Multimed Tools Appl
DOI 10.1007/s11042-016-4195-3

Software package for measurement of quality indicators working in no-reference model

Jakub Nawała[1] · **Mikołaj Leszczuk**[1] · **Michał Zajdel**[1] · **Remigiusz Baran**[2]

Received: 29 July 2016 / Revised: 4 November 2016 / Accepted: 21 November 2016

Abstract The key objective of No-Reference (NR) visual metrics (indicators) is to predict the end-user experience concerning remotely delivered video content. Rapidly increasing demand for easily accessible, high quality video material makes it crucial for service providers to test the user experience without the need for comparison with reference material. In this paper, we present a versatile measurement system and describe various optimisation strategies utilised to reach real-time operation. Furthermore, several calculation automation scripts are described, along with a dedicated graphical user interface, which gives a more comprehensive insight into the presented system. On top of that, we show the results of crowd-sourcing experiments used to estimate subjective threshold values for quality indicators. Additionally, integration with the IMCOP system is introduced.

Keywords QoE · QoS · OTT · No-Reference · Quality assessment · Crowd-sourcing · Content discovery · Metadata enrichment · IMCOP

The work was co-financed by The Polish National Centre for Research and Development (NCBR), as a part of the EUREKA Projects no. C 2013/1-5/MITSU/2/2014, & E! II/PL-IL/10/02A/2012 IMCOP & E! II/PL-IL/10/03A/2012 IMCOP.

✉ Jakub Nawała
jakub.tadeusz.nawala@gmail.com

Mikołaj Leszczuk
leszczuk@agh.edu.pl

Michał Zajdel
em.zajdel@gmail.com

Remigiusz Baran
r.baran@tu.kielce.pl

[1] AGH University of Science and Technology, Al. Mickiewicza 30, 30059 Kraków, Poland

[2] Kielce University of Technology, Al. 1000-lecia P.P. 7, 25541 Kielce, Poland

Published online: 13 December 2016

1 Introduction

Providing not only a high level of traditional Quality of Service (QoS), but also Quality of Experience (QoE) is a real challenge for ISPs (Internet Service Providers), audiovisual service providers, broadcasters and new Over-The-Top (OTT) service providers. Therefore, objective audiovisual data metrics are often carried out in order to monitor, troubleshoot, analyse and establish patterns of content applications working in real-time or offline scenarios. Since 2000, the work bound with the concept of QoE, in the context of different applications, has gained momentum and achieved business recognition.

A number of researchers focus on different ways to assess the quality of vision applications, taking into account additional information used in the evaluation process. Usually, two main approaches (metrics classes) are distinguished. The first approach is called Full-Reference (FR), and assumes unlimited access to the original (reference) video sequences. FR metrics are usually the most accurate at the expense of higher computational effort. The second class is commonly referred to as a No-Reference (NR) approach and is based on the quality assessment without knowledge of the original material. Due to the missing original signal, NR metrics may be less accurate than their FR counterparts, but tend to provide much better computational efficiency.

In this paper, we present a software package to measure quality indicators, operating in a difficult NR model. This software package is the realisation of a previously developed concept of monitoring the quality of vision, by Key Performance Indicators (KPI) [19]. The idea proposed here goes by the name: Monitoring Of Audio Visual quality by Key Performance Indicators (MOAVI). MOAVI artefacts (or KPIs) are divided into four categories, depending on their origin: a category of capturing, processing, transmission and display. The MOAVI based application is able to isolate and improve incident investigation, aid algorithm configuration, extend the periods to monitor and ensure better prediction of QoE.

Most models of quality are based on the measurement of typical artefacts/KPIs, such as blur, blockiness or jerkiness, and produce MOS (Mean Opinion Score) forecasts. Therefore, many of the algorithms generating an expected value of MOS use a blend of blur, blockiness and jerkiness metrics. Weighting between each KPI can be a simple mathematical function. However, if one KPI is not correct, the global result of prediction is completely wrong. Other KPIs – such as exposure, noise, block-loss, freezing, slicing, etc. – are usually not taken into account in prognosis of the MOS [18].

ITU-T has been working on a similar noise measurement model for many years [7], but only for the FR and with the Reduced Reference (RR) approach. The history of ITU-T recommendations for image quality metrics is presented in Table 1. Table 2 shows the synthesis of a set of standard indicators that are based on video signals [18]. As can be seen from both tables, there are no achievements for the NR approach.

Although not standardised, NR video quality assessment methods do exist. Zhu et al. presented in [29] model based on discrete cosine transform (DCT) and non-linear sequence-level features to subjective scores mapping by the usage of trained multilayer neural network. Authors of [29] used experimental results to show that NR metrics can compete with their FR and RR counterparts. However, due to its nature, the NR approach is both distortion specific and data driven, as compared to the more universal FR algorithms. This conclusion is not surprising, considering the fact that authors focused solely on the H.264/AVC compression as a fundamental source of distortions. On the other hand, findings shown in [20] suggest the possibility to introduce a data independent NR solution. Li, Guo and Lu use spatiotemporal 3D-DCT to extract features both in space and time. This

Table 1 History of ITU-T Recommendations (on the basis of: [18])

Model Type	Format	Rec.	Year
FR	SD	J.144 [8]	2004
FR	QCIF–VGA	J.247 [10]	2008
RR	QCIF–VGA	J.246 [9]	2008
FR	SD	J.144 [8]	2004
RR	SD	J.249 [11]	2010
FR	HD	J.341 [12]	2011
RR	HD	J.342 [13]	2011
Bit-stream	VGA–HD	P.1202 [14]	2013
Hybrid	VGA–HD	J.343 [15]	2014

information is further used to calculate a small set of parameters, which after temporal pooling for the entire sequence, get mapped to subjective scores. Thanks to thorough training and testing on various databases, authors of [20] verified data independence of their solution. Nonetheless, the best results were obtained for sequences distorted with only a single artefact source, making this solution not globally applicable.

It is worth mentioning that both [29] and [20] use the luminance channel solely. This concept is also applied in presented work due to a higher human visual system (HVS) sensitivity for luminance (rather than colour) changes.

