Redefining ITU-T P.912 Recommendation Requirements for Subjects of Quality Assessments in Recognition Tasks

Mikołaj I. Leszczuk[1], Artur Koń[1], Joel Dumke[2], and Lucjan Janowski[1]

[1] AGH University of Science and Technology, Department of Telecommunications, al. Mickiewicza 30, Kraków, Poland,
{leszczuk,janowski}@kt.agh.edu.pl, arturkon87@gmail.com
[2] National Telecommunications and Information Administration (NTIA), Institute for Telecommunication Sciences (ITS),
Boulder CO, USA
jdumke@its.bldrdoc.gov

Abstract. 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. Therefore it is crucial to measure, and ultimately, optimize task-based video quality. To develop accurate objective measurements and models for video quality assessment, subjective experiments must be performed. 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,") that mainly introduce basic definitions, methods of testing and requirements for subjects taking part in psychophysical experiments. Nevertheless, to the best of the authors' knowledge, the issue of requirements for subjects has been not verified in any specific academic research. Consequently, in this paper, we compare groups of subjects assessing video quality for task-based video. Once a comparison has been made for task-based video, specifications amendments for P.912 are developed. These will assist researchers of task-based video quality in identifying the subjects that will allow them to successfully perform the psychophysical experiment required.

Keywords: ITU-T, standards, systems, video, quality.

1 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

A. Dziech and A. Czyżewski (Eds.): MCSS 2012, CCIS 287, pp. 188–199, 2012.

largely separate from the objective video quality useful in computer vision [8]. Therefore it is crucial to measure and ultimately optimize task-based video quality. This is discussed in more detail in [9].

There exist 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.

Enormous work, mainly driven by the Video Quality Experts Group (VQEG) [12], 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 [12]. Unfortunately, many of the VQEG and ITU methods and recommendations (like ITU's Absolute Category Rating – ACR – described in ITU-T P.800 [2]) are not appropriate for the type of testing and research that task-based video, including closed-circuit television (CCTV), requires.

European Norm number 50132 [6] was created to ensure that CCTV systems are realized under the same rules and requirements in all European countries. 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 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.

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 [5], [3]. 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, "Subjective Video Quality Assessment Methods for Recognition Tasks," 2008 [4,7]) 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