Optimising task-based video quality

A journey from subjective psychophysical experiments to objective quality optimisation

Mikołaj Leszczuk

Published online: 13 July 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Development of techniques for assessing video quality is reviewed. Examples have been provided on the quality of video applications ranging from popular entertainment to new trends such as applications in wide-reaching public systems, not just those used by security forces but also for medical purposes. In particular, two typical usages of task-based video: surveillance video for accurate licence plate recognition, and medical video for credible diagnosis prior to bronchoscopic surgery were introduced by the author. The problem of task-based video quality assessment starting from subjective psychophysiological experiments to objective quality models is discussed. Example test results and models are provided alongside to the descriptions. Finally, a quality optimisation approach, driven by recognition rates is presented.

Keywords Video • Quality • Experiments • Models • Optimisation

1 Introduction

The storage and transmission of video is used for many applications outside of the entertainment sector; generally, this class of video is used to perform a specific task (task-based video). Examples of these applications include public safety including surveillance, medical services, remote command and control, and sign language.

Development efforts reveal a significant potential behind platforms allowing access to digital recordings of surveillance or medical video sequences. Video compression and transmission are the most widespread problems in those platforms. Surveillance and medical applications add a new dimension, as lossy compression techniques need to be both resource-effective and credible from the point of view of practitioners in the field of public safety and medical services.

M. Leszczuk (✉)

AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland
e-mail: leszczuk@agh.edu.pl

Anyone who has experienced artefacts or freezing play while watching a film or live sporting event on TV is familiar with the frustration accompanying sudden quality degradation at a key moment. Video services with blurred images may have far more severe consequences for medical or law enforcement practitioners. Therefore it is crucial to measure and ultimately optimise task-based video quality.

This paper introduces two typical usages of task-based video (Section 2): surveillance video used for accurate licence plate recognition, and medical video used for credible diagnosis prior to bronchoscopic surgery. The remainder of the article introduces the field of task-based video quality assessment from subjective psychophysical experiments (Section 3) to objective quality models (Section 4). Example test results and models are provided alongside the descriptions. Section 5 presents the most important contribution of the paper: a quality optimisation approach, driven by recognition rates. Finally Section 6 outlines the conclusions and plans for further work, including standardisation.

2 Use cases

This section introduces two typical usages of task-based video: surveillance video used for accurate licence plate recognition, and medical video used for credible diagnosis task prior to bronchoscopic surgery.

Accurate licence plate recognition task Recognizing the growing importance of video in delivering a range of public safety services, let us consider a licence plate recognition task based on video streaming in constrained networking conditions. Video technology should allow users to perform the required function successfully. The paper presents people's ability to recognise car registration numbers in video material recorded using a CCTV camera and compressed with the H.264/AVC codec. An example frame from the licence plate recognition task is shown in Fig. 1. The usage case is presented in literature [13].

Fig. 1 Example frame from the licence plate recognition task

Credible bronchoscopic diagnosis task The presented task targets video bronchoscopy, a type of recording made during surgery. In such videos the image can remain almost motionless for prolonged periods of time. This predisposes such recordings to compression. However, it is possible to use just those motion images in which lossy compression has not caused any distortion visible to the physicians [1, 9, 15, 16]. The paper describes the degree of compression which introduces quality impairment visible to the physicians. An example frame from the bronchoscopic diagnosis task is shown in Fig. 2. The usage case is presented in literature [2, 9].

Video streaming services still face the problem of limited bandwidth access. While for wired connections bandwidth is generally available in the order of megabits, higher bit rates are not particularly common for wireless links. This poses problems for mobile users who cannot expect a stable high bandwidth.