Another thing to consider about the solution described in this article is the lack of temporal pooling and subjective scores mapping, what makes it difficult to directly compare our work with others. Those missing concepts remain to be implemented and tested in the near future. Nevertheless, as described in VQEG's (Video Quality Experts Group) MOAVI project [28], KPIs approach is defined to be complementary and more universal as compared to classical QoE measurement based on overall quality prediction.

The remainder of this paper is structured as follows. A general overview of software structure and quality metrics listing is given in Section 2. Section 3 presents experimental threshold values for metrics, along with a methodology used to obtain them. A detailed description of the operation of the presented software is given in Section 4, which is further divided into Subsections 1 to 5, all of which provide a comprehensive guide to the development process. Integration of quality evaluation software package with the IMCOP system is provided in Section 5. Section 6 concludes the paper.

Table 2 Synthesis of MOS models for: FR, RR and NR approaches (on the basis of: [18])

		ITU-T Model Type		
		FR	RR	NR
Resolution	HDTV	J.341 [12]	n/a	n/a
	SDTV	J.144 [8]	n/a	n/a
	VGA	J.247 [10]	J.246 [9]	n/a
	CIF	J.247 [10]	J.246 [9]	n/a
	QCIF	J.247 [10]	J.246 [9]	n/a

2 Structure

Aiming to allow easier evaluation and debugging of the software, the authors decided to design it in a modular manner. This basically means that each of the metrics may be easily detached or attached to the whole topology. Utilising such a strategy makes it possible to comfortably and efficiently modify the functionality of the package. In this way, the final shape of the application may be precisely carved to fit the desired use-case scenario.

The software consists of 15 visual metrics, which together form KPIs that could be used to model predicted quality of experience, as seen from the perspective of the end-user. The following set of metrics was developed:

1. Exposure [17],
2. Freezing,
3. Interlacing [16],
4. Spatial activity [23],
5. Temporal activity [23],
6. Letterboxing,
7. Pillarboxing,
8. Blockiness [23],
9. Noise [16],
10. Slicing [19],
11. Block-loss [19],
12. Blur [22, 23],
13. Contrast,
14. Flickering [23],
15. Blackout.

References next to the above-mentioned metrics lead to experimental set-ups providing concept verification. As an addition, one can refer to the work of Søgaard et al. [25], which uses some of those indicators to objectively measure the quality of a video sequence with variable bitrate.

It is worth mentioning that all quality indicators presented here were developed either by the authors themselves or by other members of a team which the authors are part of.

3 Investigating room for crowd-sourcing quality evaluation

This section presents a practical solution to the problem of automatic detection of low quality. It is based on a previously developed system for quality assessment (properly trained), which evaluates Blockiness, Blur, Contrast and Noise impairments in the NR model. The choice of artefacts was made by a cooperating industrial partner.

The study of the possibility of training the quality evaluation system was conducted by a crowd-sourcing test, which is the process of acquiring knowledge from a large number of (mainly on-line) subjects. The development of information technology and the high popularity of social networking led to a conclusion that the Internet has become one of the main methods for collecting and distributing information. To perform the test a dedicated website was developed. It contained a data base of images with different degrees of degradation. For the sake of test simplicity, the authors decided to use images rather than video sequences. Test participants were asked to answer questions concerning the quality of sequentially displayed images. The site has been made available on social networks and sent via e-mail

to various audiences, including subjects dealing with issues of image analysis. With these results, gathered from a diverse background, the threshold of perception of artefacts have been designated for four types of image distortions, namely: Blockiness, Blur, Contrast and Noise.

Usage of images, rather than videos, may be justified due to the nature of artefact indicators developed. All of them operate on a single-frame basis and may later be used as an input to the selected temporal pooling algorithm, yielding quality indication for the video sequence.

The test was conducted according to best-practices taken from VQEG activities and the white paper published based on QUALINET task force experience [5].

3.1 Examined artefacts and image assessment methodology

We studied the effects of four types of artefacts: Blockiness, Blur, Contrast and Noise.

Three questions were asked in the conducted test. The first related to whether the subject saw any artefact in the displayed image. The second required the subject to score images on a Mean Opinion Score (MOS) scale. And the third related to the type of distortion present. The subject could determine if the image contained any of the following impairments: Blockiness, Blur, Contrast or Noise. In case the subject did not see any artefact, they could choose the answer "none". The "other" option has been put as well. The final question was asked only to groups active in the image processing field.

The Mean Opinion Score (MOS), referred to in a previous paragraph, is a scale related to subjective, numerical indication of quality of the medium obtained after compression, decompression or transmission. MOS consists of levels from 1 to 5, where each denotes: 1 – bad quality, 2 – poor quality, 3 – average quality, 4 – good quality, and 5 – excellent image quality [6]. The crowd-sourcing experiment subject could select only one of these levels.

3.2 Crowd-sourcing process

The first step in implementing the crowd-sourcing test was to prepare images with various degrees of artefacts. Materials designed in this way were later uploaded to the website and scored by the test subjects. Eight (8) properly transformed images have been selected for the test.

Before uploading, the image size was modified in order not to exceed nine hundred pixels in the horizontal direction. Thanks to this, photos on the website could be viewed in their entirety on a fifteen-inch screen, being the most popular screen size amongst laptop users. The images were then distorted with artefacts. For each of the eight images, the following artefacts were applied: thirteen (13) levels of Blockiness artefact type, ten (10) levels of Blur artefact type, seventeen (17) levels of Contrast artefact type, and nine (9) levels of Noise artefact type. In total, a database of four hundred (400) images was compiled. Additionally, eight common set (warm-up) images were chosen to be displayed during the first run of a test. This treatment was due to the fact that during the first visit, the subject had to learn the web interface provided. As a result, the first eight scores of the test images were not taken into account when analysing the results.

The next stage of the test was to put the images on the website and allow users to start the evaluation process. Each subject could complete the test once; it was impossible to log in again using the same user name. After log in the user was presented with a warm-up sequence, followed by four hundred (400) relevant photos.

