Task-Based Tactical and Surveillance Video Quality Tests

Homeland Security

DHS-TR-PSC-10-07

Department of Homeland Security

Public Safety Communications

Technical Report

This page intentionally left blank.

Command, Control and Interoperability:

Office for Interoperability and Compatibility

Defining the Problem

Emergency responders—police officers, fire personnel, emergency medical services-need to share vital voice and data information across disciplines and jurisdictions to successfully respond to day-to-day incidents and large-scale emergencies. Unfortunately, for decades, inadequate and unreliable communications have compromised their ability to perform mission-critical duties. Responders often have difficulty communicating when adjacent agencies are assigned to different radio bands, use incompatible proprietary systems and infrastructure, and lack adequate standard operating procedures and effective multi-jurisdictional, multi-disciplinary governance structures.

OIC Background

The Department of Homeland Security (DHS) established the Office for Interoperability and Compatibility (OIC) in 2004 to strengthen and integrate interoperability and compatibility efforts to improve local, tribal, state, and Federal emergency response and preparedness. Managed by the Science and Technology Directorate within the Command, Control and Interoperability Division, OIC helps coordinate interoperability efforts across DHS. OIC programs and initiatives address critical interoperability and compatibility issues. Priority areas include communications, equipment, and training.

OIC Programs

OIC programs address voice, data, and video interoperability. OIC is creating the capacity for increased levels of interoperability by developing tools, best practices, technologies, and methodologies that emergency response agencies can immediately put into effect. OIC is also improving incident response and recovery by developing tools, technologies, and messaging standards that help emergency responders manage incidents and exchange information in real time.

Practitioner-Driven Approach

OIC is committed to working in partnership with local, tribal, state, and Federal officials to serve critical emergency response needs. OIC's programs are unique in that they advocate a "bottom-up" approach. OIC's practitioner-driven governance structure gains from the valuable input of the emergency response community and from local, tribal, state, and Federal policy makers and leaders.

Long-Term Goals

Long-term goals for OIC include:

- Strengthen and integrate homeland security activities related to research and development, testing and evaluation, standards, technical assistance, training, and grant funding.
- Provide a single resource for information about and assistance with voice and data interoperability and compatibility issues.
- Reduce unnecessary duplication in emergency response programs and unneeded spending on interoperability issues.
- Identify and promote interoperability and compatibility best practices in the emergency response arena.

This page intentionally left blank.

Public Safety Communications Technical Report

Task-Based Tactical and Surveillance Video Quality Tests

DHS-TR-PSC-10-07
July 2010

Reported for: **The Office for Interoperability and Compatibility by NIST/OLES**

Homeland Security

This page intentionally left blank.

Publication Notice

Disclaimer

The U.S. Department of Homeland Security's Science and Technology Directorate serves as the primary research and development arm of the Department, using our Nation's scientific and technological resources to provide local, state, and Federal officials with the technology and capabilities to protect the homeland. Managed by the Science and Technology Directorate, the Office for Interoperability and Compatibility (OIC) is assisting in the coordination of interoperability efforts across the Nation.

Certain commercial equipment, materials, and software are sometimes identified to specify technical aspects of the reported procedures and results. In no case does such identification imply recommendations or endorsement by the U.S. Government, its departments, or its agencies; nor does it imply that the equipment, materials, and software identified are the best available for this purpose.

Contact Information

Please send comments or questions to: S&T-C2I@dhs.gov

This page intentionally left blank.

Contents

This page intentionally left blank.

Abstract

This report describes laboratory studies of the quality required for the following public safety video applications:

- Tactical
- Live surveillance
- Recorded surveillance

Requirements for these applications are based on the studies described here, and are given in Section 4 of the Public Safety Statement of Requirements (PS SoR) Volume II, Version 1.2 [1].

The tests described in this report were executed as task-based subjective tests, following the test methods described in ITU-T Recommendation P.912 [2].

Key words: video quality, subjective test methods

1 Introduction

A previous technical report [3] describes laboratory studies, referred to as PS1–PS3, which investigated the level of quality required for the following public safety video applications:

- Narrow field of view, tactical
- Wide field of view, tactical
- Narrow field of view, live surveillance
- Wide field of view, live surveillance
- Narrow field of view, recorded surveillance
- Wide field of view, recorded surveillance

Requirements for these applications are based on the studies described in [3] and are given in Section 4 of the Public Safety Statement of Requirements (PS SoR) Volume II, Version 1.1 [4].

The PS1–PS3 studies addressed generalized perceptual quality reporting. Viewers were asked to report a level of quality (Excellent to Bad), based on ITU Recommendations for video quality assessment [5] [6], and indicate their opinion of whether or not the quality was sufficient to do their job. These tests focused on tactical and surveillance video applications with a narrow, wide, or normal field of view. The tests investigated a large variation of quality to narrow the range for subsequent testing.

The next phase of testing for the PS SoR shifted the focus of quality reporting from a subjective impression to a task-based format. The goals of this shift were (1) to closely match the actual use of the video applications and (2) narrow the list of requirements from the previous recommendations. Public safety applications necessitate that the video quality is sufficient to recognize people, lettering, or other objects in the video content, regardless of the subjective perceptual impression of quality. Therefore, a series of tests was designed to determine what degree of video quality was sufficient to perform specific recognition tasks. This task-based phase included two subjective tests:

- PS4 covered live surveillance applications

- PS5 covered recorded surveillance applications

2 Experimental Method

The methods used in PS4 and PS5 followed [2]. Viewers were shown video clips at varying quality levels and asked to perform specific tasks, using either the multiple-choice or the single-answer method. The scenes were presented in a manner consistent with the intended application:

- PS4: For live applications, the viewer watched the video clip only once before performing the recognition task.
- PS5: For recorded applications, the viewer could replay the video clip as many times as he/she preferred, and perform frame freezes.

Table 1 shows the parameters and levels used during the study. The parameter value names, Px.y, are defined individually for each test, PS4 (Section 3) and PS5 (Section 4).

Table 1: Test parameters for PS4 and PS5

Parameters for Each Test	Levels	Level Names
Target size (Section 2.1.2.1)	2	Small, Large
Scene complexity (Section 2.1.2.2)	2	Low, High
Packet Loss (Section 2.1.2.3)	4	P1.1, P1.2, etc.
Compression (H.264) (Section 2.1.2.3)	3	P2.1, P2.2, etc.

The video clips included several scenarios. A scenario is directions for the actions and contents of a scene (e.g., man walks by camera carrying an object). Each scene was organized into a scenario group. A scenario group is a collection of scenes of the same basic scenario, with very slightly controlled differences between the scenes.

2.1 Scenes

The target item in a scene is the area within the video frame that must be recognized to perform the application task (e.g., face, alphanumeric characters, or object). Video analysts use target-recognition video (TRV) to accomplish a specific goal by recognizing specific targets of interest in a video stream. Since video analysts generally use TRV to perform a recognition task, the scenes created for the tests contained targets that were consistent with the application being studied.

2.1.1 Scenario Groups

Because the test measurements focused on a viewer’s ability to identify objects and actions, the test’s design had to address the possibility that a viewer may memorize the scene content and use other visual clues to remember the identity of the target. Therefore, the tests replaced individual scenes with a set of scenes (i.e., a scenario group) containing multiple versions, with controlled differences between the versions. The number of scenes in a scenario group had to be large enough to prevent scene memorization. Figure 1 shows an example of two scenes from one scenario group. This scenario group featured a person walking across the field of view carrying an object. Within the same field of view, some scenes showed the same person walking but carrying different objects, while other scenes showed a different person walking.

