A browser based system for remote evaluation of subjective quality of videos

Robert Furman Rafał Hadzicki Damian Karwowski Krzysztof Klimaszewski

Poznan University of Technology
Pl. M. Skłodowskiej-Curie 5
Poland, 60-965, Poznań
krzysztof.klimaszewski@put.poznan.pl

ABSTRACT

The evaluation of the quality of videos is posing a significant problem in research on video processing, including compression. Different algorithms and their modifications aim to improve the quality of the video, while keeping the necessary bitrate as low as possible, and every time, the performance of the method must be evaluated. The bitrate can be measured readily, while for quality there are different metrics and, therefore – different results. The ultimate verdict always has to be the subjective opinion about the quality that is expressed by the viewers. For now, the only reliable way to measure the quality of the video is to perform a subjective test with a group of viewers. Subjective tests are difficult to perform – not only one has to gather all viewers in one place, but also the evaluation schemes may be difficult to follow. In the paper we present an implementation of a system that greatly simplifies the process of performing the subjective quality assessment of videos. We discuss strengths and weaknesses of the system when compared to the commonly used procedures. Finally, we also provide the results of a quality assessment for the HEVC video encoder. Presented data proves that the idea of remote quality evaluation together with the usage of the two-level grade is a valid one and provides reliable results.

Keywords
subjective video quality

1. INTRODUCTION

In the process of developing the new video compression algorithms or modifying some processing steps in the video compression pipeline, several factors must be accounted for. Surely, one of the factors is the computational complexity, that translates to the time required to perform the processing step, the ability to perform the calculations in parallel, then there is the bitrate gain or loss that influences the size of encoded data stream, and finally, there is the quality of the reconstructed video. The last factor is difficult to evaluate because it has to take into account the way that humans perceive video, since most of the time the goal is to have as high a quality perceived by the viewers as possible. It is difficult to obtain the results similar to the perceived quality in software. Many commonly used metrics, like PSNR or SSIM are aimed at static image quality evaluation, and even then, for some specific cases, they provide results dramatically different than subjective tests performed by viewers. Therefore, new metrics are being developed, like VMAF that aim to model the perception mechanisms of video and provide results that closely correlate with the subjective tests.

The importance of being able to use a piece of software to evaluate the quality of the video and get results that agree well with the subjective tests is obvious, when we consider the applications in rate-distortion evaluation in video compression algorithms. For such applications one needs to perform many evaluations during a compression run.

In the absence of such algorithms, the only reliable way to evaluate the quality of the video is to perform the subjective tests, which is a very difficult process.

The methods for performing the subjective quality assessment has been formalized a long time ago and the details can be found in official recommendations, like ITU-R BT.500 [BT500] as well as the ITU-T P.910 [P910], P.911 [P911] and P.913 [P913]. Most of the recommendations are regularly updated. The update of recommendations is necessitated mostly by the developments in the video delivery methods.

In the recent years, the viewers are no longer limited to cinemas and stationary TV-sets in order to view videos. Nowadays the viewers often wish to watch videos on their smartphones or laptops in places, that usually provide less than perfect viewing environment (e.g. lighting, external distractions).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

https://www.doi.org/10.24132/CSRN.3201.36

2. SUBJECTIVE QUALITY EVALUATION

The evaluation process

The quality tests that follow the abovementioned recommendations require to gather volunteers to act as viewers and present the videos to be evaluated in a certain way, usually in a controlled environment, that is free from external distractions. There are several main types of quality evaluation procedures, and the most important ones are presented below.

Single stimulus (absolute category rating) in which the viewers are shown a single sequence and are asked to evaluate its quality in a 1 to 5 scale, 1 being “bad” and 5 being “Excellent”. Each viewer is asked to evaluate many videos in such a way. Since the results of such a simple evaluation depend on the sequence content in a broad sense, the test is usually augmented by adding a hidden reference to the test to normalize the results. Such a reference would, for example, be the original, uncompressed video.

