A Framework for Generalizing Public Safety Video Applications to Determine Quality Requirements

Carolyn Ford and Irena Stange
Public Safety Communications Research project
US Department of Commerce
Boulder, Colorado, USA

Abstract—**Recognizing the growing importance of video in delivering a range of public safety services, the Video Quality in Public Safety (VQiPS) working group has been created under the U.S. Department of Homeland Security's Office for Interoperability (OIC), and the U.S. Department of Commerce's Public Safety Communications Research program (PSCR). VQiPS is focused on developing guidelines to assist state and local officials in the selection and procurement of video technology. Developing guidelines, including a framework for assessing an agency's video application and selecting technology related to its use, is critical, especially as millions of dollars in public monies are spent on video initiatives. The VQiPS project is working to fill that gap.**

Keywords-task-based video; video quality assessment

I. Introduction

Public safety agencies are increasingly using video as a major tool for accomplishing their goals. These agencies include, but are not limited to law enforcement, firefighting, emergency medical services, traffic and port authorities, prisons, and schools. Many agencies and organizations have hosted projects geared towards setting minimum performance specifications for the selection of video equipment for a specific application. These include, but are not limited to the U.S. Army [1], the US Intelligence community [2], the International Association of Chiefs of Police [3], United Kingdom Home Office [4], and the American Public Transportation Association [5]. Additionally, information has been developed in the research community regarding the effects of video system components on the delivered video quality [6], [7], and many different industry and governmental projects have developed a large number of recommendations, specifications, and standards. The large number of separate efforts has created a great deal of confusion among end users attempting to understand which specifications will help them meet their needs.

Each public safety agency may have one or more very specific video application. However, public safety video applications all have something in common at a higher level: performing a *recognition task* – enabling the user to recognize a desired target to a particular level of discrimination. Therefore, seemingly different applications may have similar quality requirements for video equipment. That is, upon closer examination, seemingly disparate video applications may actually have the same minimum requirements to perform their individual desired recognition tasks. If a correct framework is established, a method can be created for generalizing applications based on certain aspects that affect video quality requirements. This will greatly assist agencies in more quickly determining which recommendations, specifications, and standards apply to their specific application.

II. A Framework for Describing Video Applications

In support of creating such a framework, the Video Quality in Public Safety (VQiPS) working group (WG) was formed in February 2009 in an attempt to bring together past, current, and future public safety video quality projects. The aim was to leverage the work that has already been accomplished, establish commonalities between the applications, identify open research areas, and provide a central point from which to feed and disseminate information. The VQiPS WG was created under the U.S. Department of Homeland Security's Office for Interoperability and Compatibility (DHS/OIC) and the U.S. Department of Commerce's Public Safety Communications Research (PSCR) program. It includes local, state, and Federal representatives from a variety of disciplines, representatives from non-profit research organizations, academic institutions, and the video industry.

A. Scope

The term "video quality" can be interpreted in many different ways. For the purpose of the VQiPS project, video quality refers to the delivered visual intelligibility of the video, given a target of interest and desired discrimination level. The scope of the VQiPS effort is to map public safety applications to appropriate performance specifications for components of a video system that can affect or change the video stream. These components include lens configurations, image capture systems, video stream transport systems, video processing and storage systems, and displays.

In June 2010, VQiPS will release version 1.0 of the framework document for the WG, the *Public Safety Video System Procurement Guide*. The goal of this guide is to help match public safety video applications to the existing research and specifications that apply to them and, in the process, identify areas for which more research needs to be accomplished. Instead of attempting to provide individual answers to each and every application, this guide seeks to combine commonalities. Each application consists of a recognition task. The ability to recognize targets in video is affected, for the most part, by a few key parameters. This guide is intended to help the end users identify the values of these parameters for their applications and link them to existing standards that match these parameters. To reach this goal, the end user (or contractor or integrator) is led through the following steps:

1. Define the application in terms of several specific aspects, by answering a series of questions that lead to a generalized use class (GUC)
2. Understand the components of a video system and how each may affect quality
3. Receive qualitative guidance
4. Be directed to quantitative guidance (existing recommendations, specifications and standards) based on their GUC

This paper describes step 1.

Version 1.0 of the procurement guide is qualitative – it describes the framework and the concept of a GUC, and provides some general qualitative guidance. Future versions will contain links and references to quantitative specifications and standards.