In this way, the scenario group consisted of multiple shots. In the examples in Figure 1, the object being carried in the left scene is a video camera, and the object being carried in the right scene is a can. The scene content is almost identical except for the single change of the object the person is holding.

Figure 1: Two example scenes from one scenario group

2.1.2 Scene Parameters

The application of TRV is directly related to the task being performed and the need of the user to recognize targets at different levels of detail-referred to as discrimination levels. A discrimination level is one of four levels of visual recognition at which a target can be analyzed [7]:

- General Elements of the Action: High-level description of the actions that took place
- Target Class: Large-scale detail recognition (e.g., person, car, type of object)
- Target Characteristics: Medium-scale detail recognition (e.g., gender, markings, scars, tattoos, dents, color)
- Target Positive Recognition: Fine-scale detail recognition (e.g., recognition of a person, specific object, or exact alphanumeric sequence)

In most applications, for a given a discrimination level, two factors will have the most effect on the ability to recognize targets in a video scene: target size and scene complexity.

2.1.2.1 Target Size

The ability to recognize a target at a particular discrimination level may correlate to the relative size of the target within the video frame. To study that assumption, scenario groups were designed to contain both large and small targets.

Large and small target size categories were created by counting the number of pixels for each target that the viewer was asked to identify. The two categories are separated by a threshold indicating a significant grouping between the larger and smaller targets.

Testing illustrated a marked difference in the performance requirements for targets that occupy fewer than 0.3 percent of the total pixels in the video frame, and those that occupy greater than 0.3 percent of total pixels. Therefore, the requirements for the video applications described here are specified for two cases, which appear in Sections 3.3 and 4.3:

- Large target (greater than 0.3 percent of total pixels)
- Small target (less than 0.3 percent of total pixels)

2.1.2.2 Scene Complexity

The specific definition of "complexity" applies to the effect of the homogeneity of the spatial and temporal video information on the parameters being studied. For example, if camera optics are being studied, the dynamic range of color and light levels may comprise the "complexity" of the scene. If compression is under study, scene variance may be the "complexity."

For PS4 and PS5, complexity was a function of motion and small details in the video content. The complexity score is a measure of object motion, diverse clutter, and spatial frequency information in a video scene or clip. The complexity score indicates how much processing the video coder/decoder (codec) will have to perform to encode and decode the video clip. A high complexity rating means that a scene or clip contains multiple independent motions, clutter, and high spatial frequency information.

For each scene, the median complexity score was computed based on the data available. A complexity score ranked "low" when the value was less than or equal to the median complexity score. A complexity score ranked "high" when the value was greater than the median complexity score.

2.1.2.3 Compression and Packet Loss

The network parameters studied for PS4 and PS5 included compression and packet loss. These parameters were chosen because of their influence on the quality of digital video when sent across a network. H.264 was the only compression method used in PS4 and PS5 because an earlier study [3] showed it was approximately twice as efficient as MPEG-2.

2.2 Clip Creation

The test clips were created and impaired by taking the following steps: filming the clips in high definition (HD) video format with a frame size of 1920 x 1080 pixels and a frame rate of 30 frames per second (fps); down-converting the HD clips to standard definition (SD) video format with a frame size of 720 x 486 pixels; and impairing the SD video by forcing various values for the encoder bit rate and the packet loss rate while the video streamed across a laboratory network. For consistency, the frame rate was kept constant at 30 fps.

2.3 Viewer Response

The viewers were shown short video clips and asked to perform a specific recognition task, using either a multiple-choice format or single-answer entry format.

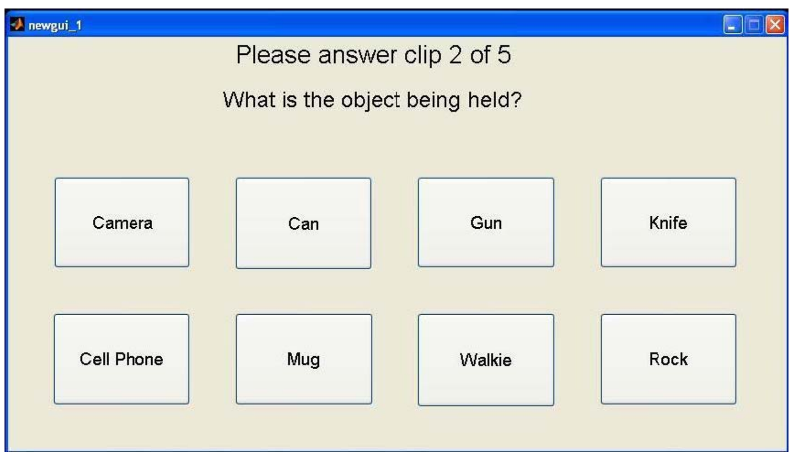
2.3.1 Multiple Choice

For the multiple-choice method, a clip from a particular scenario group was shown above a list of written labels representing the possible answers. After presenting the video, the viewer was asked to choose the label closest to what they recognized in the clip.

The number of choices offered to the viewer depended on the number of alternative scenes within the scenario group. The test did not offer an “Unsure” optional answer [8]. Future tests may benefit from the inclusion of a certainty rating option for the viewer. Figure 2 shows an image capture from an object identification scene, while Figure 3 shows an input screen example for the multiple-choice method.

Figure 2: Image capture example from object identification scene

Figure 3: Example input screen for multiple-choice method



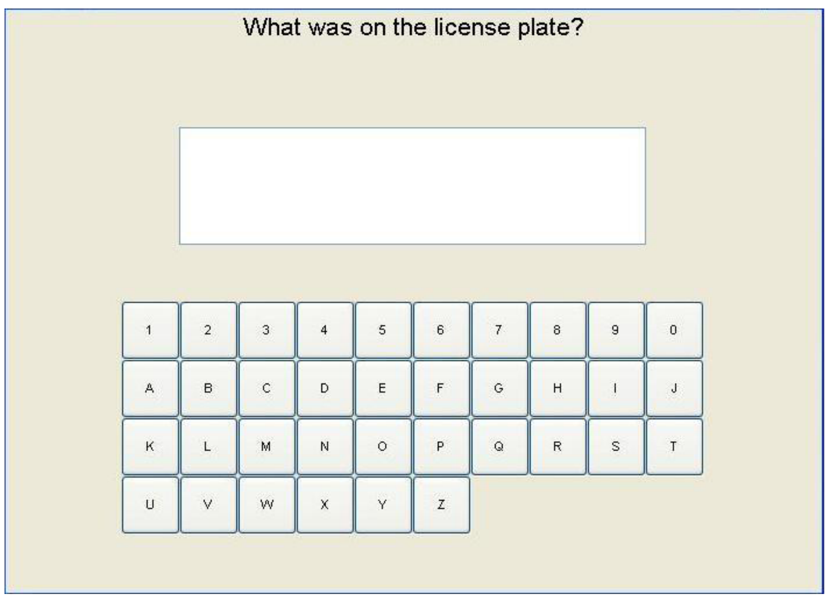
2.3.2 Single Answer

For recognition tasks that involve a non-ambiguous answer to an identification question for a given scenario group (e.g., alphanumeric-character recognition scenarios), the test used the single-answer method. For this method the viewer was asked what letters or numbers are present in a specific area of the video. The answer was evaluated as either correct or incorrect.