A more reliable method is a double stimulus one, where the viewer is shown two sequences and asked to evaluate the quality. One of the sequences that are shown is a reference (i.e. original, uncompressed) and the other one is the evaluated sequence. Depending on the variation of the method, the viewer does know or does not know which sequence is the reference one. In the first case, the viewer evaluates the impairment of the evaluated sequence in the scale of 1 (annoying artifacts) to 5 (imperceptible differences) (DSIS: double stimulus impairment scale), in the second case, the viewer evaluates the comparative quality of the second video in the scale of -3 (much worse quality) to +3 (much better quality) (DSCS: double stimulus comparison scale).

The result of the evaluation can be presented in a form of an average quality (MOS – mean opinion score) and with the confidence interval calculated with the use of the standard deviation of the results.

For DSIS, the results can be directly used to arrange a set of many different sequences with respect to their quality, while for DSCS there is no such direct possibility. This should make the DSIS to be the superior evaluation model, however, there are some inherent problems with this model, that will be discussed below.

Challenges in the process

The methods described above pose some challenges. First of all, a significant number of viewers has to be involved in the evaluation process, usually well above 10 participants are required. The participants should not be experts in video processing, since this could bias their evaluation. The participants need to be trained so that they know what kind of distortions to expect and where to put their attention during the viewing. Also, the whole process of evaluation needs to be explained. The evaluation needs to be performed in a somewhat controlled environment, and this limits the number of people that can view and evaluate the sequences at once. The viewing is usually time consuming, as the viewers are usually expected to evaluate significant number of sequences. Especially in double stimulus scenarios this consumes a lot of time and is very arduous for the viewers.

In some circumstances it is extremely difficult to gather that many people willing to spare a significant amount of their time to perform the evaluation. In the recent years it has become even more challenging, due to the sustained pandemic situation. Therefore, the idea to perform the evaluation on-line emerged, that will be presented in the paper.

Another challenge is interpretation of the results. The wide scale of the evaluation poses a significant risk for the viewers, since it is really difficult to decide whether to give the just viewed video the mark of 4 or 5. Also, there is a question whether the video is slightly worse, worse or much worse than the previously seen one. Those are difficult questions that the viewers have to consider, and the result is not easily predicted. Any attempts to show the viewers what artifacts to look for and how to judge them is only biasing the viewers and should be avoided.

Therefore, the interpretation of the results is difficult. The results can be influenced by the order of the videos during the evaluation, since the viewers may start to modify the marks given to consecutive videos. As long as the number of viewers is not high enough, those factors may significantly bias the results, even when random order of sequences is used.

3. THE PROPOSED METHOD

Motivation

During the pandemic, it is difficult to gather viewers to evaluate sequences. Even before the pandemic it was difficult, since the viewers can rarely expect to be paid for their time spent on the evaluation. Therefore, an online tool is required to perform the evaluation.

Also, the time required for the evaluation by a single user needs to be decreased. In the ITU recommendations it is suggested that a single session should be kept as short as possible, the suggested limit being 20 minutes for short sequences. In the contemporary research the test sequences are usually short, in the order of 10-20 seconds each, so the advised limit of 20 minutes does apply.

This limit means that about 20 pairs of sequences can be shown to a single viewer. This may cause fatigue for the viewer and also means that significant amounts of data need to be transferred to the viewer. Therefore, we suggest to limit the number of sequences shown to a single viewer. This way we are able to recruit more volunteers. The most of the time is spent on training and explaining the judging procedure. This can easily be done online for all the participants at once. The viewing can be done later by each viewer individually, therefore not requiring that much time from the viewers and no waiting for their viewing turn.

We also observe the problem with the marks that was signaled earlier – many users hesitate about the rating they should give and their judgement can easily saturate during the test or be adjusted halfway through the test. To alleviate those problems, we suggest to use the DSCS scheme but to make the judging procedure much easier and significantly limit the number of possible marks.

To summarize this section, the motivation for developing the described system was threefold: to increase the number of recruited viewers, to make the judging simpler and to make the whole process less time consuming, both for the viewers and for the researchers.