Each image to be assessed was presented in its original resolution and accompanied on the right by a panel displaying all three questions along with the username, progress bar, and the interface for moving between test images. When a subject passed to the next image, results were saved to the database.

For all the questions displayed on the page, one could select only a single answer. If the subject failed to answer all three and tried to move to the next image, a message asking to address the remaining questions was displayed.

The user could end the test at any time, either by logging off from the front end of the interface or simply leaving the web page.

3.3 Results

A total of one hundred seventy-three (173) subjects took part in the crowd-sourcing test in a single month. Forty (40) subjects simply logged in and did not participate in the evaluation process. Forty-two (42) people gave evaluation scores for less than nine images, assessing just a collection of common set images not included in the analysis of the test results. Ninety-one (91) subjects issued scores for more than eight images. On average, ten (10) scores were obtained for each image. The number of scores made it possible to separate the results of various user groups participating in the test. This kind of division allowed to carry out separate analysis for a few distinct user profiles.

Operation under the time constraint made it impossible to gather more results. Nonetheless, number of answers acquired has proven to be sufficient for further analysis.

3.4 Analysis of results

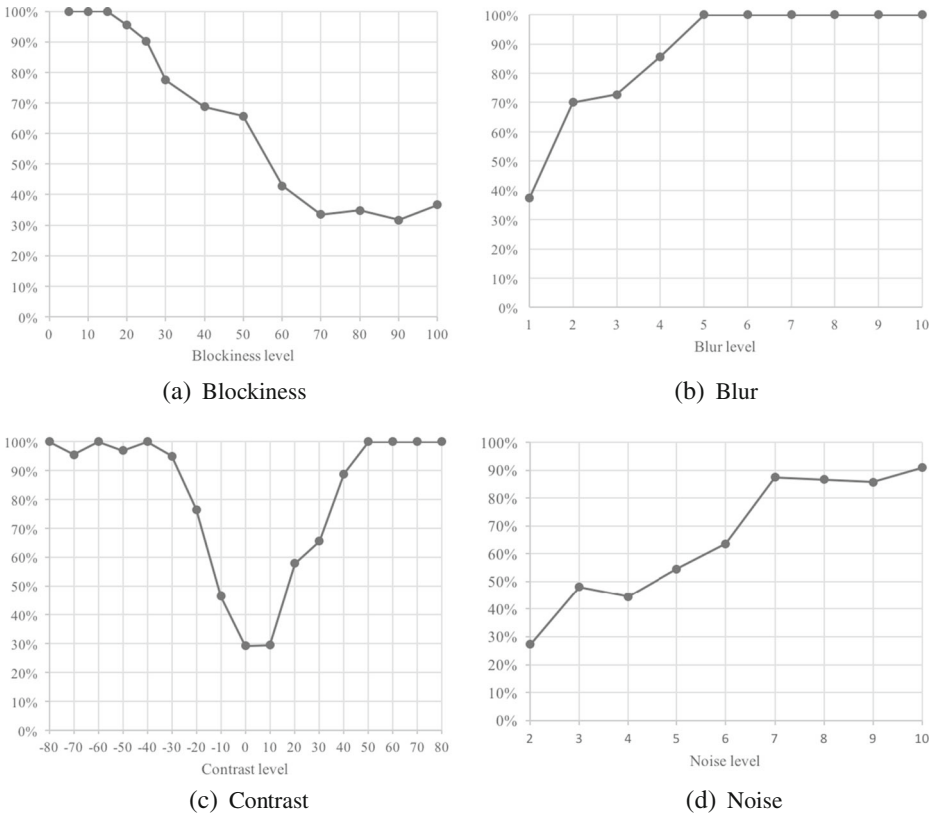
Based on the test results, artefact perception percentages were determined for all levels of each single distortion. This quantity denotes the percentage of test subjects, who properly noticed an artefact's presence. The number of scores received does not allow for a separate analysis of each image. Figure 1 presents perception percentages plotted versus quality metrics outcomes yielded by the measurement software package.

On the basis of those results, artefact perception thresholds were calculated for each type of impairment respectively. A threshold value was chosen to represent a situation when half of the test subjects saw the distortion, and the other half did not. For the Blockiness artefact type, the artefact was visible for less than half of the respondents, above the level equal to 50. For the Blur artefact type, impurities were detected for a level greater than 1. In the Contrast artefact type, degradation of an image was not detected above the level of -10 and below 20. For the Noise artefact, distortions were visible above the level equal to 4.

Designated threshold error has been estimated. The data set was divided into the training and test subsets, which was necessary to perform cross-validation of a model. The following percentages of accuracy of various types of artefacts were achieved: Blockiness – 77.09 %, Blur -87.5 %, Contrast – 75 % and Noise – 78.57 %.

The calculated threshold for the Noise artefact type did not take into account the results for one of the tested images. This was due to the inconsistency of the data obtained for different levels of impurity. Results for this single image had a significant impact on the final threshold value. Separating them from the rest of the data yielded much better model performance, which was further proven by cross-validation.

In the case of simultaneous imposition of different types of artefacts on images, of which no one is dominant, the calculation of metrics cannot be made properly because of mutual



(a) Blockiness (b) Blur

(c) Contrast (d) Noise

Fig. 1 Crowd-sourcing experiment results – artefact perception percentages

masking of the distortions. To enable an accurate assessment of a single artefact, one must first apply the appropriate compensation algorithms [4].

3.5 Summary

The main problem encountered during the research was the number of test subjects per tested image. Out of one hundred seventy-three (173) subjects, nearly half did not participate in the test, logging in or assessing common set images only. The vast majority of those who evaluated more than eight common set images failed to complete half of the test. Hence, proper examination of results was not a trivial task. It was impossible to clearly determine visible artefact thresholds or analyse the results for each group of artefacts separately. The latter difficulty arose from an insufficient number of scores for a single image in the artefact group.