Therefore a solution for streaming video services across such access connections is transcoding of video streams. The result is transcoding bit-rate (and quality) scaling to personalise the stream sent according to the current parameters of the access link. Scaling video sequence quality may be provided in compression, space and time. Scaling of compression usually involves operating the codec Quantisation Parameter (QP). Scaling of space means reducing the effective image resolution resulting in increased granularity when attempts are made to restore the original content on the screen. Scaling of time amounts to rejection of frames, i.e. reducing the number of frames per second (FPS) sent. However, frame rates are commonly kept intact as their deterioration does not necessarily result in bit-rate savings due to inter-frame coding [8].

The abovementioned scaling methods inevitably lead to lower perceived quality of end-user services (Quality of Experience, QoE). Therefore the scaling process should be monitored for QoE levels. This makes it possible to not only control but also maximise QoE levels depending on the prevailing transmission conditions. In the event of failure to achieve a satisfactory QoE level, the operator may intentionally interrupt the service, which may help preserve network resources for other users.

Fig. 2 Example frame from the bronchoscopic diagnosis task (source: literature [12])

3 Procedures for subjective psychophysical experiments

To develop accurate objective measurements (models) for video quality, subjective experiments must be performed. The ITU-T¹ P.910 Recommendation “Subjective video quality assessment methods for multimedia applications” (1999) [6] addresses the methodology for performing subjective tests in a rigorous manner.

However, such methods are currently only targeted at entertainment video. In task-based applications, video is used to recognise objects, people or events. Therefore the existing methods, developed to assess a person’s perception of quality, are not appropriate for task-based video.

The QoE concept for video content used for entertainment differs considerably from the quality of video used for recognition tasks. This is because in the latter case subjective user satisfaction is improved by achieving a given functionality (event detection, object recognition). Additionally, the quality of video used by a human observer for recognitions tasks is considerably different from objective video quality used in computer processing (Computer Vision).

Task-based videos require a special framework appropriate to the video’s function—i.e. its use for recognition tasks rather than entertainment. Once the framework is in place, methods should be developed to measure the usefulness (the ability to perform a specific task) of the reduced quality video rather than its entertainment value.

Issues of quality measurements for task-based video are partially addressed in the ITU-T P.912 Recommendation “Subjective video quality assessment methods for recognition tasks” (2008) [7]. This Recommendation introduces basic definitions, methods of testing and ways of conducting psychophysical experiments (e.g. Multiple Choice Method, Single Answer Method, and Timed Task Method), as well as the distinction between Real-Time- and Viewer-Controlled Viewing scenarios. While these concepts have been introduced specifically for task-based video applications in ITU-T P.912, more research is necessary to validate the methods and refine the data analysis methods.

Section 7.3 of ITU-T P.912 (“Subjects”) says that, “Subjects who are experts in the application field of the target recognition video should be used”. Nevertheless, based on literature [11], both expert and non-expert subjects can maintain the same response characteristics.

3.1 Psychophysical experiment for accurate licence plate recognition

A subjective experiment was carried out in order to perform the analysis. A psychophysical evaluation of the video sequences scaled (in the compression or spatial domain) at various bit-rates was performed. The aim of the subjective experiment was to gather the results of human recognition capabilities. Thirty non-expert testers rated video sequences influenced by different compression parameters. ITU’s Absolute Category Rating (ACR), described in ITU-T P.800 [5], was selected as the subjective test methodology.

30 Source Reference Circuit (SRC) video sequences used in the test were recorded in a car park using a CCTV camera. In this scenario, the camera was located 50 m

¹International Telecommunication Union—Telecommunication Standardisation Sector

from the parking lot entrance in order to simulate typical video recordings. Using ten-fold optical zoom, 6.0 m × 3.5 m field of view was obtained. The camera was placed statically without changing the zoom throughout the recording time, which reduced global movement and lighting conditions to a minimum. All the video content collected in the camera was analysed and cut into 20-s shots including cars entering or leaving the car park. The licence plate was visible for a minimum 17 s in each sequence. The H.264 (x264) codec was selected as the reference as it is a modern, open, and widely used solution. Video compression parameters were adjusted in order to cover the recognition ability threshold. The compression was done resulting in 900 Processed Video Sequences (PVSs) with the bit-rate ranging from 40 kbit/s to 440 kbit/s [10].

