Determining Image Quality Requirements for Recognition Tasks in Generalized Public Safety Video Applications: Definitions, Testing, Standardization, and Current Trends

Mikołaj I. Leszczuk, *Member, IEEE*, Irena Stange, and Carolyn Ford

***Abstract*—The Quality of Experience (QoE) concept for video content used for entertainment differs materially from the QoE of video used for public safety tasks because in the latter case, the subjective satisfaction of the user depends upon achieving the given task. Yet currently there are hardly any quality standards for task-based video applications. This is an important problem as the transmission and analysis of video is used for many applications outside the entertainment sector, and generally this class of video is used in the performance of a specific task. To address this lack, in this paper we introduce new methods of assessing video quality for task-based video that have been under development. Once a framework and measurement system have been developed for task-based video, performance specifications and standards can be developed to assist users of task-based video to identify the technology that will allow them to successfully perform the required function.**

***Index Terms*—Image quality, Objective evaluation techniques, Standardization, Subjective evaluation techniques.**

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 video is used to perform a specific task. Examples of these applications include security, public safety, remote command and control, telemedicine, 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 processing—Computer Vision.

There are hardly any quality standards for task-based video applications. Therefore it is necessary to define the requirements for such systems from the camera, to broadcast, to display. The nature of these requirements will depend on the recognition scenario.

In this paper, we introduce new methods of assessing video quality for task-based video that have been under development. Once a framework and measurement system have been developed for task-based video, performance specifications and standards can be developed that will assist users of task-based video to identify the technology that will allow them to successfully perform the function required.

In the area of entertainment video, much research has been performed into the content parameters that most affect perceptual quality [1]-[4]. These parameters form a framework in which predictors can be created, and thus objective measurements developed, through the use of subjective testing. For task-based video, a different framework must be created, appropriate to the function of the video—i.e., its use for recognition tasks, not entertainment. Once a framework is in place, methods should be developed to measure the usefulness of the video, not its entertainment value.

Enormous work, mainly driven by the Video Quality Experts Group (VQEG) [5], 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 [5]. Unfortunately, many of the VQEG and ITU methods and recommendations (like ITU's Absolute Category Rating—ACR—described in ITU-T P.800) are not appropriate for the type of testing and research that task-based video, including closed-circuit television (CCTV), requires.

Manuscript received May 6, 2011. The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n° 218086, and from the U.S. Department of Homeland Security's Office for Interoperability and Compatibility.

M. I. Leszczuk is with the Department of Telecommunications, AGH University of Science and Technology, Krakow, PL 30059 Poland (phone: +48-607-720-398, fax: +48-12-634-23-72, e-mail: leszczuk@agh.edu.pl).

I. Stange is with the National Telecommunications and Information Administration, Institute for Telecommunication Sciences (NTIA/ITS), working on behalf of the Public Safety Communications Research project under the U.S. Department of Commerce, Boulder, CO 80305 USA (phone: 303-497-4611; fax: 303-497-5969; irena@its.bldrdoc.gov).

C. Ford was, until February 2011, with the National Telecommunications and Information Administration, Institute for Telecommunication Sciences (NTIA/ITS), working on behalf of the Public Safety Communications Research project under the U.S. Department of Commerce, Boulder, CO 80305 USA.

European Norm number 50132 [7] was created to ensure that CCTV systems are realized under the same rules and requirements in all countries in Europe. The existence of a standard has opened an international market of video surveillance 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: from acquisition, to transmission, to storage and playback of surveillance video. The standard consist 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.

To develop accurate objective measurements and models for video quality assessment, subjective experiments must be performed. The ITU has Recommendations that address the methodology for performing subjective tests in a rigorous manner [8], [9]. 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.

Current efforts to remedy this lack of video quality standards and measurement methods for task-based video are presented in this paper. Section II presents a framework for describing public safety video applications. Section III discusses evaluation and optimization of quality. Section IV reports current research approaches. Standardization activities are described in Section V. Section VI discusses ethical problems; the paper is concluded in Section VII.

