Quality Assessment for Recognition Tasks (QART)

Mikołaj Leszczuk
AGH University of Science and Technology
Department of Telecommunications
Kraków, Poland
Email: leszczuk@agh.edu.pl

Joel Dumke
National Telecommunications and Information Administration
Institute for Telecommunication Sciences
Boulder CO, USA
Email: jdumke@its.bldrdoc.gov

***Abstract*—Users of video to perform tasks require sufficient video quality to recognize the information needed for their application. Therefore, the fundamental measure of video quality in these applications is the success rate of these recognition tasks, which is referred to as visual intelligibility or acuity. One of the major causes of reduction of visual intelligibility is loss of data through various forms of compression. Additionally, the characteristics of the scene being captured have a direct effect on visual intelligibility and on the performance of a compression operation-specifically, the size of the target of interest, the lighting conditions, and the temporal complexity of the scene. This paper presents a Work in Progress (WIP) Quality Assessment for Recognition Tasks (QART) project, which is performing a series of tests to study the effects and interactions of compression and scene characteristics. An additional goal is to test existing or develop new objective measurements that will predict the results of the subjective tests of visual intelligibility.**

***Index Terms*—Video; compression; MOS (Mean Opinion Score), WIP, QART**

I. Introduction

The transmission and analysis of video is often used for a variety of applications outside the entertainment sector, and generally this class of (task-based) video is used to perform a specific recognition task. Examples of these applications include security, public safety, remote command and control, tele-medicine, and sign language. The Quality of Experience (QoE) concept for video content used for entertainment differs materially from the QoE of video used for recognition tasks because in the latter case, the subjective satisfaction of the user depends upon achieving the given task, e.g., event detection or object recognition. Additionally, the quality of video used by a human observer is largely separate from the objective video quality useful in computer vision [1]. Therefore, it is crucial to measure and ultimately optimize task-based video quality. This is discussed in more detail in [2].

Enormous work, mainly driven by the Video Quality Experts Group (VQEG) [3], has been carried out for the past several years in the area of consumer video quality. The VQEG is a group of experts from various backgrounds and affiliations, including participants from several internationally recognized organizations, working in the field of video quality assessment. The group was formed in October of 1997 at a meeting of video quality experts. The majority of participants are active in the International Telecommunication Union (ITU) and VQEG combines the expertise and resources found in several ITU Study Groups to work towards a common goal [3]. Unfortunately, many of the VQEG and ITU methods and recommendations (like ITU's Absolute Category Rating – ACR – described in ITU-T P.800 [4]) are not appropriate for the type of testing and research that task-based video, including CCTV, requires.

This paper is organized as follows. Section II describes related work and motivation. In Section III, the QART Project and standardisation is discussed. Section IV concludes the paper and details the future work.

II. Related Work and Motivation

Some subjective recognition metrics, described below, have been proposed over the past decade. They usually combine aspects of Quality of Recognition (QoR) and QoE. These metrics have been not focused on practitioners as subjects, but rather on naïve participants. The metrics are not context specific, and they do not apply video surveillance-oriented standardized discrimination levels.

One of the metrics being definitively worth to mention is Ghinea's Quality of Perception (QoP) [5], [6]. However, the QoP metric does not entirely fit video surveillance needs. It targets mainly video deterioration caused by frame rate (fps), whereas fps does not necessarily affect the quality of Closed-Circuit Tele-Vision (CCTV) and the required bandwidth [7]. The metric has been established for rather low, legacy resolutions, and tested on rather small groups of subjects (10 instead of standardized 24 valid, correlating subjects). Furthermore, a video recognition quality metric for a clear objective of video surveillance context requires tests in fully controlled environment [8], with standardized discrimination levels (avoiding ambiguous questions) and with minimized impact of subliminal cues [9].

Another metric being worth to mention is QoP's offshoot, Strohmeier's Open Profiling of Quality (OPQ) [10]. This metric puts more stress on video quality than on recognition/discrimination levels. Its application context, being focused on 3D, is also different than video surveillance which requires rather 2D. Like the previous metric, this one also does not apply standardized discrimination levels, allowing subjects to use their own vocabulary. The approach is qualitative rather than quantitative, whereas the latter is preferred by public safety practitioners for, e.g., public procurement. The OPQ model is somewhat content/subject-oriented, while a more generalized metric framework is needed for video surveillance.