The testers who participated in this study provided a total of 960 answers. Each answer could be interpreted as the number of per-character errors, i.e. 0 errors meaning correct recognition. The average probability of a licence plate being identified correctly was **0.548 (526/960)**, and **0.641** recognitions had no more than one error. **0.720** of all characters were recognised [10].

For further analysis it was assumed that the threshold detection parameter to be analyzed is the probability of plate recognition with no more than one error. For detailed results, please refer to Fig. 3.

3.2 Psychophysical experiment for credible bronchoscopic diagnosis

Well-established methods of subjective assessment are based on Receiver Operating Characteristic (ROC) curves. In this case, a different and tested [1, 9, 15, 16] subjective method for qualifying lossy compressed still images to a visually undistorted subset is based on sorting compressed images by their quality. The same approach

Fig. 3 Example of the obtained detection probability and model of the licence plate recognition task

was adopted in the investigation of 3 SRC video sequences used in the test. The length of sequences was similar to the licence plate recognition scenario. Expert testers (clinicians) were randomly presented with several video sequences: the original, and seven copies compressed with various bit-rate values. As a result of ordering, it was generally possible for an experiment supervisor to clearly distinguish two subsets of video sequences [1, 9, 15, 16]. The first consisted of the highest quality video sequences in a random order. The other video sequences appeared in the second lowest quality subset. Only the video sequences belonging to the first subset were considered to be of a quality suitable for diagnostic purposes [2].

A subjective evaluation of video sequences compressed at various bit-rates was performed. The MPEG-4 codec was selected as the reference as it is still the most widely used solution for telemedicine. The codec has also been successfully applied to surgery video compression [1, 9, 15, 16]. A test based on the abovementioned quality-based ordering method was carried out in order to obtain the bit-rate for which a clinician cannot distinguish between the original and the compressed video sequences. Each of the three original (uncompressed) video sequences was supplemented with a few compressed video sequences with the same content, thus constituting three investigated video sequences [2]. The compression was done with resulting in 24 PVSs with the bit-rate ranging approximately from **80 kbit/s** to **1280 kbit/s**.

Eight clinicians were asked independently to arrange the video sequence in each of the sets following the well-known bubble sort algorithm. The only information gathered was the order of the sequences. The clinicians used purpose-developed software in the sorting. The software was run on an ordinary personal computer located at the clinic. No time restrictions were applied to the evaluations [2].

For further analysis it was assumed that the threshold detection parameter to be analyzed is the likelihood that a clinician cannot distinguish between the original

Fig. 4 Example of the obtained detection probability and model of the bronchoscopic diagnosis task (source: literature [10])

and compressed video sequences. For detailed results, please refer to Fig. 4 and literature [13].

4 Modelling perceptual video quality

In the area of entertainment video, a great deal of research has been carried out on the parameters of the contents that are the most effective for perceptual quality. These parameters form a framework in which predictors can be created such that objective measurements can be developed through the use of subjective testing [17].

Assessment principles for task-based video quality are a relatively new field. Solutions developed so far have been limited mainly to optimizing network Quality of Service (QoS) parameters. Alternatively, classical quality models such as PSNR [3] or SSIM [19] have been applied, although they are not well suited to the task. The paper presents an innovative, alternative approach, based on modelling detection threshold probabilities.

4.1 Quality modelling of accurate license plate recognition

It was possible to fit a logarithmic function in order to model the quality (expressed as detection threshold probability) of the licence plate recognition task. This is an innovative approach. The achieved R^2 is 0.81 (see Fig. 3). According to the model, one may expect hundred percent correct recognition for bit-rates of around **350** kbit/s and higher can be expected. The accuracy of recognition depends on many external conditions and also on the size of image details, therefore hundred percent correct recognition can only be expected under ideal conditions.