Figure 4 shows an image capture from an alphanumeric-identification scene, while Figure 5 shows an example of an input screen using the single-answer method.

Figure 4: Image capture example from alphanumeric identification scene

Figure 5: Example input screen for alphanumeric single-answer method



2.4 Data Analysis

Because test viewers were likely to guess, each score was normalized for the probability of a correct guess based on the equation:

$$R_A = R - \frac{W}{n-1}$$

where R_A is the adjusted number of right answers, R is the number of right answers, W is the number of wrong answers, and n is the number of answer choices [9].

The PS4 and PS5 experiments were designed so that the results could be reported as percentages of questions answered correctly, or the mean number of targets identified for clips with multiple targets present (e.g., multiple characters of an alphanumeric sequence). Viewers were considered to be interchangeable samples. PS4 calculations included "loss of detection ability" as an additional factor. This was calculated by subtracting the viewers' scores for a particular scene with a particular impairment level from the viewers' scores for the same scene, but unimpaired. The difference between the scores is equivalent to the "loss of detection ability." This data provides additional information about whether the network conditions impaired the viewers' detection ability.

3 Live Surveillance Test (PS4)

The application focus of PS4 was live surveillance. Seven different scenario groups were used. Public safety practitioners (viewers) were shown several variations from each scenario group and asked to perform specific recognition tasks. Viewers viewed each scene just once in real time before they were asked to make the requested identification.

3.1 PS4 Scenario Group Summary

The test used seven scenario groups for PS4, with three to nine variations (i.e., scenes) for each scenario group. The total number of scenes was 45. Table 2 provides descriptions of the groups.

Table 2: Scenario groups for PS4

		Complexity[a]			
Scenario Group: **Number of Variations:**	**Size[b]**	**Low**	**High**	**Discrimination Level:** **Experimental Method:**	**Scene Description:**
Scenario Group: Object thrown from stationary car **Number of Variations:** 9	**Small:** **Large:**	6 3	N/A[c] N/A	**Discrimination Level:** Target Positive Recognition **Experimental Method:** Multiple Choice	A car is standing still with the camera positioned behind it. An object is thrown from the car. Viewer identifies the object, given a number of choices.
Scenario Group: Moving car **Number of Variations:** 4	**Small:** **Large:**	N/A N/A	4 N/A	**Discrimination Level:** Target Positive Recognition **Experimental Method:** Single Answer	A car is moving, with the camera in a following car. Viewer enters the license plate number.
Scenario Group: Car at toll booth **Number of Variations:** 9	**Small:** **Large:**	3 N/A	6 N/A	**Discrimination Level:** Target Positive Recognition **Experimental Method:** Single Answer	A car drives up to a toll booth monitored by the camera. Viewer enters the license plate number.
Scenario Group: Person walking, holding an object **Number of Variations:** 7	**Small:** **Large:**	7 N/A	N/A N/A	**Discrimination Level:** Target Positive Recognition **Experimental Method:** Multiple Choice	A person walks past the camera holding an object. Viewer identifies the object, given a number of choices.

Table 2: Scenario groups for PS4 (Continued)

Scenario Group: Number of Variations:	Size[b]	Complexity[a] Low	Complexity[a] High	Discrimination Level: Experimental Method:	Scene Description:
Scenario Group: Person walking, holding an object **Number of Variations:** 7	**Small:** **Large:**	7 N/A	N/A N/A	**Discrimination Level:** Target Positive Recognition **Experimental Method:** Multiple Choice	A person takes an object out of a backpack and holds it briefly. Viewer identifies the object, given a number of choices.
Scenario Group: Hazardous materials sign **Number of Variations:** 6	**Small:** **Large:**	2 N/A	3 1	**Discrimination Level:** Target Positive Recognition **Experimental Method:** Single Answer	A hazardous materials sign is affixed to the back of a stationary truck. The camera zooms in on the sign in some scenes. Viewer enters the four numbers on the hazardous materials sign.
Scenario Group: Robot search **Number of Variations:** 3	**Small:** **Large:**	N/A N/A	N/A 3	**Discrimination Level:** Target Positive Recognition **Experimental Method:** Multiple Choice	A camera mounted on a robot moves through a cluttered room. Viewer selects the time on a clock in the room as the robot moves by, given a number of choices.

a. Scene Complexity arguments are “Low” or “High”
b. Target Size arguments are “Small” or “Large” for which High/Low sums equal Number of Variations
c. N/A means not applied

3.2 Experimental Design

Two types of codec systems were used. One type performed no special processing on video streams to compensate for packet loss. Two different manufacturers’ codecs of this type were used (each approximately half of the time). A second type employed a proprietary algorithm (generically known as Error Concealment (EC)) to intelligently fill in missing information from video streams with packet loss. Table 3 shows the network conditions (i.e., combinations of compression bit rates and packet loss) that were applied to each scene.

Table 3: Network conditions: compression and packet loss combinations

System\ bit rate (kbps)	512	256
With EC (percent packet loss)	15.0	5.0, 0.0
Without EC (percent packet loss)	0.5	0.1, 0.0

Certain scenario groups had additional network conditions applied to each scene variation they contained. Table 4 summarizes these additional network conditions.

Table 4: Additional network conditions tested for selected scenario groups

Scenario Groups	■ Car at toll booth ■ Person walking ■ Person standing		■ Car at toll booth	
System\ bit rate (kbps)	512	384	512	384
With EC (percent packet loss)	10.0, 5.0	0.0	0.0	N/A
Without EC (percent packet loss)	0.1, 0.0	0.0	N/A	0.2

The number of scenes tested for a particular network condition and scene parameter combination varied. Tables 5 and 6 summarize these combinations. Refer to these tables for additional information when reviewing the results presented in Section 3.3. As can be calculated from Tables 5 and 6, the total number of impaired scenes with error concealment presented to viewers is 69, and the total number of impaired scenes without error concealment presented to viewers is 144. Viewers also saw all 45 scenes unimpaired, and had two training sessions (ten questions for the multiple-choice format and six questions for the alphanumeric format) to practice using the test software. Each viewer examined 274 scenes, of which 258 determined the scores and analysis, and 16 were used for training purposes only. The test randomized presentation order for each viewer.

Table 5: Summary of number of scenes tested per network condition without error concealment

Percent Packet Loss	**at 512 kbps**	0.5		0.1, 0.0		N/A	
	at 384 kbps	N/A		0.0		0.2	
	at 256 kbps	0.1, 0.0		N/A		N/A	
Complexity		**Low**	**High**	**Low**	**High**	**Low**	**High**
Large Object	**Multiple Choice**	2	2	N/A	N/A	N/A	N/A
	Numeric Entry	N/A	1	N/A	N/A	N/A	N/A
Small Object	**Multiple Choice**	14	N/A	10	N/A	N/A	N/A
	Alphanumeric Entry	2	6	2	4	2	4
	Numeric Entry	1	2	N/A	N/A	N/A	N/A

Table 6: Summary of number of scenes tested per network condition with error concealment

Percent Packet Loss	**at 512 kbps**	15.0		10.0, 5.0		0.0	
	at 384 kbps	N/A		0.0		N/A	
	at 256 kbps	5.0, 0.0		N/A		N/A	
Complexity		**Low**	**High**	**Low**	**High**	**Low**	**High**
Large Object	**Multiple Choice**	1	N/A	N/A	N/A	N/A	N/A
Small Object	**Multiple Choice**	6	N/A	4	N/A	N/A	N/A
	Alphanumeric Entry	1	4	1	2	1	2
	Numeric Entry	1	1	N/A	N/A	N/A	N/A

3.3 Results

Fourteen viewers completed the test,[1] viewing all of the clips in the test design. Correct answers were tabulated and grouped according to the four parameters under study (target size, scene complexity, compression, and packet loss). Sections 3.3.1 and 3.3.2 show the results graphically.