 ISBN: 978-1-61208-239-4

OPQ partly utilizes free sorting, as used in [11], but also applied in the method called Interpretation Based Quality (IBQ) [12], [13], adapted from [14], [15]. Unfortunately, these approaches allow mapping relational, rather than absolute, quality.

Furthermore, there exists only a very limited set of quality standards for task-based video applications. Therefore, it is still necessary to define the requirements for such systems from the camera, to broadcast, to display. The nature of these requirements will depend on the task being performed.

European Norm №. 50132 [16] was created to ensure that European CCTV systems are realized under the same rules and requirements. The existence of a standard has opened an international market of CCTV devices and technologies. By selecting components that are consistent with the standard, a user can achieve a properly working CCTV system. This technical regulation deals with different parts of a CCTV system including acquisition, transmission, storage, and playback of surveillance video. The standard consists of such sections as lenses, cameras, local and main control units, monitors, recording and hard copy equipment, video transmission, video motion detection equipment, and ancillary equipment. This norm is hardware-oriented as it is intended to unify European law in this field; thus, it does not define the quality of video from the point of view of recognition tasks.

The Video Quality in Public Safety (VQiPS) Working Group, established in 2009 and supported by the U.S. Department of Homeland Securitys Office for Interoperability and Compatibility, has been developing a user guide for public safety video applications. The goal of the guide is to provide potential public safety video customers with links to research and specifications that best fit their particular application, as such research and specifications become available. The process of developing the guide will have the desired secondary effect of identifying areas in which adequate research has not yet been conducted, so that such gaps may be filled. A challenge for this particular work is ensuring that it is understandable to customers within public safety, who may have little knowledge of video technology.

The approach taken by VQiPS is to remain application-agnostic. Instead of attempting to individually address each of the many public safety video applications, the guide is based on their common features. Most importantly, as mentioned above, each application consists of some type of recognition task. The ability to achieve a recognition task is influenced by many parameters, and five of them have been selected as being of particular importance. They are:

- **Usage time-frame.** Specifies whether the video will need to be analysed in real-time or recorded for later analysis.
- **Discrimination level.** Specifies the level of detail required from the video.
- **Target size.** Specifies whether the anticipated region of interest in the video occupies a relatively small or large percentage of the frame.
- **Lighting level.** Specifies the anticipated lighting level of the scene.
- **Level of motion.** Specifies the anticipated level of motion in the scene.

These parameters form what are referred to as Generalized Use Classes, or GUCs [17]. Fig. 1 is a representation of the GUC determination process.

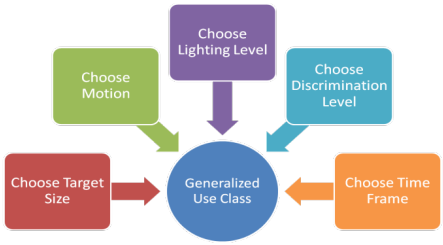

Fig. 1. Classification of video into GUC as proposed by VQiPS (source: [1]).

To develop accurate objective measurements and models for video quality assessment, subjective tests (psychophysical experiments) must be performed. The ITU has recommendations that address the methodology for performing subjective tests in a rigorous manner [8], [18]. These methods are targeted at the entertainment application of video and were developed to assess a person's perceptual opinion of quality. They are not entirely appropriate for task-based applications, in which video is used to recognize objects, people or events.

Assessment principles for the maximization of task-based video quality are a relatively new field. Problems of quality measurements for task-based video are partially addressed in a few preliminary standards and a Recommendation ITU-T P.912 [9], [19] that mainly introduce basic definitions, methods of testing and psycho-physical experiments. ITU-T P.912 describes multiple choice, single answer, and timed task subjective test methods, as well as the distinction between real-time and viewer-controlled viewing, and the concept of scenario groups to be used for these types of tests. Scenario groups are groups of very similar scenes with only small, controlled differences between them, which enable testing recognition ability while eliminating or greatly reducing the potential effect of scene memorization. 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.