Existing solutions

The problem of the subjective quality evaluation is not new, therefore there exist several solutions for the remote evaluation of subjective quality of video. A review of such solutions is presented in [Uhr20a]. The author present their own solution of such a system as well. Some other systems for remote evaluation of videos are described in [Jai13], where a system is developed that enables an online gathering of voting results and is available for download and use. Another system for evaluation of video quality is presented in [Rai13]. This system is very close to the system described in our paper, however it seems that it does not standardize the coding format of the videos and therefore can only be used to evaluate the quality for video coders for which a plug-in or codec pack already exists. It is therefore not suited for research on new codecs or new, non-standard modifications of existing codecs. No comments on possible use for any kind of video processing (like post processing of videos) are given. Another advanced system is described in [Che10]. This system is fully based on an Adobe Flash technology, that is outdated and not supported any more.

The comparison of the remote and local video quality assessment results are presented in the [Uhr20b]. The comparison of the results of the same evaluations performed remotely and locally show very high correlation of the MOS values and authors claim that the remote quality evaluation can replace the laboratory tests.

Many of the systems mentioned above are, unfortunately, either not accessible any more or use outdated technology (like Adobe Flash scripts).

Our system is developed to be able to evaluate any kind of video processing technique, including postprocessing or any kind of modification. It is meant to work based entirely on web browser, not requiring any extensions nor codecs. It is also designed so that no access to commandline on the server is required and a simple web server services with a database access are sufficient.

The implementation

The schematic of our proposed system is shown in Figure 1.

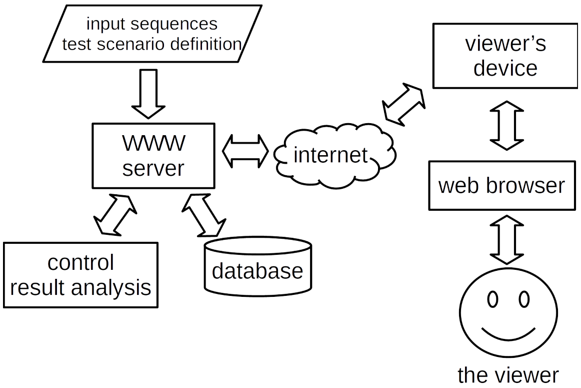


Figure 1. The schematic of the proposed system.

From the point of view of the viewer, the whole system is based on the web browser. The videos to be viewed are sent from the web server to the viewer's device and presented to the viewer in the correct order. The viewer is expected to evaluate the sequences by selecting the proper mark in the browser. The rating, along with the information about the user, such as the reported screen resolution of the user's device, is sent back to the server and is stored in the database for further analysis. Also the basic information about the viewer are stored: the age and sex of the viewer.

We decided that in order to make the whole process for the viewers as nonintrusive as possible, we should aim to prepare the whole system to use a web browser as an interface. This means that the user does not have to install any software in order to take part in the evaluation. Although this is a significantly limiting factor (a custom-made application would provide far more possibilities), we believe that the tradeoff is in favor of the browser based solution. Our choice also means that no non-standard libraries nor codecs should be required and the videos have to be encoded in a way that enables most of the web browsers to decode them natively.

In fact, video sequences that are on the web server are in compressed form. An important factor for this compression is not to introduce any artifacts to the videos. Therefore the coder should provide a lossless or nearly-lossless compression. The lossless video compression, even for very short clips (10 seconds long) produces very large amounts of data, therefore the nearly-lossless compression is preferred from the point of view of the feasibility of the system. The results of the tests show that the use of nearly-lossless compression should significantly reduce the amount of data to be sent to the user while the difference in perception of the data before and after such compression is expected to be insignificant.

The review of possible codecs was performed and the final choice was to use the VP9 codec using a WebM container. Currently, the lossless coding is used, although the system can accept nearly-lossless coded videos readily. VP9 supports both coding methods.

The system is prepared in the even more popular and capable JavaScript language and the React library. The database system used is MySQL.