For single-answer-method questions, where each alphabetic or numeric character was considered to be one target to identify, the results are presented as the mean number of targets identified out of the total number of targets present in each clip. Ninety-five percent confidence intervals were calculated assuming a normal distribution.

For the multiple-choice questions, the results are presented as the total number of questions correctly answered. Ninety-five percent confidence intervals were calculated using the normal approximation.[2]

The results for the cases including EC show a significant improvement over the no-EC cases. However, the algorithm used was proprietary so the results should not be expanded to include all EC cases.

3.3.1 Large Target Size

3.3.1.1 Targets Identified

The results for multiple-choice and numeric-entry questions are calculated as the percentage correct for multiple-choice questions, and the mean number of targets identified out of the four targets presented in the numeric-entry clips. Figure 6 shows that for large targets without error concealment at a 512 kbps bit rate, the recognition level is high even with a packet loss rate that is the highest level tested (when error concealment is not present; 0.5 percent). The high-complexity numeric-answer results at 256 kbps are counter-intuitive, and merit further study. If there is an apparent absence of confidence intervals in these figures, this indicates there was no variation in the scores.

1. Two additional viewers took portions of the test but were not able to complete it. Their scores were similar to viewers who completed the entire test, and are not included in the data presented here.
2. The data followed the conditions $np - \sqrt[2]{npq}$ and $0 < np + \sqrt[2]{npq} < n$, where n is the number of questions, p is the number of questions answered correctly and q is the number of questions answered incorrectly [10].

Figure 6: Percent targets identified, target size = large, without error concealment (four graphs)

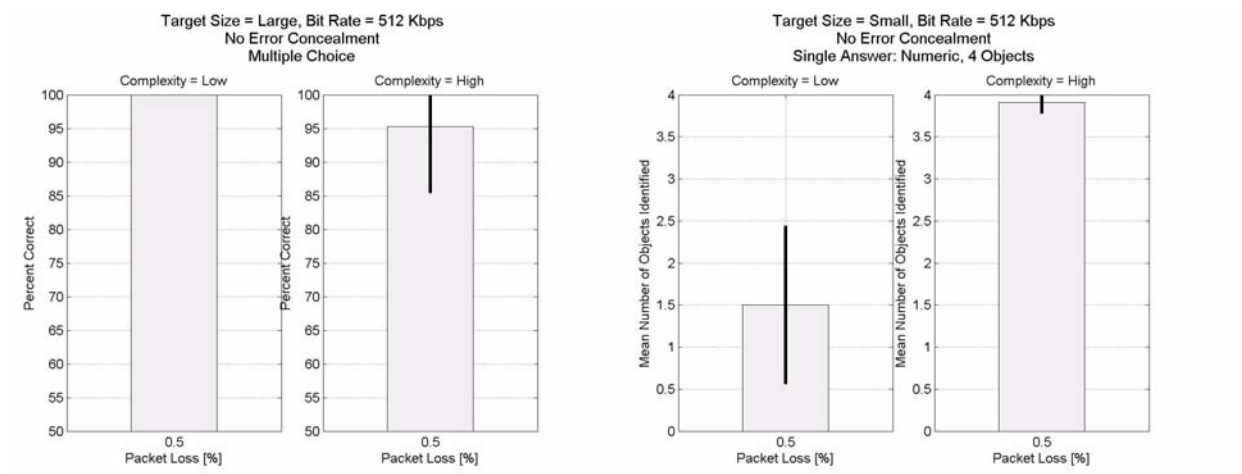
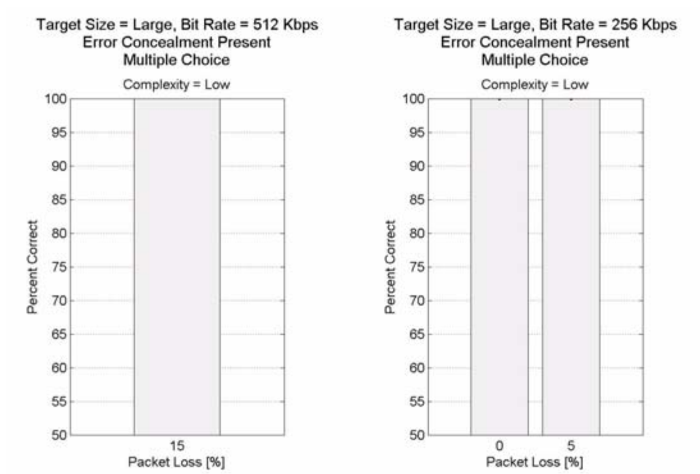


Figure 7 shows that all viewers answered all multiple-choice questions correctly, given a large target size at the three bit rate/packet loss combinations tested when error concealment was present. The test did not include high-complexity scenes, therefore all data reflects low-complexity scenes.

Figure 7: Percent targets identified, target size = large, with error concealment (two graphs)



3.3.1.2 Loss of Detection

Table 7 presents the calculated loss of detection ability data (see Section 2.4) for large target sizes without error concealment for multiple-choice and numeric-entry questions. Scores represent a percentage lost for multiple choice and the mean number of objects lost for numeric entry.

Table 7: Loss of detection ability, target size = large, without error concealment

		Multiple Choice [Percent]		Numeric [Mean Number of Objects / 4]
	Complexity	**Low**	**High**	**High**
Bit Rate	**Percent Packet Loss**			
512 kbps	**0.5**	0.00	5.35	0.00
256 kbps	**0.0**	5.36	0.00	2.05
	0.1	5.36	5.36	0.00

For the error concealment case, Table 8 shows the loss of detection ability as zero for all cases, as would be expected from results presented in Figure 7.

Table 8: Loss of detection ability, target size = large, with error concealment

		Multiple Choice [Percent]
	Complexity	**Low**
Bit Rate	**Percent Packet Loss**	
512 kbps	**15.0**	0.0
256 kbps	**0.0**	0.0
	5.0	0.0

3.3.2 Small Target Size

3.3.2.1 Targets Identified

Figure 8 shows the effect of complexity on the ability to perform recognition tasks for small targets, given an alphanumeric-recognition task. This effect of complexity on viewers is most clearly observed at the 256 kbps bit rate. The multiple-choice data contains low-complexity scenes only and it shows a consistent slight decline as network conditions are further impaired. The numeric-detection task shows counter-intuitive results, suggesting an error in testing or scene selection.

For low-complexity scenes with multiple-choice recognition tasks, viewers achieved high correct response levels, even at the largest impairment level tested: 256 kbps compression at 0.1-percent packet loss. For alphanumeric tasks, compression of 512 kbps showed a marked decline at 0.5 percent for high-complexity scenes versus low-complexity scenes. At 256 kbps, there is a marked decline for high complexity at 0.0-percent and 0.1-percent packet loss as well.