II. A Framework for Describing Public Safety Video Applications

The term *video quality* can be interpreted in many different ways. For public safety video applications, video quality generally refers to the delivered visual intelligibility of the video, given a target of interest and desired discrimination level. Establishing video quality parameters for this type of video application should enable a mapping of public safety applications to appropriate performance specifications for components of a video system that can create or affect a video stream. These components include lens configurations, image capture systems, video stream transport systems, video processing and storage systems, and displays.

The VQiPS (Video Quality in Public Safety) Working Group, established in 2009 and supported by the U.S. Department of Homeland Security's 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.

In July 2010, Volume 1.0 of the framework document "Defining Video Quality Requirements: A Guide for Public Safety" was released [10]. This document provides qualitative guidance, such as explaining the role of various components of a video system and their potential impact on the resultant video quality. The information in this document as well as quantitative guidance will start to become available at the VQiPS website in June 2011 [11].

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 commonalities between them. Most importantly, as mentioned above, each application consists of some type of recognition task. The ability to achieve a recognition task is impacted by many parameters, and five of them have been selected as being of particular importance. They are:

1. **Usage timeframe.** Specifies whether the video will need to be analyzed in real-time or will be recorded for later analysis.
2. **Discrimination level.** Specifies how fine a level of detail is sought from the video.
3. **Target size.** Specifies whether the anticipated region of interest in the video occupies a relatively small or large percentage of the frame.
4. **Lighting level.** Specifies the anticipated lighting level of the scene.
5. **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. Figure 1 is a representation of the GUC determination process.

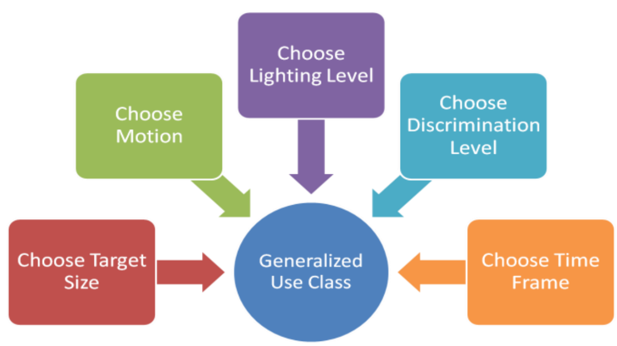

Figure 1. Classification of video into generalized use classes as proposed by VQiPS.

The VQiPS user guide is intended to help the end users determine how their application fits within these parameters. The research and specifications provided to users is also to be framed within those parameters. The end user is thus led to define their application within the five parameters and will in turn be led to specifications and other information most appropriate for their needs.

III. Evaluation and Optimization of Quality

Assessment principles for the maximization of task-based video quality are a relatively new field. Solutions have thus far been limited mainly to optimizing the network QoS parameters, and classical methods, like the peak signal-to-noise ratio (PSNR) [1] or structural similarity (SSIM) index [2], were applied. Some work has been devoted to developing new methods [6]. Problems of quality measurements for task-based video are partially addressed in a few preliminary standards and a Recommendation (ITU-T P.912, "Subjective Video Quality Assessment Methods for Recognition Tasks," 2008 [12]) 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 P.912, more research is necessary to validate the methods and refine the data analysis.

IV. Current Research

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.

The Public Safety Video Quality (PSVQ) project, undertaken by the Public Safety Communications Research (PSCR) program, sponsored by the U.S. Department of Homeland Security's Office of Interoperability and Compatibility, recently completed a subjective test of various levels of compression and resolution reduction following the methods suggested in ITU-T P.912 and the VQiPS GUCs [13]. 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 Figure 2.

Figure 2. User interface for subjective target recognition task test performed by PSCR.

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 "Intelligent information system supporting observation, searching and detection for security of citizens in urban environment" (INDECT) project, undertaken by the Seventh Framework Program, sponsored by the European Commission (Grant No. 218086), recently completed a subjective test of people's ability to recognize car registration numbers in video material recorded using a CCTV camera and compressed with the H.264/AVC codec [14].

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. ITU's Single Stimulus (SS) described in ITU-R BT.500-11 [8], was selected as the subjective test methodology [14]. An example of the user interface is shown in Figure 3.