B. The Generalized Use Class Concept

The first step to getting sufficient video quality is to clearly define the application's functional requirements, or "use case.[1]" At the most basic level, the requirements depend on the answers to two questions:

- *Use Characteristics*: What is the desired recognition task to be accomplished from viewing the video scene?
- *Scene Content*: What is in the scene of interest?

The answers to these two questions define the use case. All public safety video systems must present a scene of interest to a remote user with sufficient detail for the user to make informed decisions or to perform tasks. The end user must be able to, for example, read the characters in a license plate or determine the identities of individuals while performing surveillance.

The concept behind a GUC is that a few key aspects of the scene and the video's purpose can predict the majority of the quality needs for an application. Below are five key GUC aspects. Each can impact the quality of the video content as seen and used by the end user. For each parameter, a finite number of levels were chosen to attain a balance between completeness and manageability.

1) Use Characteristics

a) Discrimination Level (Recognition Task)

Video may be used for identifying a wide range of detail, from motion detection to positive identification of a person for forensic evidence. Not every video system needs to be capable of performing positive identification. The video system selected should conform to the quality requirements needed for the application, without over- or under- specifying the system.

Levels:

- General elements of the action: high-level description of the actions that took place
- Target class: large-scale detail recognition
- Target characteristics: medium-scale detail recognition
- Target positive ID: fine-scale detail recognition

b) Usage Timeframe

Video can be used for real-time applications, such as live surveillance or live tactical control of a situation. Video can also be recorded for later use, such as video used in its recorded version as evidence in a case. Recorded video carries with it the ability to rewind, freeze, zoom, and further process the video, while live video is used only in real-time. Video systems are often used for both recorded and real-time applications; in these cases, both applications should be examined for their implied requirements.

Levels:

- Real-time
- Recorded

2) Scene Content

a) Target Size

The size of the region of interest (target) with respect to the size of the field of view directly affects the ability to recognize that target. The larger the target, the more details can be discerned. Target size, as a percentage of field of view, is a direct function of the actual object size, distance from camera, and field of view of the camera.

Levels:

- Large (example, Fig. 1)
- Small (example, Fig. 2)

[1] In software and systems engineering, a "use case" is a description of a system's behavior in response to external stimuli. This technique is used to develop functional requirements by specifying the system's behavior through scenarios. This concept can be expanded to apply to video systems that are used to perform specific tasks.

Figure 1. Identifying a person: target size large.

Figure 2. Identifying a person: target size small.

b) Motion

Motion can originate from the target (e.g., a car driving by), the background (e.g., a large crowd), or from the camera itself moving (e.g., a dash-mounted camera in a police car). Motion affects the length of time a desired target is shown in the video frame, and can cause the target to blur.

Levels:

- Low
- High (example, Fig. 3)

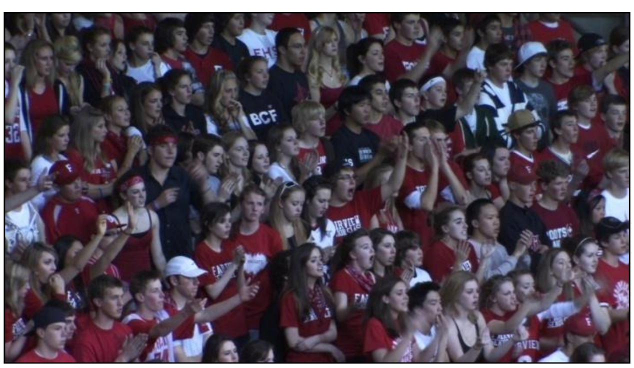

Figure 3. Sporting event crowd: a high-motion scene

c) Lighting Level

Lighting levels can vary from very dark (nighttime, indoors) to very bright (daylight, spotlight), affecting the ability of the camera to capture the image. The presence of both very bright areas and very dark areas in the frame simultaneously can impair target recognition, as shown in Fig. 4.

Levels:

- Bright
- Dim
- High dynamic range (Fig. 4)

Figure 4. Identifying a vehicle at night: variable lighting in scene

The combinations of the end user's specific needs in these five areas comprise the GUC. It should be emphasized that there are many, many more aspects that affect video, and each of these parameters have an infinite number of possible values. The VQiPS guide addresses the most important aspects, and defines a manageable number of choices for each one. A concerted effort was made to provide a balance between the competing needs of completeness and simplification.