Figure 8: Percent targets identified, target size = small, without error concealment (eight graphs)

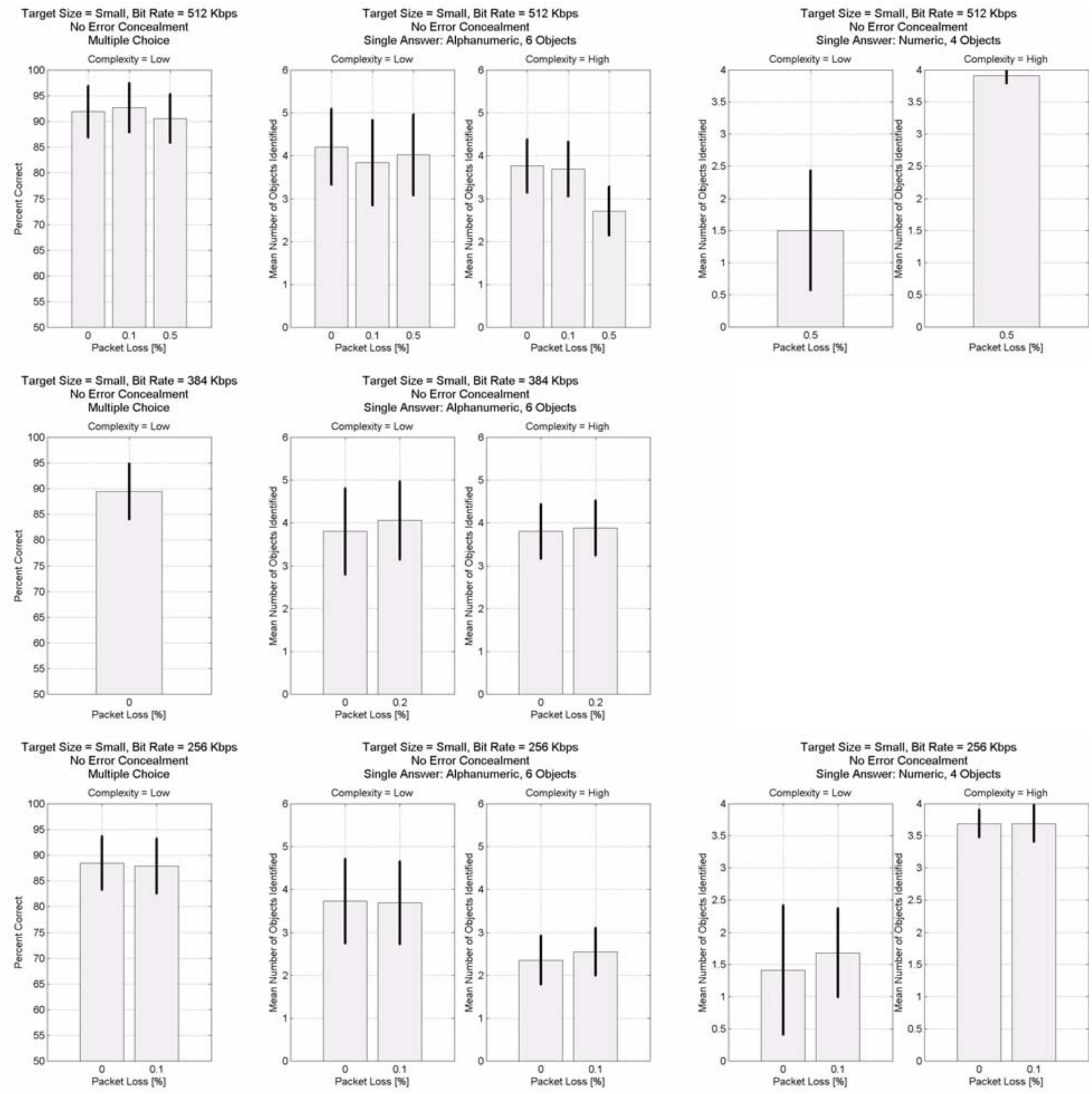
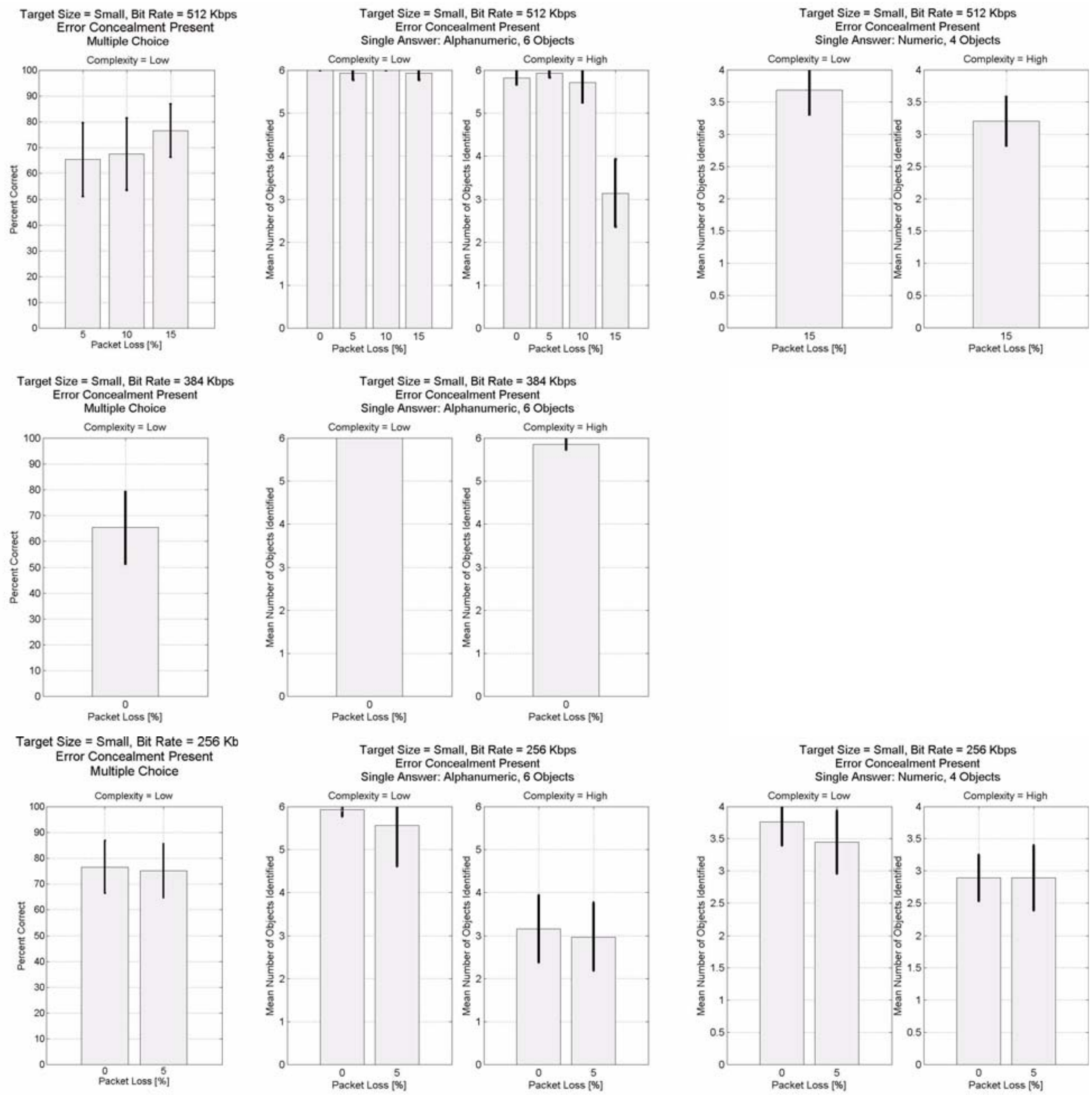


Figure 9 again shows the effect of complexity on the ability to perform recognition tasks for small targets, given the alphanumeric- and numeric-recognition tasks with high levels of impairment — in this case 15.0-percent packet loss at 512 kbps, and 256 kbps at both 0.0-percent packet loss and 5.0-percent packet loss. The multiple-choice data shows that viewers were able to complete the object-recognition task relatively well even at high packet loss rates of up to 15.0-percent.