The fact that the system is fully dependent on internet browser results in some difficulties caused by incompatibility of some platforms that require special approach, but, on the other way, allows many different platforms that are compatible to be used during evaluation. Still, the browser approach seems to be superior to preparing applications for wide variety of platforms, especially for smartphones. The evaluation can, in principle, be done on a smartphone and personal computer. This provides a wide variety of screen resolutions, viewing distances and screen sizes, together with different surroundings. One can even expect to get results from people that were performing the session in means of public transport on a loud street. Such a possibility means that we get results for the actual surroundings in which the user is usually watching video content. This seems to be a significant advantage of the proposed method.

To differentiate between the most important types of devices, the system stores the screen resolution reported by the browser.

The evaluation process

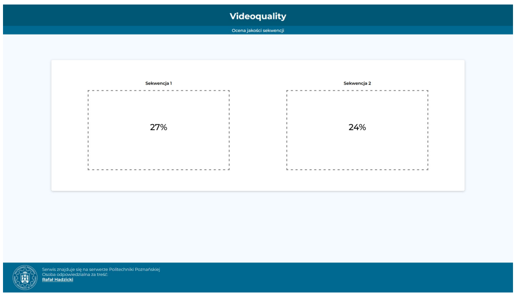
When user enters the website dedicated to the system, the basic information about the age and sex of the viewer is gathered. Then, a specific set of sequences to be shown to the user are randomly chosen.

Next, the pairs of videos are shown and the user is asked to evaluate the comparative quality of the videos. The pairs of videos are shown on full screen. This is a problematic feature, since there are compatibility issues between different systems and web browsers that still need to be addressed separately for some combinations. Before the playback of the pair of sequences, they are buffered in the viewer's device. This is done in order to decrease the probability of pauses during the playback. This is another challenging issue, since even the compressed streams are large (hundreds of megabytes each) and the need to download and buffer them poses a set of challenges for the network connection as well as the buffer memory for the web browser on the viewer's device. The challenge is much smaller if almost-lossless compression for the evaluated videos are used.

In the current form, the system is used to perform a direct comparison between two sequences and only two choices are given – the viewer is asked to select the video with better quality from the two shown. This makes the process much simpler for the viewer, since a simple question needs to be answered. It needs to be stressed that videos of the same quality are never shown, therefore there always are differences between the two sequences. The same approach is suggested in [Che10] and agrees well with our observations described above.

Each user can perform multiple sessions and since the videos for each session are randomly selected, such an approach is welcome and provides additional data.

The screenshots of the consecutive pages of the developed webpage as seen on a personal computer using a 1920x1080 screen are presented in Figures from 2 to 5. First, the data is buffered to avoid stalls during the playback. At this stage, the page displays the progress, as shown in Figure 2.



Videoquality

Ocena jakości sekwencji

Sekwencja 1

27%

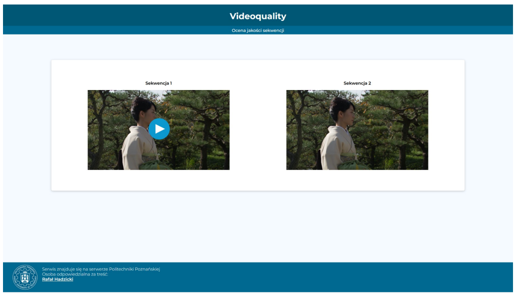
Sekwencja 2

24%

Serwis znajduje się na serwerze Politechniki Poznańskiej
Osoba odpowiedzialna za treść:
Rafał Hadzicki

Figure 2. Loading the video data.

When the data is buffered, the first video can be played. Playing of the second video is not possible now. When the play button is clicked (see Figure 3), the video is shown on the full screen.



Videoquality

Ocena jakości sekwencji

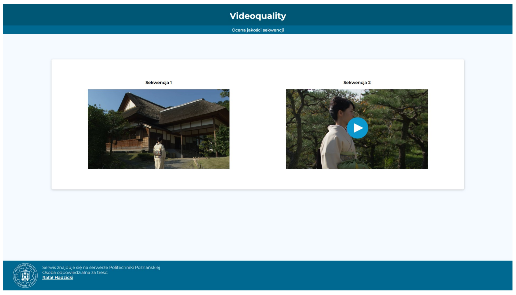
Sekwencja 1

Sekwencja 2

Serwis znajduje się na serwerze Politechniki Poznańskiej
Osoba odpowiedzialna za treść:
Rafał Hadzicki

Figure 3. Data loaded - ready to show first video.