III. The QART Project and Standardisation

Internationally, the number of people and organizations interested in this area continues to grow, and there is currently enough interest to motivate the creation of a task-based video project under VQEG. At one of the recent meetings of VQEG, a new project was formed for task-based video quality research. The Quality Assessment for Recognition Tasks (QART) project addresses precisely the problem of lack of quality standards for video monitoring [20]. The initiative is co-chaired by Public Safety Communications Research (PSCR) program, U.S.A., and AGH University of Science and Technology in Krakow, Poland. Other members include

 ISBN: 978-1-61208-239-4

research teams from Belgium, France, Germany, and South Korea. The purpose of QART is exactly the same as the other VQEG projects – to advance the field of quality assessment for task-based video through collaboration in the development of test methods, performance specifications and standards for task-based video, as well as predictive models based on network and other relevant parameters [21].

There has been some QART work conducted so far. The research has answered the practical problem of a network link with a limited bandwidth and detection probability is an interesting parameter to find. QART have presented the results of the development of critical quality thresholds in licence plate recognition by human subjects, based on a video streamed in constrained networking conditions. Many video sequences originated from the database of the Consumer Digital Video Library (CDVL) [22]. QART have shown that, for a particular view, a model of detection probability based on bit rate can work well. Nevertheless, different views have very different effects on the results obtained. QART have also learned that for these kinds of psycho-physical experiments, licence plate characteristics (such as illumination) are of great importance, sometimes even prevailing over the distortions caused by bit-rate limitations and compression [23].

One important conclusion is that for a bit rate as low as 180 kbit/s the detection probability is over 80% even if the visual quality of the video is very low. Moreover, the detection probability depends strongly on the Source Reference Channel/Circuit. (SRC, over all detection probability varies from 0 (sic!) to over 90%) [23].

Furthermore, a study of the ability to recognize a moving or stationary object given several lighting and target size combinations, and a study of license plate recognition, both processed at a number of compression rates, have been completed. These are the first in a planned series of studies with the similar goal of studying the ability to recognize objects given various network conditions [1].

Recently, a subjective test has been completed, consisting of various levels of compression and resolution reduction following the methods suggested in ITU-T P.912 and the VQiPS GUCs [3]. The test method was the multiple choice method. Bit-rates from 64 kbit/s to 1536 kbit/s using H.264 encoding were studied, in combination with either VGA or CIF resolution. A total of 10 bit-rate/resolution combinations were tested. The recognition task for the viewer was the identification of an object within a simulated real-time environment (i.e., pausing or replaying the video was not allowed.) An example of the user interface is shown in Fig. 2.

The objects were either stationary or moving, and were filmed under three lighting conditions and at two distances from the camera. The test results thus can be categorized into several of the GUCs. Results were presented as recognition rates; in other words, the percentage of objects correctly identified (after normalization for guessing). Recognition rates of 90% and 50% were chosen as significant thresholds for which recommendations were suggested based on test results.

The accuracy of answers given by subjects was growing during the test. It suggests that subjects were aided by memory effects as the test progressed.

Fig. 2. User interface for subjective target recognition task test (source: [1]).

Finally, QART recently completed a test of subjects' ability to recognize car registration numbers in video material recorded using a CCTV camera and compressed with the H.264/AVC codec [2].

A subjective experiment was carried out in order to perform the analysis. A psycho-physical 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. ITUs Single Stimulus (SS) described in ITU-R BT.500-11 [8], was selected as the subjective test methodology [2].

The recognition task was threefold: 1) type in the licence plate number, 2) select car colour, and 3) select car make. Testers were allowed to control playback and enter full screen mode. A more detailed description of the recognition task is available in [2].

The tests were conducted using a web-based interface connected to a database. In the database both information about the video samples and the answers received from the testers were gathered. An example of the user interface is shown in Fig. 3.

Video sequences used in the test were recorded in a car park using a CCTV camera. The H.264 codec with x264 implementation 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 with the bit-rate ranging from 40 kbit/s to 440 kbit/s [2].

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., zero errors meaning correct recognition. The average probability of a license plate being identified correctly was 54.8% with 526 recognitions out of 960, 64.1% recognitions had no more than one error, and 72% of all characters were recognized [2].