Figure 9: Percent targets identified, target size = small, with error concealment (eight graphs)



3.3.2.2 Loss of Detection

Table 9, which shows loss of detection ability data (see Section 2.4) for the small-target-size-without-error-concealment cases, and confirms the trends seen in Figure 8 for alphanumeric-recognition tasks. The loss-of-detection scores are consistently higher for the high-complexity scenes for alphanumeric-recognition tasks in all network conditions. Table 9 does not show the expected decline in detection ability for the multiple-choice task. The numeric-detection task shows counter-intuitive results that correlate with Figure 8.

Table 9: Loss of detection ability, target size = large, without error concealment

		Multiple Choice [Percent]	Alphanumeric [Mean number of objects / 6]		Numeric [Mean Number of Objects / 4]	
	Complexity	**Low**	**Low**	**High**	**Low**	**High**
Bit Rate	**Percent Packet Loss**					
512 kbps	**0.0**	–0.82	1.58	2.06	N/A	N/A
	0.1	–1.62	1.95	2.13	N/A	N/A
	0.5	–3.05	1.76	3.13	2.50	0.04
384 kbps	**0.0**	1.63	1.98	2.02	N/A	N/A
	0.2	N/A	1.73	1.95	N/A	N/A
256 kbps	**0.0**	–0.92	2.06	3.49	2.59	0.27
	0.1	–0.36	2.09	3.29	2.32	0.27

Table 10 shows the loss-of-detection ability data for the small target size with error concealment cases, and confirms the trends seem in Figure 9 with the high-complexity scenes showing a greater loss of detection ability for all impairment levels for alphanumeric and numeric tasks.

Table 10: Loss of detection ability, target size = small, with error concealment

		Multiple Choice [Percent]	Alphanumeric [Mean number of objects / 6]		Numeric [Mean Number of Objects / 4]	
	Complexity	**Low**	**Low**	**High**	**Low**	**High**
Bit Rate	**Percent Packet Loss**					
512 kbps	**0.0**	N/A	–0.07	0.15	N/A	N/A
	5.0	8.16	0.00	0.37	N/A	N/A
	10.0	6.12	–0.07	0.26	N/A	N/A
	15.0	5.90	0.00	2.81	0.14	1.19
384 kbps	**0.0**	8.16	–0.07	.011	N/A	N/A
256 kbps	**0.0**	5.87	0.00	2.79	0.06	1.11
	5.0	7.31	0.37	2.98	0.38	1.11

3.4 Summary

Table 11 summarizes the results of PS4 drawn from the figures in Section 3.3. The summary considers both types of tasks: multiple-choice and alphanumeric.

Table 11: Tactical and live surveillance video performance requirements

Parameters	Performance Requirements	
Interaction of Minimum Bit Rate (BR) (Section 3.5.3.1 in [1]) and a Maximum Packet Loss Ratio (PLR) (Section 3.5.3.2 in [1]) without Error Concealment	**Small Target**	
	Low Complexity ■ PLR of 0.1 Percent with a BR of 256 kbps	***High Complexity*** ■ PLR of 0 (None) Percent with a BR of 512 kbps
	Large Target	
	Low Complexity ■ PLR of 0.1 Percent with a BR of 256 kbps ■ PLR of 0.5 Percent with a BR of 512 kbps	***High Complexity*** ■ PLR of 0.1 Percent with a BR of 256 kbps ■ PLR of 0.5 Percent with a BR of 512 kbps
Interaction of Minimum Bit Rate (BR) (Section 3.5.3.1 in [1]) and a Maximum Packet Loss Ratio (PLR) (Section 3.5.3.2 in [1]) with Error Concealment	**Small Target**	
	Low Complexity ■ PLR of 15 Percent with a BR of 512 kbps	***High Complexity*** ■ PLR of 10 Percent with a BR of 512 kbps
	Large Target	
	Low Complexity ■ PLR of 5 Percent with a BR of 256 kbps	***High Complexity*** ■ PLR of 5 Percent with a BR of 256 kbps

4 Recorded Surveillance Test (PS5)

PS5 focused on recorded surveillance video analysis. Professional video analysts performed specific recognition tasks using three different scenario groups. Viewers could pause and replay the video as many times as necessary.

4.1 PS5 Scenario Group Summary

Table 12 describes the three scenario groups created for PS5. Occasionally, variations were created by asking the viewers to answer multiple questions pertaining to a single clip.

Table 12: Scenario groups for PS5

Scenario Group: Number of Variations:	Size[b]	Complexity[a] Low	High	Discrimination Level: Experimental Method:	Scene Description:
Scenario Group: Bank Surveillance **Number of Variations:** 111	**Small:** **Large:**	32 28	20 31	**Discrimination Level:** Elements of the Action: 24 Target Class: 24 Target Characteristics: 72 Target Positive Recognition: 11 **Experimental Method:** Multiple Choice	Typical bank surveillance footage, depicting different types of robberies, from a single robber quietly passing a note to a teller to armed robberies undertaken by either an individual robber or a group of robbers.
Scenario Group: Store Surveillance **Number of Variations:** 87	**Small:** **Large:**	28 16	16 27	**Discrimination Level:** Elements of the Action: 8 Target Class: 65 Target Characteristics: 14 Target Positive Recognition: 0 **Experimental Method:** Multiple Choice	Store surveillance recordings of customers taking items off of a shelf and either placing them back on the shelf, in a pocket, inside clothing, or in a bag.
Scenario Group: Outdoor Surveillance **Number of Variations:** 51	**Small:** **Large:**	23 20	3 5	**Discrimination Level:** Elements of the Action: 18 Target Class: 10 Target Characteristics: 22 Target Positive Recognition: 1 **Experimental Method:** Multiple Choice	Outdoor surveillance camera filming pedestrian activity on a busy sidewalk.

a. Scene Complexity arguments are "Low" or "High"
b. Target Size arguments are "Small" or "Large" for which High/Low sums equal Number of Variations

4.2 Impairments to the Video

The encoder bit rates were varied from 256 kbps to 512 kbps, and the packet loss rate was varied from 0.0 to 0.5 percent, as Table 13 shows.

Table 13: Compression and packet loss combinations for PS5

System\ bit rate (kbps)	512	384	256
Percent packet losses used	0.5, 0.1, 0.0	0.2, 0.1, 0.0	0.1, 0.0

4.3 Results

Twenty-five viewers viewed all of the clips in the test design. Correct answers were tabulated and grouped according to the four parameters under study (target size, scene complexity, compression, and packet loss). The results in Figure 10 through Figure 15 show that under all impairment levels, analysts could answer at least 80 percent of the questions correctly, regardless of the target size, complexity score, compression, or packet loss rate. These results indicate that most analysts should correctly identify pertinent details in the footage at these levels of impairment. For Figure 10 through Figure 15, results include 95-percent confidence intervals (intervals calculated using the Wald method for binomial distributions).