Each image was almost identically rated for the Blur artefact type. The most varied ratings were obtained for the Noise distortion type. As was previously mentioned, to achieve a high threshold accuracy here, results for one of the test images had to be ruled out, significantly increasing the reliability of the final threshold level.

Despite the problems encountered during the test, it was completed successfully. High accuracy thresholds of artefact perception for specific types of impurities were received.

4 Measurement software package

As was already mentioned, the presented software package performs a remote NR quality assessment. The main goal accompanying its design and implementation was the idea to create an application that is platform-independent and does not include proprietary software. Consequently, the source code of the program was written entirely in C programming language and none of the metrics utilised any external libraries. This approach resulted in a longer development timeframe but at the same time allowed us to create a versatile, portable and stable measurement system.

4.1 Input and output interfaces

The presented software package operates within the NR model, meaning that the measurement is performed without any knowledge of the original sequence. As a consequence, input material must be analysed in pixel-by-pixel fashion. This in turn imposes the necessity of decompression of the video file or stream, before any computation may be performed. Due to the fact that the algorithms used operate solely on the luminance channel (Y), YUV420p format is utilised to store the input files for the application. It makes it possible to save memory by omitting part of the information related to colours, further referred to as chrominance channels. Data stored in this manner incorporates complete information about the grayscale representation, but allocates only one value of chrominance channels (U and V) for each 4 pixels of the original material. An additional advantage of using the previously mentioned format is contiguous alignment of image data, which constitutes a very basic optimisation strategy. Most hardware platforms perform best when operated on linearly stored information. Reading out sequentially ordered memory blocks yields the lowest possible access times and thus leaves more headroom for the actual computation.

In addition to the uncompressed video sequence, the application also expects the parameters describing width, height and number of frames per second of the tested material. Supplementary input arguments result from the specification of YUV420p format. It does not contain any header for storing detailed information about the included material. In most cases, however, this is not a problem, since data used for processing exists in compressed form, which along with the video material, contains all the essential information.

The application generates a detailed report concerning each frame of the input material. Alongside frame number, one can also see the result of each single metric. Presentation of the output information is twofold:

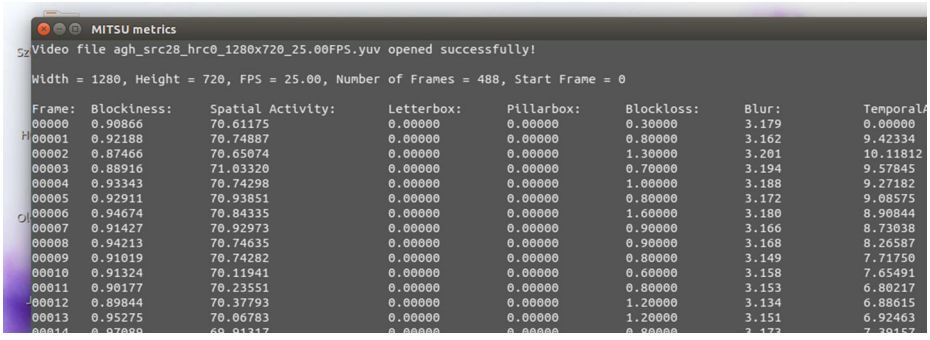


Fig. 2 Exemplary standard output generated by QoE software package

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Frame:	Blockiness:	SA:	Letterbox:	Pillarbox:	Blockloss:	Blur:	TA:	Blackout:	Freezing:	Exposure(bri):	Contrast:	Interlace:	Noise:	Slice:	Flickering:
2	0	0.90866	70.61175	0	0	0.3	3.179	0	0	0	154	39.37412	0.04054	3.6521	3.604	0
3	1	0.92188	70.74887	0	0	0.8	3.162	9.42334	0	0	154	39.2981	0.04021	3.6534	3.459	-1
4	2	0.87466	70.65074	0	0	1.3	3.201	10.11812	0	0	155	39.27022	0.04153	4.2078	3.066	-1
5	3	0.88916	71.0332	0	0	0.7	3.194	9.57845	0	0	155	39.27171	0.04038	4.1942	2.99	-1
6	4	0.93343	70.74298	0	0	1	3.188	9.27182	0	0	155	39.13484	0.03865	2.6308	3.279	-1
7	5	0.92911	70.93851	0	0	0.8	3.172	9.08575	0	0	155	38.99756	0.03804	4.1895	3.553	-1
8	6	0.94674	70.84335	0	0	1.6	3.18	8.90844	0	0	155	38.80327	0.03764	4.1612	3.283	-1
9	7	0.91427	70.92973	0	0	0.9	3.166	8.73038	0	0	155	38.57045	0.03818	4.1616	2.821	-1
10	8	0.94213	70.74635	0	0	0.9	3.168	8.26587	0	0	154	38.35791	0.03738	4.175	2.864	0.76707
11	9	0.91019	70.74282	0	0	0.8	3.149	7.7175	0	0	153	38.14938	0.03762	4.1742	3.305	-1
12	10	0.91324	70.11941	0	0	0.6	3.158	7.65491	0	0	154	37.88908	0.04033	3.6368	3.658	-1
13	11	0.90177	70.23551	0	0	0.8	3.153	6.80217	0	0	152	37.82033	0.04059	3.6457	4.086	-1
14	12	0.89844	70.37793	0	0	1.2	3.134	6.88615	0	0	153	37.79031	0.03988	4.1575	3.956	-1
15	13	0.95275	70.06783	0	0	1.2	3.151	6.92463	0	0	154	37.6995	0.03715	4.6124	3.495	-1
16	14	0.97089	69.91317	0	0	0.8	3.173	7.39157	0	0	154	37.69124	0.03807	4.1687	3.523	-1
17	15	0.91599	69.91529	0	0	0.3	3.166	7.65922	0	0	154	37.61771	0.03674	4.6144	3.589	-1
18	16	0.96421	69.52181	0	0	0.9	3.218	8.04491	0	0	155	37.51045	0.03854	3.6791	3.249	0.73351
19	17	0.92153	70.97548	0	0	0.7	3.172	8.06154	0	0	154	37.37474	0.03491	3.6485	3.048	-1
20	18	0.95315	71.73471	0	0	0.4	3.196	9.01084	0	0	153	37.20493	0.03561	4.0865	3.237	-1
21	19	0.9044	72.52641	0	0	1.3	3.177	8.11429	0	0	154	37.09685	0.03137	4.0733	3.02	-1
22	20	0.95984	71.90306	0	0	1.6	3.196	8.01805	0	0	154	36.97626	0.03094	4.0952	3.515	-1

Fig. 3 Exemplary CSV file generated by QoE software package

– Standard output (stdout) – results get printed in the terminal session used to invoke the software (see Fig. 2),
– CSV (Comma-Separated Values) file – outcome stored in the form suitable for usage in spreadsheets and automated calculation scripts (see Fig. 3).