4.2 Quality modelling of bronchoscopic diagnosis

Before presenting the results for the second quality modelling case, it should be noted that a common method of presenting results has been used. This is possible through the application of appropriate transformations, allowing the fitting of diverse recognition tasks into a single quality framework.

Again, it was possible to fit a logarithmic function in order to model the quality (expressed as detection threshold probability) of the bronchoscopic diagnosis task. This is also an innovative approach. The achieved R^2 is 0.71 (see Fig. 4). According to the model, hundred percent correct recognition for bit-rates of around **900** kbit/s and higher can be expected.

Unfortunately, due to the relatively high diversity of subjective answers, no better fitting was achievable in either case. However, a slight improvement is likely to be possible by using other curves.

5 Optimisation

A quality optimisation problem consists of maximising a multidimensional quality function. As in these experiment, multidimensional quality scaling took place only in the licence plate scenario, the second, bronchoscopic diagnosis scenario, cannot be discussed any-more.

Once quality levels (expressed, as in the licence plate scenario, as recognition rates) were established, the next stage was optimisation. When streaming a task-based video thorough a constrained bandwidth channel, and striving to achieve the highest possible recognition rate, it is possible to adapt the streamed video sequences in both the quality and spatial domains (keeping the temporal domain constant, as explained in Section 2). By operating with the compression QP and scaling down video resolutions, virtually all desired bit-rates can be achieved. However, the question as to which combination of the above-mentioned parameters produces the highest possible recognition rate remains unanswered.

In the remaining part of this Section, the generation of Hypothetical Reference Circuits (HRCs, Section 5.1) is introduced first, followed by example optimisation models using logarithmic (Section 5.2) and logistic (Section 5.3) functions.

5.1 Generation of hypothetical reference circuits

The authors tested multi-dimensional (more specifically, two-dimensional) optimisation. An experiment was conducted involving collecting, pooling and comparing recognition rates for various combinations of QP and resolution. The experiment was based on the licence plate recognition scenario solely, due to the higher availability of collected results. In order to verify whether the final results are related to the SRC video content, another set of video sequences was introduced. This set of sequences contained cropped versions of full frame videos, with the cropped area always including the Region of Interest (ROI), in this case the licence plate.

Prior to encoding, some modifications involving quadruple resolution changes and cropping were applied in order to obtain diverse aspect ratios between the car plates and the video size (see Fig. 5 for details related to the processing). Each SRC video sequence was modified into four versions, and each version was encoded with 5 different, graduating QPs. Selected QP values were adjusted to different video processing paths in order to cover the number plate recognition ability threshold. As a result, 20 different hypothetical reference circuits (HRC) were obtained.

Fig. 5 Generation of HRCs (based on literature [13])

Figure 6 presents example frames of four SRC versions. All the HRCs within the “Full frame” scenario (Fig. 6a and b) can be considered as various attempts to scale down the bit-rate of the given SRC. Nevertheless, HRCs of the “Cropped frame” scenario (Fig. 6c and d) should be considered as a separate set, used solely for verification of the results of the first scenario. Cropping cannot be considered as another scaling domain, as it requires prior decisions to be made on which particular areas of the frame will be irrevocably lost.

5.2 Optimisation using the logarithmic function

In this subsection, proposed optimisation models using logarithmic function are presented. Both the “Full frame” (Section 5.2.1) and “Cropped frame” (Section 5.2.2) scenarios are targeted. The scenarios are then combined (Section 5.2.3).

5.2.1 “Full frame” scenario with logarithmic modelling

The results of the “Full frame” scenario (Fig. 7) should be considered first. For bit-rates higher than approximately 200 kbit/s, only video sequences scaled in the quality domain have been achieved. They exhibit a high performance in terms of recognition

Fig. 6 Example frames of four SRC versions (with relative sizes maintained)

Fig. 7 Example of the obtained detection probabilities and logarithmic models for the “Full frame” scenario

rates (threshold probabilities). However, for bit-rates lower than **200** kbit/s, video sequences that have been additionally pre-scaled in the spatial domain achieve better quality. The performance gain can reach up to **0.2** of the threshold probability.