Given the aspects, and their possible values given above, Fig. 5 illustrates the GUC determination process.

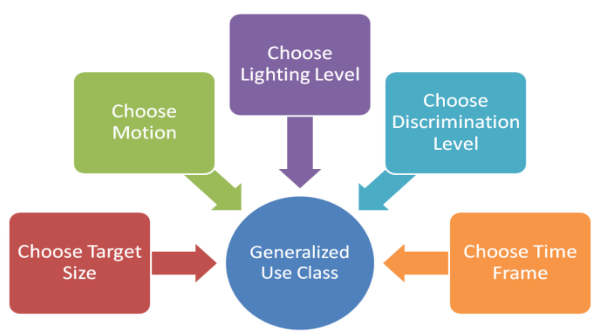

Figure 5. The GUC determination process

TABLE 1. SAMPLE APPLICATIONS

Scene Content	End Use	Actors	Sequence of Events	Target	Motion Sources	Lighting
License plate reading from a moving car	Accurately read a license plate of a moving car, from a moving car, at speeds of up to 70mph	Police in car, dispatcher, command & control	Police car follows target car, back of target car is filmed, acquisition of license plate is attempted in real-time	Alphanumeric character	Target, camera	Daytime
Structural smoke reading from a helicopter	Determine color, velocity of smoke plumes	Firefighter in air, command & control	Helicopter circles building looking for smoke plumes, sends footage back to command & control	Smoke plume	Target, camera	Daytime
Store surveillance as evidence	Count the number of people in the scene	Forensic analysts	People walk through a store during a period of 2 minutes	Bodies	People at walking speed	Indoor

C. *Examples*

This section describes three examples for determining GUCs for specific applications. The first two examples demonstrate how two seemingly different applications may actually have the same video quality needs, because they belong to the same GUC. The third demonstrates a different resulting GUC.

- Application 1: Police in-car camera used to read license plates during the day, in real time
- Application 2: Daytime footage from a helicopter circling a fire, used to recognize smoke features from small plumes (velocity, color, etc.) in real time
- Application 3: Store surveillance footage used as evidence to count the number of people in an aisle during an incident

Table 1 breaks out these three applications as use cases, and identifies the applicable targets and scene conditions.

For the first two use cases, Fig. 6 shows the choices (indicated in yellow) that would be made from the GUC selection process:

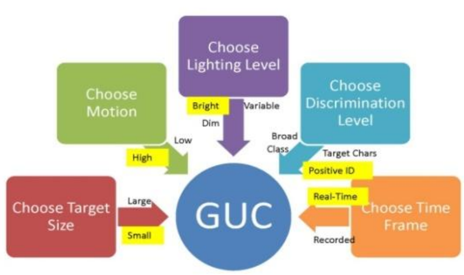

Figure 6. Example license plate identification and smoke reading GUC.

The third example use case belongs to a different class, as shown in Fig 7.

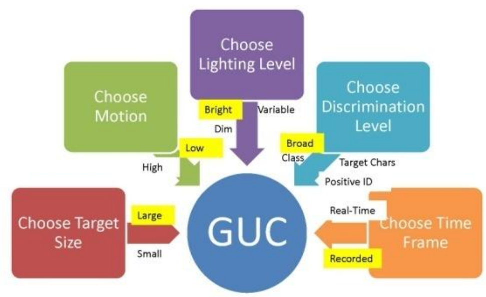

Figure 7. Example store surveillance GUC.

III. CONCLUSIONS

The VQiPS guide provides a centralized resource for any agency or organization that uses video to perform recognition tasks. It assists the end user in describing their application in terms of parameters that have the most effect on the video system quality requirements, and provides qualitative guidance and links to quantitative specifications results that are appropriate for their application. As research continues, the guide will be a central clearinghouse as such specifications and standards are developed.

IV. FUTURE RESEARCH

One goal of the VQiPS WG effort is to collect the results of previous efforts together, leveraging the research to explore relationships between size, motion and lighting for target recognition at different discrimination levels, the effects of video system component characteristics on these relationships, and to map the various results to the GUC framework. An additional goal is to identify gaps in the spectrum of available standards and specifications, in order to direct the work of the PSCR's Video Quality project to develop more where needed.

The next step towards the development of quantitative definitions for the GUCs is the study of interactions between the scene content aspects (lighting, motion, and target size) within the GUCs. It is expected that recognition ability will be a continuous function of these aspects; however, a quantization of the aspects into a finite number of Use Classes, as depicted in Section II, will be easier for the end users to characterize their applications. Guidance will be established through testing for establishing those ranges.