4.2 Planned and applied optimisation schemes

The careful reader may notice that operations performed on uncompressed video sequences require large memory bandwidth, as well as high computational power. This kind of restriction becomes especially important when operating in real-time or nearly real-time scenarios. Figure 4 shows the relative execution times for each metric, when processing video with a resolution of 1920×1080 pixels. Average computation time for such a material oscillates around 119 ms. At this point it is worth mentioning that this test was conducted using a single thread version of the application on the machine featuring an *Intel Core i7 CPU 950@.3.07 GHz x 8* processor.

The average processing time indicates the necessity of further optimisation if one requires real-time execution of the software. Assuming the video sequence gets refreshed

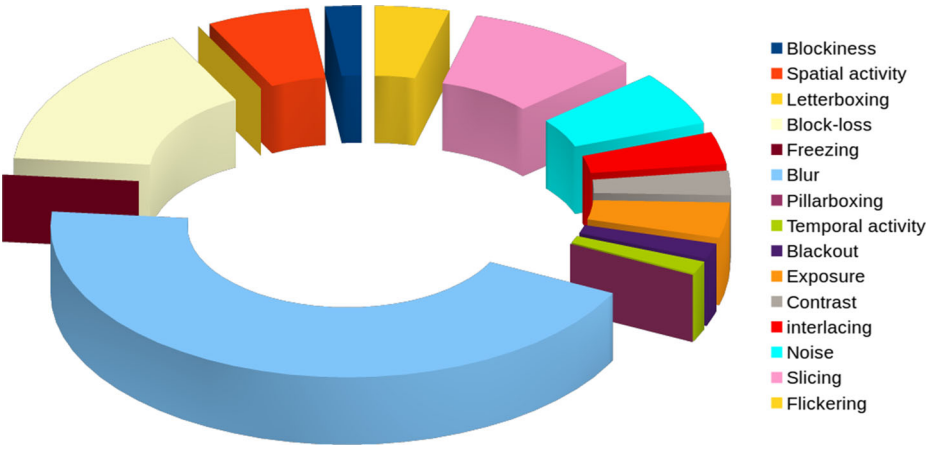


Fig. 4 Relative metrics execution time for a single 1920×1080 image frame

30 times per second, fetching image data and preforming computations must not exceed 33 ms. Should dropping any of the provided indicators prove impossible, another optimisation technique would be to utilize a multiprocessor and thus, multithread architecture of contemporary platforms. Performing the test once again - this time employing a multithread version of the application - allowed to reduce the time needed for calculations to 59 ms. Even though it does not guarantee real-time operation, there is still more optimisation strategies to be implemented.

If, on the other hand, eliminating some of the indicators is possible, ruling out Blur and Block-loss metrics yields an execution time below 33 ms (provided that multithread version of the software is used).

It is worth mentioning that many image processing algorithms use precisely defined, and more importantly, finite set of operations, which may be performed on the image. As a consequence, once processed, an image or parameter may be stored and used again in other metrics. This strategy works best if the amount of data to be stored does not exceed some threshold value, which defines the balance point for a trade-off between memory usage and computational complexity.

Yet another possible optimisation scenario is to move as much computations as possible into the domain of integer numbers. This is justified only if one plans to use the central processing unit (CPU) exclusively. Due to its internal topology, it performs best when used with this kind of data.

All optimisation methods described operate in the software layer of the system design. Apart from those, one can always try to port the code to another hardware platform like the GPU (Graphics Processing Unit) or FGPA (Field-Programmable Gate Array). Both solutions allow to massively parallelise the execution and thus reduce the time needed for processing. However, advantageous features of both these solutions come at a price of thorough source code rebuilding that is necessary to gain maximum performance boost.

4.3 Additional scripts

As an addition, several automated calculation scripts are provided. In order to achieve a high level of portability, all of the scripts were written both for Unix-like and Microsoft Windows systems. Obtaining this extent of versatility required the creation of two separate implementations. One written in Bash (Linux, Mac OS) and one in Batch (Windows). Utilisation of FFmpeg tools allowed to reduce the input interface to a single parameter, namely the path to video sequence or folder containing video materials to process. Automation scripts are based on the assumption that all input data is stored in the form including detailed information about its content. This mechanisation allows one to seamlessly apply the presented measurement techniques to a large set of input data, be it images or videos.

4.4 Versions

One of the most important aspects accompanying the development process was the assumption that if possible, the application should be platform independent. As a result, the software package was released for all of the most popular operating systems: Linux, Mac OS and Windows. Though multi-sided, the software's implementation remains consistent, meaning that a single source code may be used to compile into all supported binaries. Minute changes

in the configuration file is enough to quickly switch between the desired OS (operating system) and architecture type (32 or 64-bit).

The described software is provided free of charge (for non-commercial usage) and may be downloaded from the WWW web page [26].

4.5 Graphical user interface

Keeping in mind that presentation of the software is of key importance, the authors decided to additionally implement a graphical user interface. Its main advantage is the possibility of simultaneous observation of results and the currently processed video sequence. Figure 5 shows an example of the described software. The graphical version of the measurement system is capable of processing any video stream, provided its content is made available in a shared memory. Thus, it is necessary to introduce a thin integration layer decompressing video stream and uploading raw frames into memory shared with measurement application. This kind of solution was developed and tested inside the MITSU project [21]. Merging transcoding software with the measurement system allowed to create dynamically changing video delivery architecture that aimed to maximise user experience in terms of QoE.