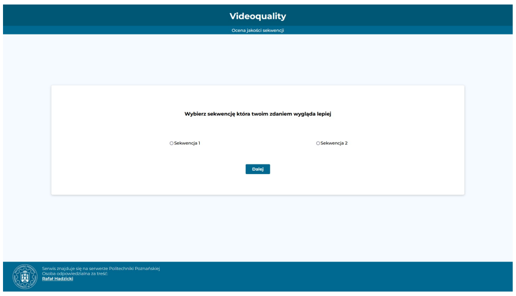
After viewing the first video, the second video can be played. The play button is shown on the second video and replaying the first video is not possible (see Figure 4).



Serwis znajduje się na serwerze Politechniki Poznańskiej
Osoba odpowiedzialna za treść:
Rafał Hadzicki

Figure 4. Ready to show the second video.

Directly after viewing the second video, the voting starts. The voting page is shown on Figure 5. After voting, the new pair of videos is loaded, or, when the last pair was graded, the session ends and a “thank you” page is displayed.



Wybierz sekwencję która twoim zdaniem wygląda lepiej

○ Sekwencja 1 ○ Sekwencja 2

Serwis znajduje się na serwerze Politechniki Poznańskiej
Osoba odpowiedzialna za treść:
Rafał Hadzicki

Figure 5. Voting.

Result analysis

After a significant number of viewers finish their voting sessions, the results can be analyzed.

For cases when viewers select a sequence with a better quality, the direct comparison of the results is not straightforward and cannot be performed by a simple comparison of values, as it is the case for DSIS method. A more sophisticated method is required. Here, we adopt the method for result analysis using the appropriate preference matrix, similar to the one used in [Che10].

The chosen way of grading the videos does not, at this time, allow to order a number of sequences with respect to their perceived quality. Only comparative results are possible. This, however does not necessarily limit the usability significantly. Many times one is interested to compare two methods directly and either demonstrate the superiority or inferiority of one method, or prove the equal performance of two different methods. For such cases the developed system seems to be perfectly suited.

4. TEST OF THE SYSTEM

Test scenario

For a test scenario used to verify the developed system we use a set of HD sequences compressed with HEVC encoder using different QP indices. The actual encoder used was the reference implementation HM version 16.6. The QP values used were 22, 27, 32 and 37. The original (i.e. uncompressed) sequences are not used in the survey. Two test sequences were used: Kimono and ParkScene. Those sequences are a part of the set of sequences recommended by ITU-T and ISO/IEC for research on video coding [Bos12]. Every user is asked to select the better quality video for two pairs of videos. One pair of videos is two randomly selected encoded videos of Kimono sequence, the second pair is two randomly selected videos of ParkScene sequence. The sequences are transmitted to the viewers’ device in a form of losslessly compressed streams. The sequences are displayed at full screen, therefore are resized to the display resolution by the web browser.

Results

The test involved gathering 70 sets of votes. This means that 140 pairs of videos were rated and the better one for each pair was selected.

The results for the Park Scene sequence are shown in Table 1.

		Chosen as better (QP)			
		22	27	32	37
Not chosen as better (QP)	22		8	4	0
	27	8		7	2
	32	6	7		0
	37	9	11	8	

Table 1. Results for the ParkScene sequence.