The first phase of testing will objectively explore the relationships between the three scene content aspects. Testing will be done using high-quality, uncompressed video. Bar-type resolution charts will be used to characterize the smallest resolvable object size in a test scene. The test target will be recorded as it is moving at a number of known velocities, including a static condition. This will be repeated for several lighting levels ranging from dim to well-lit and including a high dynamic range condition. The smallest resolvable target size, measured in pixels, will be identified for each of the various target motion and target illumination levels. The resulting data should provide guidance for establishing quantitative values for the GUCs, as well as guidance for a second, subjective phase of testing.

This second phase of testing will involve subjective testing using appropriate experienced viewers (e.g., viewers trained in forensic video, viewers trained in smoke reading). This testing can also start to quantify the individual effects of degradations to the video stream introduced by the video system components. An initial experiment is briefly described below.

Test scenes will be filmed in a manner similar to the first phase of testing. The testing will follow the ITU P.912 standard [8], and will focus on the most stringent use characteristics: positive identification in real time. The multiple choice test method will be used, where viewers will be asked to identify objects given a number of choices. Similarly sized objects will be used, and may include:

- Guns
- Knives
- Police radios
- Cell phones
- Wallets
- Bottles

The objects will be filmed at two different speeds. Each scene will be filmed with the camera at two different distances from the target. The distances will be chosen in order for the apparent size of the target to be a predetermined size in pixels, chosen based on the outcome of the first phase test. Additionally, there will be three lighting conditions. The clips will be down-converted to standard definition (SD) video format and will be impaired as they are streamed across a laboratory network. The data gathered will measure the ability of the viewer to correctly identify the target in the scene, and the percentage of correct responses will be analyzed for each of the conditions of interest. The outcome of the research should begin to establish predictive relationships between target size, motion, and lighting and the ability of an expert viewer to detect targets within video that has undergone various levels of degradation.

The collaborators within VQiPS will be undertaking studies to determine network artifact effects on the three relationships above. For other components of the video system, research will be encouraged and results reported in the VQiPS User's Guide, which will be a continually updated resource as research progresses.

ACKNOWLEDGMENT

Funding for the VQiPS conference and the VQiPS Working Group was provided jointly by the Department of Homeland Security Office of Interoperability and Compatibility, and the Department of Commerce's Public Safety Communications Research (PSCR) program in Boulder, Colorado. The work was initiated at the VQiPS conference, hosted by the PSCR in partnership with DHS/OIC.

The authors wish to thank the VQiPS WG member organizations who all participated in the development of the materials presented. Membership information, along with ongoing project information in the VQiPS WG can be found at http://www.its.bldrdoc.gov/psvq/video_quality/video_about.php.

REFERENCES

[1] S. K. Moyer, E. Flug, T. C. Edwards, K. A. Krapels, and J. Scarbrough, "Identification of handheld objects for electro-optic/FLIR applications," Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XV, Proc. SPIE 5407, 116–126, Aug. 2004.

[2] L. P. Colburn ., J. M. Irvine, J. C. Leachtenauer, W. A. Malila, and N. L. Salvaggio, "The general image quality equation (GIQE)", Proceedings of the American Society of Photogrammetry and Remote Sensing Annual Meetings, April 1996.

[3] 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.

[4] N. Cohen, J. Gattuso, and K. MacLennan-Brown, CCTV Operational Requirements Manual 2009, Publication No. 28/09, Home Office Scientific Development Branch, United Kingdom.

[5] 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.

[6] M. Duplaga, M. I. Leszczuk, Z. Papir, and A. Przelaskowski , "Compression evaluation of surgery video recordings retaining diagnostic credibility (compression evaluation of surgery video)," Opto-Electronics Review, Stowarzyszenie Elektryków Polskich, Wojskowa Akademia Techniczna. Warszawa, ISSN 1230-3402, 2008 vol. 16 no. 4 s. 428–438.

[7] "Statement of Requirements for Public Safety Communications Interoperability", http://www.safecomprogram.gov/SAFECOM/library/technology/1258_statementof.htm

[8] ITU-T Recommendation P.912, "Subjective video quality assessment methods for recognition tasks," Recommendations of the ITU, Telecommunication Standardization Sector, August 2008.