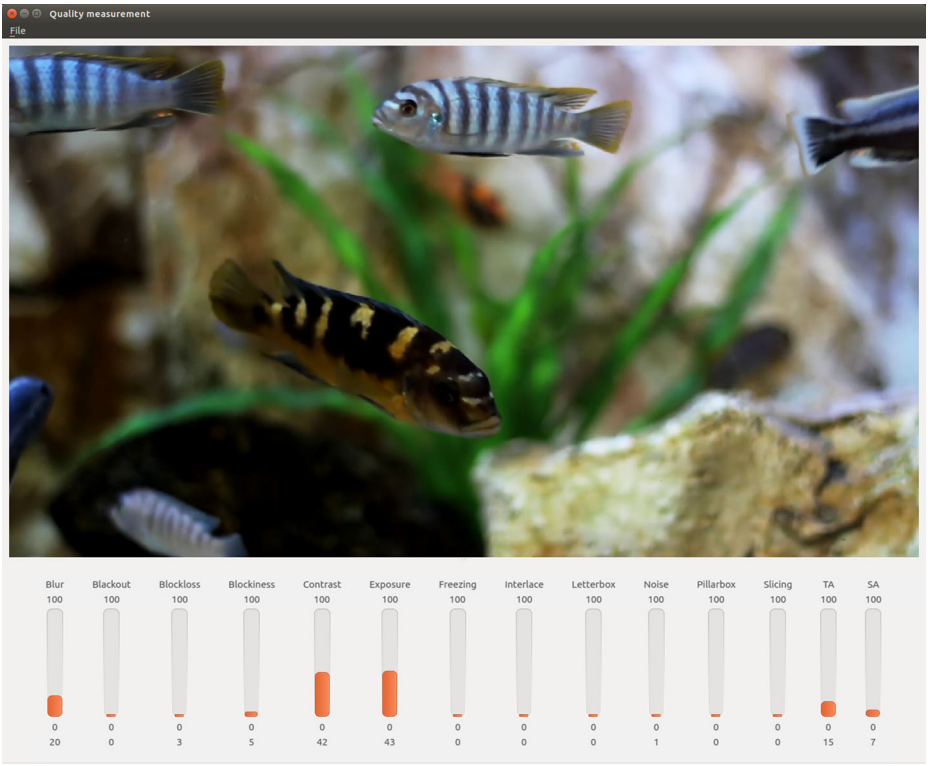


Fig. 5 The graphical user interface of application measuring QoE

5 Integration with IMCOP architecture

The IMCOP project - an "Intelligent Multimedia System for Web and IPTV Archiving. Digital Analysis and Documentation of Multimedia Content", is a joint Polish-Israeli R&D project realised by a consortium consisting of four partners. In general, IMCOP's objectives are twofold and are referred to as: (i) multimedia data analysis and content discovery on one side and (ii) data aggregation, content related binding (finding and assigning content related connections between data) and delivery on the other [3]. An overall IMCOP platform architecture is illustrated in Fig. 6.

Data analysis is performed in order to enrich the data (mainly images and video sequences) by extracting their features and classify them according to a given criteria. Components of the IMCOP system dedicated to carry out the above analysis are known as the Metadata Enhancement Services (MES), which, in fact, are REST-compliant Web services in the cloud [2]. Each MES service is intended to perform a single classification task. Selected specialized tasks of IMCOP's MES services are as follows:

- detection as well as facial recognition,
- head counting,
- bokeh effect detection,
- text detection and recognition,
- nudity detection,
- sky/landscape detection,
- detection of architectural scenes (scenes containing buildings, monuments or other kinds of artificial structures), etc.

In addition to the ones given above, IMCOP also incorporates less specialized types of MES services, which are dedicated, for example, to extracted selected low level features of analysed data, such as e.g. SURF features [1], Shape Context histogram, MPEG-7 visual descriptors, coefficients of Piecewise-linear transforms [24], etc. Services designated to perform visual quality evaluation are also of great importance to the IMCOP system. In general,