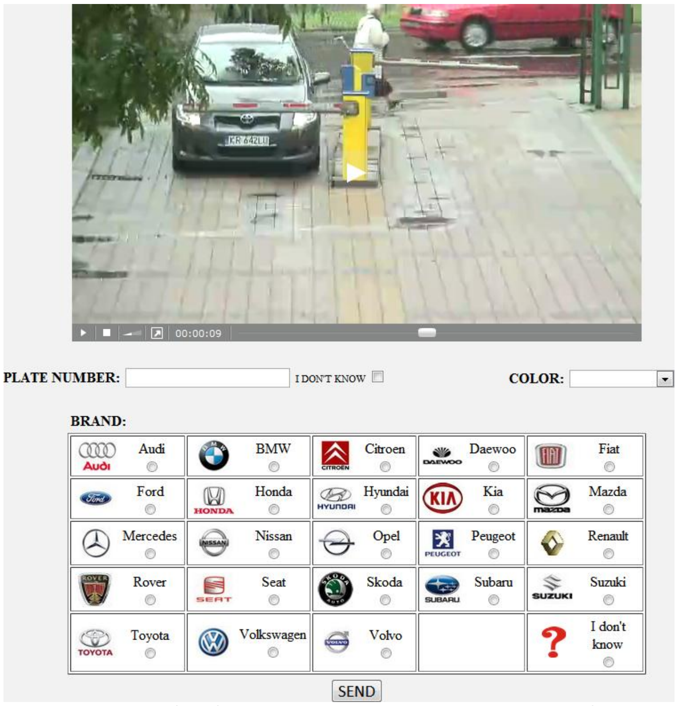

Figure 3. User interface for subjective plate recognition task test performed by FP7-INDECT (source: [16]).

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 [14].

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 [14].

V. Standardization

In addition to the user guide and the GUC framework it proposes, VQiPS is also working towards standards development. To that end, VQiPS works to coordinate the efforts of various organizations whose goal is to create standards for public safety video. VQiPS also seeks to create a consistent terminology of concepts related to the quality of video utility and related equipment. Some standardization activity has taken place within organizations directed at specific applications (in-car police cameras [17], transportation scenarios [18], etc.). The goal of the VQiPS project is to determine how this existing work can be applied to a broader class of applications, and to determine where gaps still exist in task-based video standards.

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 a recent meeting 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. The initiative is co-chaired by PSCR, and AGH University of Science and Technology in Krakow, Poland. PSCR's share of the work is funded by the U.S. Department of Homeland Security's Office for Interoperability and Compatibility via the Law Enforcement Standards Office in the National Institute of Standards and Technology. The AGH's share of the work is co-funded by the European Commission via the INDECT Project. Other members include 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 [6].

VI. Ethical Issues

Video surveillance naturally raises concerns of privacy. Numerous bodies have shown interest in protecting citizens against what could become an almost Orwellian-like "permanent surveillance." Among these, we should mention: Liberty Group, an Open Europe organization dedicated to human rights, the Electronic Frontier Foundation, and the Ethics Board of the INDECT project. The matter of citizens' security needs versus citizens' privacy was also one of the main themes of the Fourth Security Conference Research Conference organized by the European Commission in September 2009. In practice, alternative methods of at least partial protection of privacy do exist, based on the selective monitoring of figureheads, automatic erasure of faces/license plates not related to the investigation, or data hiding techniques using digital watermarking.

Those working on video quality tests, measurement methods and standards for public safety applications should be mindful of these concerns. The license plate recognition test described above, for example, ensured that the owners of the vehicles filmed were asked for their written consent, which allowed the use of the video content for testing and publication purposes [16].

VII. Conclusion and Future Research

Much research in this area is yet to be conducted. One important area of research that remains largely unexplored is determining appropriate measurement methods and thresholds to quantify the VQiPS GUCs. Along these lines, another area to be explored is the potential for automated classification of video into GUCs. As an example, quantifying the motion level in video using the Temporal Information algorithm recommended by ITU-T P.910 proves to be of little value when poor lighting conditions, which are also a factor in the GUCs, disguise the level of motion. Predictive relationships between any such automatic classification techniques and estimated recognition or acuity rates would be a highly desirable outcome of this type of research.

