruit few observers with repetitions (FOWR)," *IEEE Transactions on Multimedia*, vol. 24, pp. 3442–3454, 2022. [Online]. Available: https://doi.org/10.1109/TMM.2021.3098450

[19] Github repository. [Online]. Available: https://github.com/dutta-agh/TANGO_A-B

[20] K. Borchert, A. Seufert, E. Gamboa, M. Hirth, and T. Hossfeld, "In vitro vs in vivo: Does the study's interface design influence crowdsourced video qoe?" *Quality and User Experience*, vol. 6, 12 2021. [Online]. Available: https://doi.org/10.1007/s41233-020-00041-2

[21] M. Leszczuk, L. Janowski, J. Nawała, and A. Boev, "Objective video quality assessment method for face recognition tasks," *Electronics*, vol. 11, no. 8, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/8/1167

[22] Crowdsourcing Evaluation of Video Summarization. [Online]. Available: http://pbz.kt.agh.edu.pl/~testySubiektywne/QoE_Dutta/TANGO/