For the original scale video, the following regression function and R^2 have been achieved:

$$y = 0.79 \cdot \ln(x) - 3.37 \quad (1)$$

$$R^2 = 0.72 \quad (2)$$

For the scaled down video, the following regression function and R^2 have been achieved:

$$y = 0.77 \cdot \ln(x) - 3.14 \quad (3)$$

$$R^2 = 0.77 \quad (4)$$

5.2.2 “Cropped frame” scenario with logarithmic modelling

In order to verify whether the final results are related to the SRC video content, results from another set of video sequences have been analysed (Fig. 8). As clearly shown, the two regression curves resemble the previous case. Due to the slightly different modality of SRC data, the chart shows the same characteristics, although for shifted bit-rates (having the bit-rate threshold at around **130** kbit/s).

Fig. 8 Example of the obtained detection probabilities and logarithmic models for the “Cropped frame” scenario

For the original scale video, the following regression function and R^2 have been achieved:

$$y = 0.66 \cdot \ln(x) - 2.38 \quad (5)$$

$$R^2 = 0.78 \quad (6)$$

For the scaled down video, the following regression function and R^2 have been achieved:

$$y = 0.76 \cdot \ln(x) - 2.69 \quad (7)$$

$$R^2 = 0.91 \quad (8)$$

5.2.3 Combined scenarios with logarithmic modelling

The two examples given above could eventually lead to the kind of generalisation for the licence plate recognition task that states there is a threshold bit-rate below which it is better to start quality optimisation from spatial scaling. Unfortunately, due to different initial SRC conditions, the charts presented above cannot be directly combined in order to produce a more generalised model. However, a solution is available. The results can be combined if their bit-rates (Compressed Data Rates) are first normalised using the Compression Ratio parameter, and using the Uncompressed Data Rate as a common denominator, as shown in (9).

$$\text{Compression Ratio} = \frac{\text{Compressed Data Rate}}{\text{Uncompressed Data Rate}} \quad (9)$$

Figure 9 presents an example of the detection probabilities and models obtained for combined scenarios. It is clear that both the regression curves maintained the same characteristics as those presented in the two previous charts. The threshold for Compression Ratio can be observed at around **0.0005**. Furthermore, similarly to the individual scenarios, the performance gain can reach up to **0.2** of the threshold probability, with Compression Ratios at around **0.0002–0.0003**. The regression curve fitting for the combined data is slightly worse than for the individual scenarios due to the modality of the SRC data.

For the original scale video $R^2 = 0.64$ has been achieved. For the scaled down video, $R^2 = 0.72$ has been achieved.

5.3 Optimisation using the logistic function

While modelling using logarithmic functions allows for simply fitting the curve to average bit-rates of a given HRC, the averaging process could potentially impact the accuracy of fitting. Therefore, more detailed results could be achieved through optimisation using a logistic function.

A logistic function or logistic curve is a common sigmoid curve. It can model the “S-shaped” growth curve (abbreviated S-curve) for a population P . The initial stage of growth is approximately exponential; then, as saturation begins, the growth slows, and at maturity, stops [20].

Fig. 9 Example of the obtained detection probabilities and logarithmic models for combined scenarios

Ordinary regression deals with finding a function that relates a continuous outcome variable (dependent variable y) to one or more predictors (independent variables x_1, x_2 , etc.). Simple linear regression assumes a function of the form:

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \cdots + a_n \cdot x_n \quad (10)$$

and finds the values of a_0, a_1, a_2 , etc. (a_0 is called the “intercept” or “constant term”).

Logistic regression is a variation of ordinary regression, useful when the observed outcome is restricted to two values, which usually represent the occurrence or non-occurrence of a particular outcome event (usually coded as 1 or 0 respectively). It produces a formula that predicts the probability of the occurrence as a function of the independent variables.