The next step is designing and conducting new research experiments based on these needs. Further studies are planned to examine object recognition given various motion, lighting and target size combinations processed by a number of Hypothetical Reference Circuits (HRC). The first set of collaborative tests is currently under development by PSCR and AGH. The set of parameters under investigation consists of the scene parameters (target size, lighting, and motion) as well as network parameters (e.g., bit-rate, compression ratio, transmission errors, and delay) for live/tactical applications, spatial scaling, and temporal scaling [6]. Among the first of these tests will be a study by PSCR to determine object recognition rates for recorded video, via a subjective test similar to the one described in this paper, but which allows the test subject to pause and replay the video.

REFERENCES

[1] A.M. Eskicioglu, P.S. Fisher, “Image quality measures and their performance,” Communications, *IEEE . Communications* 43(12), 2959-2965 (1995), http://dx.doi.org/10.1109/26.477498.

[2] Z. Wang, L. Lu, A.C. Bovik, “Video quality assessment based on structural distortion measurement,” *Signal Processing: Image Communication* 19(2), 121-132 (Feb. 2004), http://dx.doi.org/10.1016/S0923-5965(03)00076-6%20.

[3] S. Wolf, M. Pinson, NTIA Report TR-02-392, “Video Quality Measurement Techniques”, June 2002.

[4] M. Pinson, S. Wolf, “A New Standardized Method for Objectively Measuring Video Quality,” *IEEE Transactions on Broadcasting*, 50(12), 312-322, Sep. 2004.

[5] VQEG, “The Video Quality Experts Group,” http://www.vqeg.org/.

[6] 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, F. Misiorek, “D1.2, Report on NS and CS hardware construction,” P. Szczuko, P. Romaniak, M. Leszczuk (eds.), INDECT : Intelligent Information System Supporting Observation, Searching and Detection for Security of Citizens in Urban Environment, European Seventh Framework Programme FP7-218086-collaborative project, Europa : The INDECT Consortium, cop. 2010, 51 pages.

[7] “Alarm systems. CCTV surveillance systems for use in security applications,” EN 50132.

[8] ITU-R Recommendation BT.500-11, “Methodology for Subjective Assessment of the Quality of Television Pictures,” Geneva, 2002 (available at http://www.itu.int).

[9] ITU-T Recommendation P.910, “Subjective Video Quality Assessment Methods for Multimedia Applications,” Geneva, 1999 (available at http://www.itu.int).

[10] “Defining Video Quality Requirements: A Guide for Public Safety, Volume 1.0,” http://www.safecomprogram.gov/NR/rdonlyres/5BCA1CBF-1500-4B29-9370-81B823575DE8/0/3aVideoUserRequirementGuidedoc.pdf, July 2010.

[11] “Guide to Defining Video Quality Requirements,” http://www.pscr.gov/outreach/vqips/vqips_guide/define_vid_qual_reqs.php, available Jun. 2011.

[12] ITU-T Recommendation P.912, “Subjective Video Quality Assessment Methods for Recognition Tasks,” Geneva, 2008 (available at http://www.itu.int).

[13] “Video Quality Tests for Object Recognition Applications,” http://www.safecomprogram.gov/SAFECOM/library/technology/1627_additionalstatement.htm, Sep. 2010.

[14] M. Leszczuk, “Assessing task-based video quality | a journey from subjective psycho-physical experiments to objective quality models,” *Multimedia Communications, Services and Security 2011*, submitted for publication.

[15] ITU-T Recommendation P.800, “Methods for subjective determination of transmission quality,” Geneva, 1996 (available at http://www.itu.int).

[16] M. Leszczuk, L. Janowski, P. Romaniak, A. Glowacz, R. Mirek, “Quality Assessment for a License Plate Recognition Task Based on a Video Streamed in Limited Networking Conditions,” *Multimedia Communications, Services and Security 2011*, submitted for publication.

[17] International Association of Chiefs of Police, “Digital video systems minimum performance specifications document for in-car video camera systems performance specifications: digital video systems module,” Version 14 – 11/21/2008.

[18] The American Public Transportation Association, IT Standards program, “Technical recommended practice for the selection of cameras, digital recording systems, digital high speed train-lines and networks for use in transit related CCTV systems,” Recommended Practice, first document, Jun. 2007.