 ISBN: 978-1-61208-239-4

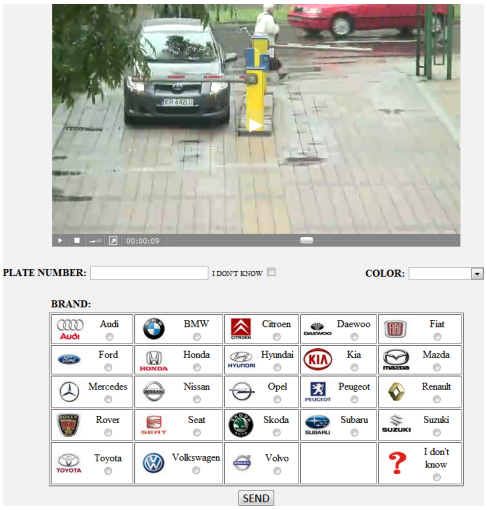

Fig. 3. User interface for subjective plate recognition task test (source: [23]).

IV. Conclusion and Future Work

In summary, QART introduced contributions to the field of task-based video quality assessment methodologies: from subjective psycho-physical experiments to objective quality models. The developed methodologies are just a single contribution to the overall framework of quality standards for task-based video. It is necessary to further define requirements starting from the camera, through the broadcast, and until after the presentation. These requirements will depend on the particular tasks users wish to perform [2]. Future work will include, e.g., quantification of GUC and extension of P.912.

Acknowledgments

For the research leading to these results, Mikołaj Leszczuk would like to thank the European Community's Seventh Framework Program (FP7/2007-2013) for received funding under grant agreement №. 218086 (INDECT) as well as to thank AGH University of Science and Technology for received funding under Department of Telecommunications statutory work. Joel Dumke would like to thank the U.S. Department of Commerce for received funding for the research under the Public Safety Communications Research project.

References

[1] M. Leszczuk, I. Stange, and C. Ford, "Determining image quality requirements for recognition tasks in generalized public safety video applications: Definitions, testing, standardization, and current trends," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2011, pp. 1–5.

[2] M. Leszczuk, "Assessing task-based video quality a journey from subjective psycho-physical experiments to objective quality models," in *Multimedia Communications, Services and Security*, ser. Communications in Computer and Information Science, A. Dziech and A. Czyżewski, Eds. Springer Berlin Heidelberg, 2011, vol. 149, pp. 91–99. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21512-4_11

[3] *The Video Quality Experts Group*, VQEG, July 2012, http://www.vqeg.org/.

[4] *ITU-T P.800, Methods for subjective determination of transmission quality*, International Telecommunication Union Recommendation, 1996. [Online]. Available: http://www.itu.int/rec/T-REC-P.800-199608-I

[5] G. Ghinea and J. P. Thomas, "Qos impact on user perception and understanding of multimedia video clips," in *Proceedings of the sixth ACM international conference on Multimedia*, ser. MULTIMEDIA '98. New York, NY, USA: ACM, 1998, pp. 49–54. [Online]. Available: http://doi.acm.org/10.1145/290747.290754

[6] G. Ghinea and S. Y. Chen, "Measuring quality of perception in distributed multimedia: Verbalizers vs. imagers," *Computers in Human Behavior*, vol. 24, no. 4, pp. 1317–1329, 2008.

[7] L. Janowski and P. Romaniak, "Qoe as a function of frame rate and resolution changes," in *Future Multimedia Networking*, ser. Lecture Notes in Computer Science, S. Zeadally, E. Cerqueira, M. Curado, and M. Leszczuk, Eds. Springer Berlin / Heidelberg, 2010, vol. 6157, pp. 34–45. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-13789-1_4

[8] *ITU-R BT.500-12, Methodology for the subjective assessment of the quality of television pictures*, International Telecommunication Union Recommendation, Rev. 12, 2009. [Online]. Available: http://www.itu.int/rec/R-REC-BT.500-12-200909-I

[9] *ITU-T P.912, Subjective video quality assessment methods for recognition tasks*, International Telecommunication Union Recommendation, 2008. [Online]. Available: http://www.itu.int/rec/T-REC-P.912-200808-I

[10] D. Strohmeier, S. Jumisko-Pyykkö, and K. Kunze, "Open profiling of quality: a mixed method approach to understanding multimodal quality perception," *Adv. MultiMedia*, vol. 2010, pp. 3:1–3:17, January 2010. [Online]. Available: http://dx.doi.org/10.1155/2010/658980

[11] M. Duplaga, M. Leszczuk, Z. Papir, and A. Przelaskowski, "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*, ser. VISUAL '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 227–230.