The results from the Table 1 are interpreted in the following way. When the pair of sequences compressed with QPs of 32 and 22 were shown to the viewers, 6 times the sequence with QP=22 was selected as better (blue frame in Table 1) and 4 times the sequence with QP=32 was selected as better (green frame in Table 1). The pair 22-32 was shown 10 times, and 6 times out of 10 the sequence with QP=22 was selected as the better one. The numbers below the main diagonal correspond to the cases when the sequence with the lower QP was selected as the better one (and, in general, that is the expected result). For ParkScene this happened 49 out of 70 times, while 21 times the video with higher QP was selected. The results for the Kimono sequence are shown in Table 2.

https://www.doi.org/10.24132/CSRN.3201.36

		Chosen as better (QP)			
		22	27	32	37
Not chosen as better (QP)	22		1	2	5
	27	8		3	0
	32	12	10		0
	37	11	10	8	

Table 2. Results for the Kimono sequence.

It can be seen that for the Kimono sequence the users most of the times selected the sequence with lower QP (59 out of 70 times). It may seem that for the Kimono sequence the differences between different QP values are more visible.

The interesting comparison can be done when analyzing the results for cases when a regular computer and a smartphone was used. The results for smartphones are shown in Table 3, and the results for regular PC (usually FullHD or bigger resolution screen) are shown in Table 4.

		Chosen as better (QP)			
		22	27	32	37
Not chosen as better (QP)	22		3	3	3
	27	3		4	1
	32	2	1		0
	37	2	3	1	

Table 3. Results for smartphones (both sequences).

		Chosen as better (QP)			
		22	27	32	37
Not chosen as better (QP)	22		6	3	2
	27	13		6	1
	32	16	16		0
	37	18	18	15	

Table 4. Results for computers (both sequences).

The ratio between the cases when the lower QP is chosen over higher QP to the total cases is 12 over 26 (only 46% of cases), while for computers this ratio is 96 over 114 (84% of cases).

Such results are not surprising, since it can be expected that for smartphones the differences in quality of the sequences compressed with different QP may not be visible at all. The screen can simply be to small to notice the differences easily. We can, for example, notice, that when Kimono sequence encoded with QP of 37 was compared to the sequence with QP of 22 (a really surprising choice!), only in 5 cases the QP37 sequence won. Out of those 5 cases, 3 cases were the smartphone users. The two remaining cases may be regarded as mistakes, since it is really difficult to believe that when viewing on a big screen the viewers would not notice the significant artifacts for the QP 37 case.

Unfortunately, such mistakes or deliberate actions cannot be avoided entirely and especially in short viewing sessions it is not possible to filter them out.

Confidence interval calculation

The results presented above are given only for a certain sample of the population. It is expected, therefore, that the ratios calculated for the results from a sample of population may differ from the ratio calculated hypothetically for the entire population. In order to measure the confidence of the calculated results one needs to perform statistical evaluation of the results. For the example described above, the only possible choices for the user are better/worse quality within a pair of sequences, therefore the viewing results may be regarded as the results of a binomial trial (Bernoulli trial). For such cases, there are established methods for calculating the confidence interval, as explained in [Ros03]. The method of choice of estimating the confidence intervals for the conducted experiments is the "exact" Clopper-Pearson method, due to its popularity and robustness. An example below is given for the case when the quality of the Kimono sequence compressed with QP 22 is compared to the quality of the sequence encoded with QP 32. From Table 2 we can see that the QP22 sequence is chosen as the better one in 12 cases and the QP32 is chosen in 2 cases. Therefore the proportion of cases when the lower QP produces a sequence that is regarded to have a higher quality is 12/14 = 85,7%. The estimate of the confidence interval at the 95% confidence level would be from 57.19% to 98.22%. Since the lower bound is higher than 50%, we can, at this level of confidence, say that majority of population would perceive the QP22 case as that of a higher quality than QP32.

The confidence interval in this case is quite wide, and this alone supports the idea about performing tests at as high a number of viewers, as possible. For example if the results were 120 cases in 140 cases total, the confidence interval would shrink to from 78,8% to 91,05%. Our system makes it much easier to gather such number of marks.

5. SUMMARY

In the paper we presented an idea and an implementation of the system for remote evaluation of subjective quality of video. The system enables the grading to be performed in the real life situations, for example on smartphones in places where people actually watch videos, and also on personal computers. This, however, may be perceived as a drawback – inability to fully control the environment and the viewing method, but this, we believe, is offset by the real life experience during the tests.