Examining the percent of correct responses versus target size and complexity level for a given bit rate yielded an upper impairment level limit that could inform future tactical and surveillance video testing. Results indicate that although a scene may be considered “complex”, the detail sought by an analyst may not occur in the “complex” part of a scene. Hence, the complexity score for a clip may be arbitrary depending upon what information the analyst needs in the clip and whether the information is in a “complex” part of the clip.

Some trends emerge from the results, indicating that viewers could correctly identify more targets at higher impairment levels; however these trends could be explained by the viewers’ ability to memorize a scenario group setting and the context of the video clips. For example, in the shoplifting footage, a viewer may notice during testing that notepads and paper are kept on the bottom shelf. If the viewer is questioned later about an item taken from the bottom shelf, he/she could answer the identification question correctly, even if the impairments are so severe that identification of the target is almost impossible without having previously memorized the aisle’s layout.

Figure 10: Percent targets identified with small target size and 512 kbps bit rate

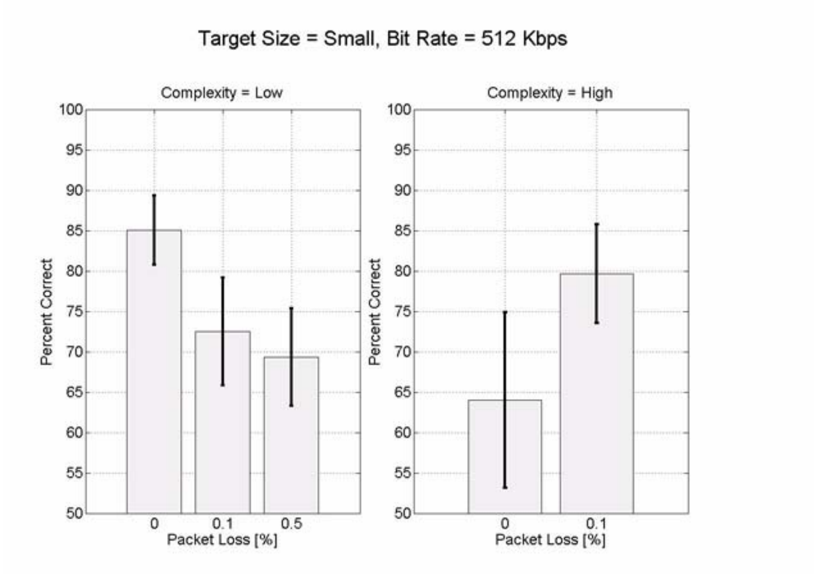


Figure 11: Percent targets identified with small target size and 384 kbps bit rate

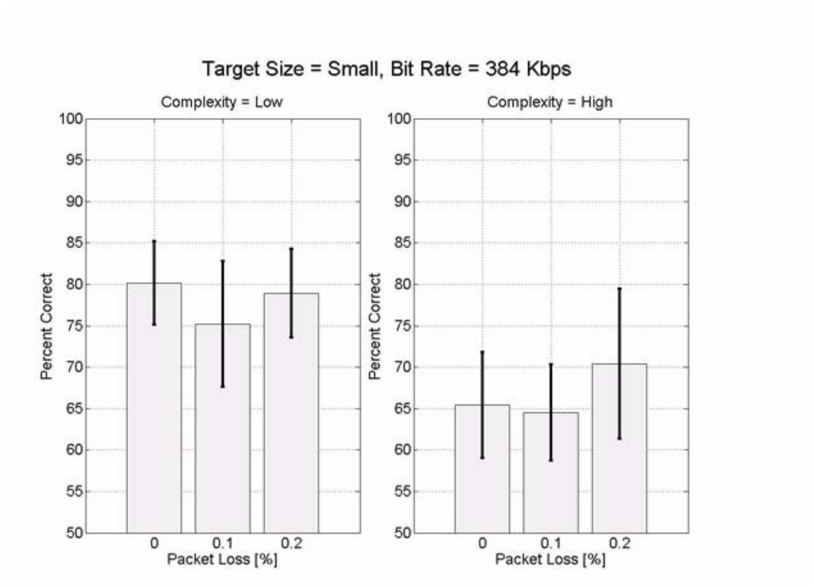


Figure 12: Percent targets identified with small target size and 256 kbps bit rate

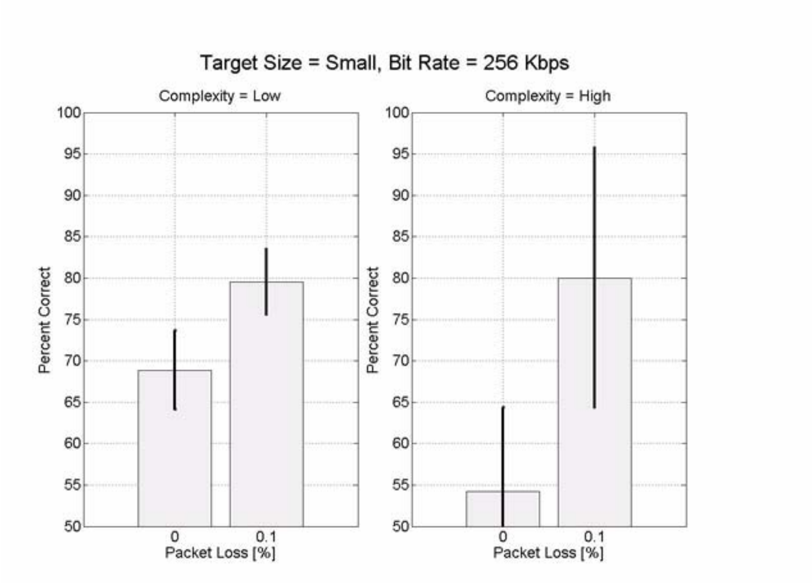


Figure 13: Percent targets identified with large target size and 512 kbps bit rate

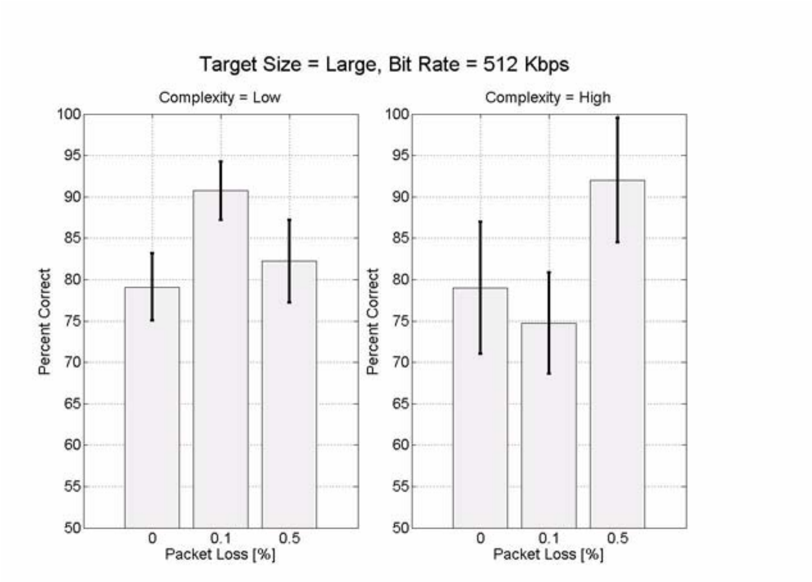


Figure 14: Percent targets identified with large target size and 384 kbps bit rate