Logistic regression fits a special S-curve by taking the linear regression (above), which could produce any y -value between minus infinity and plus infinity, and transforming it with the function

$$P = \frac{e^y}{1 + e^y} \quad (11)$$

which produces P -values between 0 (as y approaches minus infinity) and 1 (as y approaches plus infinity). This now becomes a special type of non-linear regression [4, 14].

The remaining part of this subsection presents example optimisation models using the logistic function. Both the “Full frame” (Section 5.3.1) and “Cropped frame” (Section 5.3.2) scenarios are targeted. The scenarios are then combined (Section 5.3.3).

5.3.1 “Full frame” scenario with logistic modelling

When examining the results of the “Full frame” scenario (Fig. 10), it can be seen that for bit-rates higher than approximately **200** kbit/s, there is no statistical difference between scaling with or without the spatial domain. Both scaling options have a high performance in terms of recognition rates (threshold probabilities). However, for bit-rates lower than **200** kbit/s, video sequences that have been additionally pre-scaled in the spatial domain achieve better quality. The performance gain can reach up to around **0.15** of the threshold probability.

Unfortunately, confidence intervals (denoted as “CL” on the charts) continue to overlap slightly, therefore more variants of spatial domain scaling will need to be checked in the future. This approach should allow for more evident performance gains.

For the original scale video, the following regression functions have been achieved:

$$y = -2.66 + \frac{2.52 \cdot x}{100} \quad (12)$$

$$P = \frac{e^{-2.66 + \frac{2.52 \cdot x}{100}}}{1 + e^{-2.66 + \frac{2.52 \cdot x}{100}}} \quad (13)$$

Fig. 10 Example of the obtained detection probabilities and logistic models for the “Full frame” scenario

For the scaled down video, the following regression functions have been achieved:

$$y = -1.85 + \frac{1.79 \cdot x}{100} \quad (14)$$

$$P = \frac{e^{-1.85 + \frac{1.79 \cdot x}{100}}}{1 + e^{-1.85 + \frac{1.79 \cdot x}{100}}} \quad (15)$$

5.3.2 “Cropped frame” scenario with logistic modelling

Similarly to logarithmic function modelling, in order to verify whether the final results are related to the SRC video content, results from another set of video sequences have been analysed (Fig. 11). The two regression curves clearly show the same tendency as in the previous case. Due to a slightly different modality of SRC data, the chart shows the same characteristics, although for shifted bit-rates (having the bit-rate threshold at around **180** kbit/s).

Unfortunately, the situation with confidence intervals is similar to the previous example. The confidence intervals continue to overlap slightly, although the performance gain is slightly more evident.

For the original scale video, the following regression functions have been achieved:

$$y = -2.52 + \frac{3.2 \cdot x}{100} \quad (16)$$

$$P = \frac{e^{-2.52 + \frac{3.2 \cdot x}{100}}}{1 + e^{-2.52 + \frac{3.2 \cdot x}{100}}} \quad (17)$$

Fig. 11 Example of the obtained detection probabilities and logistic models for the “Cropped frame” scenario

For the scaled down video, the following regression functions have been achieved:

$$y = -2.72 + \frac{4.19 \cdot x}{100} \quad (18)$$

$$P = \frac{e^{-2.72 + \frac{4.19 \cdot x}{100}}}{1 + e^{-2.72 + \frac{4.19 \cdot x}{100}}} \quad (19)$$

5.3.3 Combined scenario with logistic modelling

Using the Compression Ratio approach, the charts presented above have been combined in order to produce a more generalised model. Figure 12 presents an example of the detection models obtained for combined scenarios. It is clear that both the regression curves maintained the same characteristics as those presented in the two previous charts. The threshold for the Compression Ratio can be observed at around **0.0007**. Furthermore, similarly to the individual scenarios, the performance gain of the threshold probability is visible.