[12] J. Radun, T. Leisti, J. Hakkinen, H. Ojanen, J. L. Olives, T. Vuori, and G. Nyman, "Content and quality: Interpretation-based estimation of image quality," *ACM Transactions on Applied Perception*, vol. 4, no. 4, pp. 2:1–2:15, 2008. [Online]. Available: http://doi.acm.org/10.1145/1278760.1278762

[13] G. Nyman, J. Radun, T. Leisti, J. Oja, H. Ojanen, J. L. Olives, T. Vuori, and J. Hakkinen, "What do users really perceive — probing the subjective image quality experience," in *Proceedings of the SPIE International Symposium on Electronic Imaging 2006: Imaging Quality and System Performance III, Vol. 6059*, 2006, pp. 1–7.

[14] P. Faye, D. Bremaud, M. D. Daubin, P. Courcoux, A. Giboreau, and H. Nicod, "Perceptive free sorting and verbalisation tasks with naive subjects: an alternative to descriptive mappings," *Food Quality and Preference*, vol. 15, no. 7-8, pp. 781 – 791, 2004, fifth Rose Marie Pangborn Sensory Science Symposium. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950329304000540

[15] D. Picard, C. Dacremont, D. Valentin, and A. Giboreau, "Perceptual dimensions of tactile textures," *Acta Psychologica*, vol. 114, no. 2, pp. 165–184, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0001691803000751

[16] *CENELEC EN 50132, Alarm systems. CCTV surveillance systems for use in security applications.*, European Committee for Electrotechnical Standardization European Norm, 2011.

[17] VQiPS, "Video quality tests for object recognition applications," U.S. Department of Homeland Security's Office for Interoperability and Compatibility, June 2011. [Online]. Available: http://www.safecomprogram.gov/SAFECOM/library/technology/1627_additionalstatement.htm

[18] *ITU-T P.910, Subjective video quality assessment methods for multimedia applications*, International Telecommunication Union Recommendation, 1999. [Online]. Available: http://www.itu.int/rec/T-REC-P.910-200804-I

[19] C. G. Ford, M. A. McFarland, and I. W. Stange, "Subjective video quality assessment methods for recognition tasks," *Human Vision and Electronic Imaging XIV*, vol. 7240, no. 1, p. 72400Z, 2009. [Online]. Available: http://link.aip.org/link/?PSI/7240/72400Z/1

[20] M. Leszczuk and J. Dumke, *The Quality Assessment for Recognition Tasks (QART)*, VQEG, July 2012, http://www.its.bldrdoc.gov/vqeg/project-pages/qart/qart.aspx.

[21] P. Szczuko, P. Romaniak, M. Leszczuk, R. Mirek, M. Pleva, S. Ondas, G. Szwoch, P. Korus, C. Kollmitzer, P. Dalka, J. Kotus, A. Ciarkowski, A. Dąbrowski, P. Pawłowski, T. Marciniak, R. Weychan, and F. Misiorek, "D1.2, report on ns and cs hardware construction," The INDECT Consortium: Intelligent Information System Supporting Observation,

 ISBN: 978-1-61208-239-4

Searching and Detection for Security of Citizens in Urban Environment, European Seventh Framework Programme FP7-218086-collaborative project, Europa, Tech. Rep., 2010, cop.

[22] *The Consumer Digital Video Library*, CDVL, July 2012, http://www.cdvl.org/.

[23] M. Leszczuk, L. Janowski, P. Romaniak, A. Głowacz, and R. Mirek, "Quality assessment for a license plate recognition task based on a video streamed in limited networking conditions," in *4th International Conference on Multimedia Communications, Services and Security*, ser. Communications in Computer and Information Science, A. Dziech and A. Czyżewski, Eds., vol. 149. Krakow, Poland: Springer Berlin Heidelberg, 2-3 June 2011, pp. 10–18. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21512-4_2

 ISBN: 978-1-61208-239-4