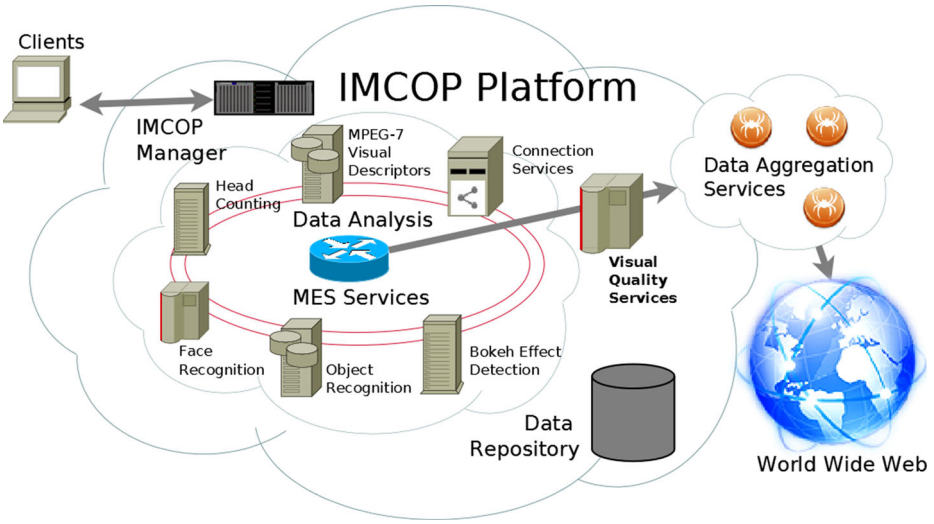


Fig. 6 IMCOP system overall architecture

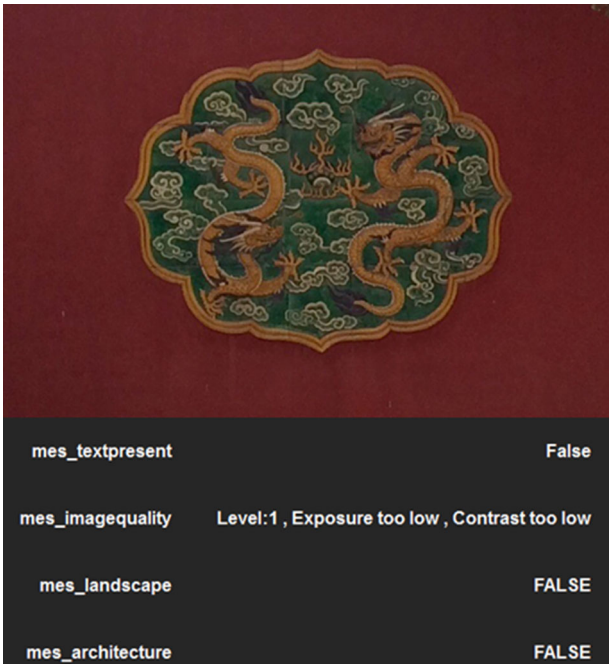


Fig. 7 An example of applying quality metrics to label images

they are used to filter out the low quality multimedia data and exclude them from further processing. The way they are used to model predicted quality of experience in the IMCOP system is as follows:

- quality of multimedia data is classified into three categories, known as quality levels 0, 1 and 2, where level 0 means data of very low, unacceptable quality and level 2, on the contrary, data of very high quality,
- data is classified into quality level 2 when at most two metrics fail (fall outside the min/max range given in [26]),
- data is classified into quality level 1 (the category of low and medium quality) when three or four metrics fail,
- when more than four metrics fail, data is classified at quality level 0.

Metadata Enhancement Services can be, inter alia, used just to label (tag) multimedia data. Examples of labels given by selected IMCOP's MES services to a chosen image from the VIME Flickr dataset [27] are depicted in Fig. 7.

6 Conclusions

Quality indicators have been successfully developed as a result of the work. All together constitute a single, universal and multi-platform measurement system, which runs entirely on the receiving side. This ability makes it especially suitable for content providers operating on a massive scale. The opportunity to remotely sense quality of experience at each user-node guarantees better system control and gives solid input for various resource utili-

sation algorithms. Moreover, measurement performed on two ends of the system allows one to quantitatively measure its impact on the content being transmitted.

A related point to consider is the fact that the software provides information regarding all indicators separately. Establishing trustworthy mapping between those KPIs and final subjective quality is a challenging task requiring more experimental data. As such, it remains to be implemented and defines the scope of the current research. A direct consequence of this shortage is the difficulty in objective assessment of algorithm performance that would allow to compare it with other state-of-the-art achievements.

On the other hand, the lack of consistent KPI to MOS mapping may also be regarded as advantageous. Due to clear and comprehensive presentation of results, the user alone may choose the meaning and importance of certain metrics, making it possible to introduce a customised quality evaluation process. Both presented use-cases [3, 25] of the software package utilised this property to aid their operation and, at the same time, prove its usability both for end-products and experimental set-ups.

References

1. Baran R, Ruść T, Rychlik M (2014) A smart camera for traffic surveillance. Springer International Publishing, Cham, pp 1–15. doi:10.1007/978-3-319-07569-3_1
2. Baran R, Zeja A (2015) The imcop system for data enrichment and content discovery and delivery. In: 2015 International conference on computational science and computational intelligence (CSCI), pp 143–146. doi:10.1109/CSCI.2015.137
3. Baran R, Zeja A, Slusarczyk P (2015) An overview of the IMCOP system architecture with selected intelligent utilities emphasized. Springer International Publishing, Cham, pp 3–17. doi:10.1007/978-3-319-26404-2_1
4. Głowacz A, Grega M, Gwiazda P, Janowski L, Leszczuk M, Romaniak P, Romano SP (2010) Automated qualitative assessment of multi-modal distortions in digital images based on glz. Annals of Telecommunications - Annales des Té,lécommunications 65(1):3–17. doi:10.1007/s12243-009-0146-6
5. Hoßfeld T, Hirth M, Redi J, Mazza F, Korshunov P, Naderi B, Seufert M, Gardlo B, Egger S, Keimel C (2014) Best practices and recommendations for crowdsourced qoe - lessons learned from the qualinet task force "crowdsourcing". Lessons learned from the qualinet task force "Crowdsourcing" COST action IC1003 European network on quality of experience in multimedia systems and services (QUALINET). https://hal.archives-ouvertes.fr/hal-01078761
6. International Telecommunication Union (1996) ITU-T P.800, Methods for subjective determination of transmission quality. http://www.itu.int/rec/T-REC-P.800-199608-I
7. International Telecommunication Union (1996) ITU-T P.930, Principles of a reference impairment system for video. http://www.itu.int/rec/T-REC-P.930-199608-I
8. International Telecommunication Union (2004) ITU-T J.144, Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference. http://www.itu.int/rec/T-REC-J.144-200403-I
9. International Telecommunication Union (2008) ITU-T J.246, Perceptual isual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference. http://www.itu.int/rec/T-REC-J.246-200808-I
10. International Telecommunication Union (2008) ITU-T J.247, Objective perceptual multimedia video quality measurement in the presence of a full reference. http://www.itu.int/rec/T-REC-J.247-200808-I
11. International Telecommunication Union (2010) ITU-T J.249, Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference. http://www.itu.int/rec/T-REC-J.249-201001-I
12. International Telecommunication Union (2011) ITU-T J.341, Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference. http://www.itu.int/rec/T-REC-J.341-201101-I