For the original scale video, the following regression functions have been achieved:

$$y = -1.63 + 6896 \cdot x \quad (20)$$

$$P = \frac{e^{-1.63 + 6896 \cdot x}}{1 + e^{-1.63 + 6896 \cdot x}} \quad (21)$$

Fig. 12 Example of the obtained detection probabilities and logistic models for combined scenarios

For the scaled down video, the following regression functions have been achieved:

$$y = -1.78 + 8160 \cdot x \quad (22)$$

$$P = \frac{e^{-1.78+8160 \cdot x}}{1 + e^{-1.78+8160 \cdot x}} \quad (23)$$

6 Conclusions and future work

This paper provides an introduction to two typical usage cases of task-based video: a surveillance video used for accurate license plate recognition, and a medical video used for credible diagnosis prior to bronchoscopic surgery. The field of task-based video quality assessment is presented, from subjective psychophysical experiments to objective quality models. Example test results and models complement the descriptions provided. Finally, a quality optimisation approach driven by recognition rates is presented. The results show that video sequences additionally pre-scaled in the spatial domain achieve better quality. This is an important result, since it contradicts the common pursuit towards high camera resolutions.

Unfortunately, in many cases, the modelled threshold probability is outside of the confidence intervals of the actual data. Confidence intervals continue to overlap slightly, therefore more variants of spatial domain scaling will need to be checked in future. This approach should allow for more evident performance gains. Nevertheless, these kind problems related to varied responses, is very common in case of dealing with data collected from human subjects.

The methodologies outlined in this paper are just a single contribution to the overall framework of quality standards for task-based video. It is necessary to define requirements starting from the camera, through the broadcast, until after the presentation. These requirements will depend on scenario recognition. Some results can be generalized in the optimisation process, despite the modality of SRC data.

The practical value of the contribution is limited thus far, since it refers to limited scenarios. Concurrently, Proof of Concept (POC) implementation of a quality-driven streaming system is under development. Therefore more subjective experiments will need to be performed in order to fine-tune the process, encompassing more diversification of the content and its quality scaling parameters.

The approach presented is a starting point of a more advanced framework of objective quality assessment, described below. Extensive work is being carried out nowadays in the area of video quality, mainly driven by the Video Quality Experts Group (VQEG) [18]. A new project, Quality Assessment for Recognition Tasks (QART), was created for task-based video quality research at a recent meeting of the VQEG. QART will address the problems of a lack of quality standards for video monitoring. The initiative is co-chaired by the NTIA (National Telecommunications and Information Administration, an agency of the United States Department of Commerce) and AGH University of Science and Technology in Kraków, Poland. The aims of QART are the same as those of other VQEG projects—to advance the field of quality assessment for task-based video through collaboration in the development of test methods (including possible enhancements of the ITU-T P.912 Recommendation [7]), performance specifications and standards for task-based video, and predictive models based on network and other relevant parameters.

Acknowledgements This work was supported by the European Commission under the Grant INDECT No. FP7-218086. The author extends his thanks to his AGH colleague, Lucjan Janowski, for his ideas and general guidance on statistics.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Bartkowiak M, Domanski M (2001) Applications of chrominance vector quantization to intraframe and interframe compression of colour video sequences
2. Duplaga M, Leszczuk M, Papir Z, Przelaskowski A (2008) Evaluation of quality retaining diagnostic credibility for surgery video recordings. In: Proceedings of the 10th international conference on visual information systems: web-based visual information search and management, VISUAL '08. Springer, Berlin, pp 227–230
3. Eskicioglu AM, Fisher PS (1995) Image quality measures and their performance. *IEEE Trans Commun* 43(12):2959–2965. doi:10.1109/26.477498
4. Hosmer D, Lemeshow S (2000) Applied logistic regression. In: Wiley series in probability and statistics: texts and references section. Wiley, New York. URL <http://books.google.com/books?id=Po0RLQ7USIMC>. Accessed 9 July 2012
5. ITU-T P.800 (1996) Methods for subjective determination of transmission quality. URL <http://www.itu.int/rec/T-REC-P.800-199608-I>. Accessed 9 July 2012
6. ITU-T P.910 (1999) Subjective video quality assessment methods for multimedia applications. URL <http://www.itu.int/rec/T-REC-P.910-200804-I>. Accessed 9 July 2012
7. ITU-T P.912 (2008) Subjective video quality assessment methods for recognition tasks. URL <http://www.itu.int/rec/T-REC-P.912-200808-I>. Accessed 9 July 2012