The system makes it much easier to gather significant number of grades for the videos, when compared to traditional ("face to face" or "local") viewing sessions. The system is configurable, any desired method of double stimulus judging can be implemented, although the main idea behind the system was to implement a "binary" method of selecting a better sequence among the two shown.

The results are available in real time, even during the ongoing tests. The results can be gathered any time and the basic statistics can be calculated.

The important feature of the system is that it is coder agnostic. The raw videos are encoded to a common format (lossless VP9 in WebM container) and therefore does not require any external codecs nor applications. Most of the contemporary web browsers supporting JavaScript are compatible with the system.

The main drawbacks of the system are concerned with the huge amounts of data, in the form of the test sequences, that not only need to be stored on the server, but also transmitted to the viewers. This limits the possible number of the sequences used during the test and limits the overall test length for a single viewer (some people are not prepared to download hundreds of megabytes of test video streams at once).

6. FUTURE WORK

The most important changes of the system, required for any further development and widespread use, would be to limit the amount of data that needs to be stored and transmitted. The only viable option here is to use a nearly-lossless compression. This, however, requires further study to choose the proper settings.

The compatibility of the system needs to be improved, especially in relation to the iOS system.

If the ordering of the quality of several sequences is required, the method similar to the Transitivity Satisfaction Rate principle described in [Che10] can be tried in the processing of the results.

Further developments, regarding the test scenario configuration flexibility, test sequences storage and the overall security and reliability of the system, are envisaged in the near future. The ability to perform different test scenarios in parallel is one of the other possible modifications.

7. ACKNOWLEDGMENTS

The work was supported by the Ministry of Education and Science, Poland.

8. REFERENCES

[Bos12] F. Bossen, Common test conditions and software reference configurations, Joint Collaborative Team on Video Coding (JCT-VC) of ITUT SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCTVC-J1100, Stockholm, Sweden, July 2012.

[Bt500] ITU-R Rec. BT.500-14, Methodologies for the subjective assessment of the quality of television images, ITU: Geneva, Switzerland, 2019.

[Che10] K. Chen, C. Chang, C. Wu, Y. Chang and C. Lei, Quadrant of euphoria: a crowdsourcing platform for QoE assessment, IEEE Network, vol. 24, no. 2, pp. 28-35, March-April 2010, doi: 10.1109/MNET.2010.5430141.

[Jai13] A. K. Jain, C. Bal and T. Q. Nguyen, Tally: A web-based subjective testing tool, 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX), 2013, pp. 128-129, doi: 10.1109/QoMEX.2013.6603224.

[P910] ITU-T Rec. P910, Subjective video quality assessment methods for multimedia applications, ITU: Geneva, Switzerland, 2021.

[P911] ITU-T Rec. P.911, Subjective audiovisual quality assessment methods for multimedia applications, ITU: Geneva, Switzerland, 1998.

[P913] ITU-T Rec. P.913, Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment, ITU: Geneva, Switzerland, 2021.

[Rai13] B. Rainer, M. Waltl and C. Timmerer, A web based subjective evaluation platform, 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX), 2013, pp. 24-25, doi: 10.1109/QoMEX.2013.6603196.

[Ros03] T.D. Ross, Accurate confidence intervals for binomial proportion and Poisson rate estimation, Computers in Biology and Medicine, vol. 33, Issue 6, 2003, pp. 509-531, ISSN 0010-4825, doi: 10.1016/S0010-4825(03)00019-2.

[Uhr20a] M. Uhrina, A. Holesova, Development of web-based crowdsourcing framework used for video quality assessment, 2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA), 2020, pp. 718-723, doi: 10.1109/ICETA51985.2020.9379172.

[Uhr20b] M. Uhrina, J. Bienik, T. Mizdos, QoE on H.264 and H.265: Crowdsourcing versus Laboratory Testing, 2020 30th International Conference Radioelektronika (RADIOELEKTRONIKA), 2020, doi: 10.1109/RADIOELEKTRONIKA49387.2020.9092424.