13. International Telecommunication Union (2011) ITU-T J.342, Objective multimedia video quality measurement of HDTV for digital cable television in the presence of a reduced reference signal. http://www.itu.int/rec/T-REC-J.342-201104-I
14. International Telecommunication Union (2013) ITU-T P.1202, Parametric non-intrusive bitstream assessment of video media streaming quality. https://www.itu.int/rec/T-REC-P.1202
15. International Telecommunication Union (2014) ITU-T J.343, Hybrid perceptual bitstream models for objective video quality measurements. https://www.itu.int/rec/T-REC-J.343
16. Janowski L, Papir Z (2009) Modeling subjective tests of quality of experience with a generalized linear model. In: International workshop on quality of multimedia experience, 2009. QoMEx 2009, pp 35–40. doi:10.1109/QOMEX.2009.5246979
17. Leszczuk M (2011) Multimedia communications, services and security. In: 4th international conference, MCSS 2011, krakow, Poland, June 2–3, 2011. Proceedings, chap. assessing task-based video quality — a journey from subjective psycho-physical experiments to objective quality models. Springer, Berlin, pp 91–99. doi:10.1007/978-3-642-21512-4_11
18. Leszczuk M, Hanusiak M, Blanco I, Dziech A, Derkacz J, Wyckens E, Borer S (2014) Key indicators for monitoring of audiovisual quality. In: Signal processing and communications applications conference (SIU), 2014 22nd, pp 2301–2305. doi:10.1109/SIU.2014.6830724
19. Leszczuk M, Hanusiak M, Farias MCQ, Wyckens E, Heston G (2014) Recent developments in visual quality monitoring by key performance indicators. Multimedia Tools and Applications:1–23. doi:10.1007/s11042-014-2229-2
20. Li X, Guo Q, Lu X (2016) Spatiotemporal statistics for video quality assessment. IEEE Trans Image Process 25(7):3329–3342. doi:10.1109/TIP.2016.2568752
21. Mitsu (multimedia efficient scalable and robust delivery) (2016). http://mitsu-project.eu/
22. Mu M, Romaniak P, Mauthe A, Leszczuk M, Janowski L, Cerqueira E (2012) Framework for the integrated video quality assessment. Multimedia Tools and Applications 61(3):787–817. doi:10.1007/s11042-011-0946-3
23. Romaniak P, Janowski L, Leszczuk M, Papir Z (2012) Perceptual quality assessment for h.264/avc compression. In: 2012 IEEE consumer communications and networking conference (CCNC), pp 597–602. doi:10.1109/CCNC.2012.6181021
24. Slusarczyk P, Baran R (2014) Piecewise-linear subband coding scheme for fast image decomposition. Multimedia Tools and Applications:1–18. doi:10.1007/s11042-014-2173-1
25. Søgaard J, Tavakoli S, Brunnström K, García N (2016) Subjective analysis and objective characterization of adaptive bitrate videos. In: IS&T international symposium on electronic imaging 2016: image quality and system performance XIII
26. Video quality (2016). http://vq.kt.agh.edu.pl
27. Vime flickr dataset (2015). https://www.flickr.com/groups/vime/
28. Vqeg moavi project (2016). http://www.its.bldrdoc.gov/vqeg/projects/moavi/moavi.aspx
29. Zhu K, Li C, Asari V, Saupe D (2015) No-reference video quality assessment based on artifact measurement and statistical analysis. IEEE Trans Circuits Syst Video Technol 25(4):533–546. doi:10.1109/TCSVT.2014.2363737

Jakub Nawała is a student and beginner researcher at AGH University of Science and Technology. Nawała works at Department of Electronics and Department of Telecommunications. His research concerns computer vision, video quality assessment and embedded systems programming.

Mikołaj Leszczuk Ph.D. He started his professional career in 1996 at COMARCH SA as manager of the Multimedia Technology Department, and then at COMARCH Multimedia as the CEO. Since 1999 has been employed at the AGH Department of Telecommunications. In 2000 he moved to Spain for a four-month scholarship at the Universidad Carlos III de Madrid. After returning to Poland, he was employed at the Department of Telecommunications as a research and teaching assistant, and in 2006, he successfully defended his doctoral dissertation as an assistant professor. His current research interests are focused on multimedia data analysis and processing systems, with particular emphasis on Quality of Experience. He (co-)authored approximately 130 scientific publications of which 23 are publications in journals of the JCR database. He has been teaching at undergraduate and graduate levels. He has cosupervised 1 Ph.D. student and supervised (promoted) approximately 40 MSc students of various nationalities. He has participated more than 20 major research projects, including FP4, FP5, FP6, FP7, Horizon 2020, OPIE, Culture 2000, PHARE, eContent+, and Eureka!. Between 2009 and 2014, he was the administrator of the major international INDECT research project, dealing with solutions for intelligent surveillance and automatic detection of suspicious behaviour and violence in urban environments. He is a member of VQEG (Video Quality Experts Group, board member), IEEE (Institute of Electrical and Electronics Engineers), and GAMA (Gateway to Archives of Media Art). The latter organization collaborates with the VQiPS (Video Quality in Public Safety) working group. More information: http://www.linkedin.com/in/miklesz

Michał Zajdel - electronics and telecommunications engineer from AGH University of Science and Technology in Cracow.

Remigiusz Baran was awarded the M.Sc. in Electrical Engineering from the Faculty of Electrical and Control Engineering, Kielce University of Technology in 1993, and the Ph.D. in Telecommunications from the Faculty of Electrical, Control, Electronic and Computer Engineering, AGH University of Science and Technology in Krakow in 2004. He is currently working as an Assistant Professor at the Kielce University of Technology. He is the author or co-author of over 50 publications in the field of digital signal processing, focusing on image compression and feature extraction. The main areas of his academic interest are feature- and appearance-based object detection and recognition techniques, and microprocessor technology and embedded systems. Apart of his scientific activity he is also an academic teacher. He has promoted approximately 60 MSc students as well at undergraduate as graduate levels. He also serves as a reviewer of international journals and conferences. Dr. Baran has participated numerous international and national (Polish) research projects including INDECT, OASIS Archive, Calibrate, INWAS, INSIGMA, TAPAS. At present he is the Project Manager of the second joint Polish-Israeli R&D project IMCOP - Intelligent Multimedia System for Web and IPTV Archiving. Digital Analysis and Documentation of Multimedia Content.