8. Janowski L, Romaniak P (2010) QoE as a function of frame rate and resolution changes. In: Zeadally S, Cerqueira E, Curado M, Leszczuk M (eds) FMN. Lecture notes in computer science, vol 6157. Springer, Berlin, pp 34–45
9. Leszczuk M (2006) Analiza mozliwosci budowy internetowych aplikacji dostepu do cyfrowych bibliotek wideo. Ph.D. thesis, AGH University of Science and Technology, Krakow
10. Leszczuk M (2011) Assessing task-based video quality—a journey from subjective psycho-physical experiments to objective quality models. In: Dziech A, Czyżewski A (eds) Multimedia communications, services and security, communications in computer and information science, vol 149. Springer, Berlin, pp 91–99
11. Leszczuk M (2012) Redefining ITU-T P.912 recommendation requirements for subjects of quality assessments in recognition tasks. In: Dziech A, Czyżewski A (eds) Multimedia communications, services and security, communications in computer and information science. Springer, Berlin
12. Leszczuk M, Grega M (2010) Prototype software for video summary of bronchoscopy procedures with the use of mechanisms designed to identify, index and search. In: Pietka E, Kawa J (eds) Information technologies in biomedicine. Advances in soft computing, vol 69. Springer, Berlin, pp 587–598
13. Leszczuk M, Janowski L, Romaniak P, Glowacz A, Mirek R (2011) Quality assessment for a licence plate recognition task based on a video streamed in limited networking conditions. In: Fourth multimedia communications, services and security (MCSS'11). Krakow, Poland
14. Pezzullo JC, Sullivan KM (2005) Logistic regression. <http://statpages.org/logistic.html>. Accessed 9 July 2012
15. Przelaskowski A (2002) Falkowe metody kompresji danych obrazowych. Ph.D. thesis, Oficyna Wydawnicza Politechniki Warszawskiej, Warsaw
16. Skarbek W (1998) Multimedia—algorytmy i standardy. PLJ, Warsaw
17. Takahashi A, Schmidmer C, Lee C, Speranza F, Okamoto J, Brunnström K, Janowski L, Barkowsky M, Pinson M, Nicolas S, Huynh Thu Q, Green R, Bitto R, Renaud R, Borer S, Kawano T, Baroncini V, Dhondt Y (2010) Report on the validation of video quality models for high definition video content. Tech. rep., Video Quality Experts Group
18. VQEG: The Video Quality Experts Group. <http://www.vqeg.org/>. Accessed 9 July 2012
19. Wang Z, Lu L, Bovik AC (2004) Video quality assessment based on structural distortion measurement. Signal Process, Image Commun 19(2):121–132. doi:10.1016/S0923-5965(03)00076-6%20
20. Wikipedia (2011) Logistic function. http://en.wikipedia.org/wiki/Logistic_function. Accessed 9 July 2012

Mikołaj Leszczuk (Phd. Eng.) is an assistant professor in the Faculty of Telecommunication at AGH University (Krakow). He is a scholarship holder of Universidad Carlos III de Madrid (Madrid, Spain), where he prepared his Master thesis “Implementation of the Recognition Engine’s Search Module in Multimedia Application”. His current research interests are focused on the development of digital video libraries, in particular video summarizing, indexing, compression and streaming of video sequences. The results of these studies were published in 2006 as a doctoral dissertation.