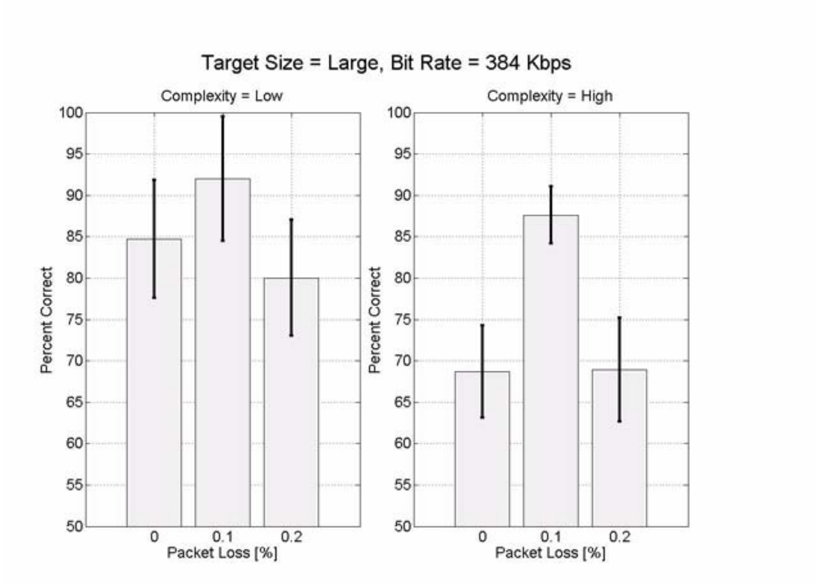
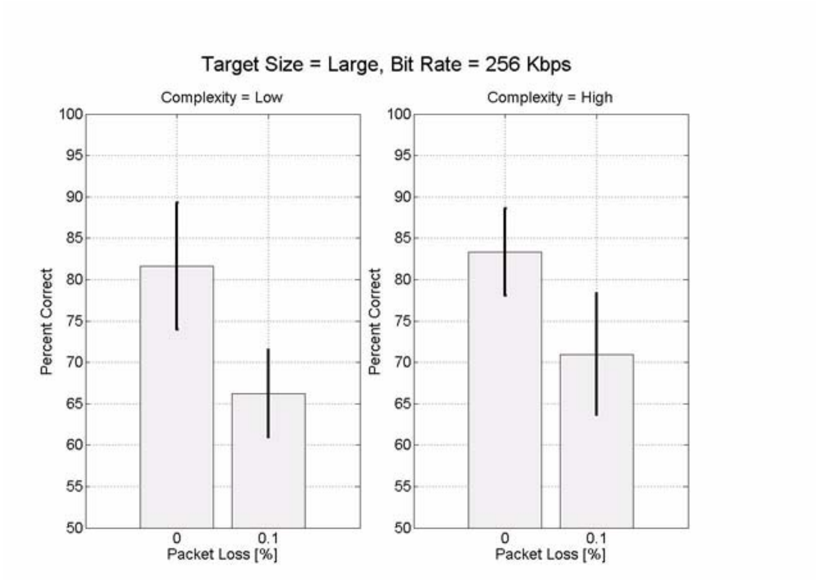


Figure 15: Percent targets identified with large target size and 256 kbps bit rate



4.4 Summary

Table 14 summarizes the results of PS5 drawn from the figures in Section 4.3.

Table 14: Recorded surveillance video performance requirements

Parameters	**Performance Requirements**	
Interaction of Minimum Bit Rate (BR) (Section 3.5.3.1 in [1]) and a Maximum Packet Loss Ratio (PLR) (Section 3.5.3.2 in [1]) without Error Concealment	**Small Target**	
	Low Complexity ■ PLR of 0.1 Percent with a BR of 256 kbps ■ PLR of 0.2 Percent with a BR of 384 kbps ■ PLR of 0.5 Percent with a BR of 512 kbps	***High Complexity*** ■ PLR of 0.1 Percent with a BR of 256 kbps ■ PLR of 0.2 Percent with a BR of 384 kbps ■ PLR of 0.5 Percent with a BR of 512 kbps
	Large Target	
	Low Complexity ■ PLR of 0.1 Percent with a BR of 256 kbps ■ PLR of 0.2 Percent with a BR of 384 kbps ■ PLR of 0.5 Percent with a BR of 512 kbps	***High Complexity*** ■ PLR of 0.1 Percent with a BR of 256 kbps ■ PLR of 0.2 Percent with a BR of 384 kbps ■ PLR of 0.5 Percent with a BR of 512 kbps

5 Limitations

The number of permutations needed to study effects of the four parameters (size, complexity, compression, and packet loss) was too large to include a full matrix of comparisons for PS4 and PS5. In addition, a vast number of variables influence the quality of surveillance video recordings. Thus, the scope of PS4 and PS5 is narrow. This test studied only two variables: encoder bit rate and packet loss rate. Hence, the test excluded such variables such as lighting conditions, camera lenses, camera placement, or video enhancement. Unlike PS4, PS5 did not study error concealment.

6 Future Work

The complex relationship between compression rates and packet loss warrants further study. In some cases in PS4, lower bit rates did not affect packet loss as severely as higher bit rates. The results of PS4 and PS5 suggest further research is necessary to study how lower bit rates affect larger targets and how higher bit rates affect complex scenes with small targets.

The study of impairment schemes is another area for further testing. A possible future test could study how impairment schemes differ depending on higher packet loss rates and decreased frame rates. Other areas for study include: filming clips in different weather conditions (e.g., overcast, rain, or snow), presenting clips in black-and-white, and presenting clips in different frame size formats (e.g., 320 x 240, 640 x 480, or 320 x 480).

7 Summary

The task-based subjective testing used in PS4 and PS5 yielded slightly more lenient requirements than previous studies. These studies used a rating system (Excellent to Bad) rather than requiring the viewer to perform a recognition task. The current results should be more accurate, since they measured a viewer's actual ability to use the video, rather than the viewer's perception of the usability of the video.

Video is increasingly used to recognize objects, people or events for public safety applications. Previous studies reported in [3] (i.e., PS1–PS3) relied on a viewer's subjective impression to measure video quality. To more closely match actual public safety video applications and to narrow requirements from previous recommendations, the task-based format of PS4 and PS5 relied on a viewer's ability to accurately perform a recognition task. PS4 and PS5 thereby rate video quality according to the usefulness of the video within its application.

8 References

[1] *Public Safety Statement of Requirements for Communications & Interoperability*, Volume II: Quantitative, Version 1.2, September 2008.

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

[3] SAFECOM, Public Safety Technical Report, "Tactical and surveillance video quality experiments."

[4] *Public Safety Statement of Requirements for Communications & Interoperability*, Volume II: Quantitative, Version 1.1, November 2007.

[5] ITU-R. 2002. Recommendation BT.500-11. Methodology for Subjective Assessment of the Quality of Television Pictures. Recommendations of the ITU, Radiocommunication Sector.

[6] ITU-T. 1999 September. Recommendation P.910. Subjective Video Quality Assessment Methods for Multimedia Applications. Recommendations of the ITU, Telecommunication Standardization Sector.

[7] J. Johnson, "Analysis of image-forming systems," Proc. Image Intensifier Symp., 249-273 (1958).

[8] D. Green, "Application of Detection Theory in Psychophysics," Proceedings of the IEEE, Vol. 58, No. 5, May 1970.

[9] ANSI S3.2, American National Standard Method for Measuring the Intelligibility of Speech Over Communications Systems, 1989.

[10] W. Mendenhall. *Statistics for Engineering and the Sciences*, Macmillan Publishing Company, 1992.

This page intentionally left blank.