Recorded-Video Quality Tests for Object Recognition Tasks

Homeland Security

Science and Technology

DHS-TR-PSC-11-01

U.S. Department of Homeland Security

Public Safety Communications

Technical Report

This page intentionally left blank.

Homeland Security
Science and Technology

Support to the Homeland Security Enterprise and First Responders:
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's Support to the Homeland Security Enterprise and First Responders Group, 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 allows for the valuable input and insights 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 unnecessary 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

Recorded-Video Quality Tests for Object Recognition Tasks

DHS-TR-PSC-11-01
September 2011

Reported for: The Office for Interoperability and Compatibility by the Public Safety Communications Research program

Homeland Security

This page intentionally left blank.

Publication Notice

Disclaimer

DHS' Science and Technology (S&T) 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 S&T, OIC currently assists 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: sandtfrg@hq.dhs.gov

This page intentionally left blank.

Contents

This page intentionally left blank.

Figures

This page intentionally left blank.

Tables

This page intentionally left blank

Abbreviations

CAVLC	Context-adaptive variable-length coding
CBR	Constant bit rate
CIF	Common Intermediate Format (352 lines x 288 pixels)
DoC	Department of Commerce
DHS	Department of Homeland Security
GOP	Group of Pictures
GUC	Generalized use class
HRC	Hypothetical Reference Circuit
HD	High Definition
ITS	Institute for Telecommunication Sciences
ITU-T	International Telecommunications Union, Telecommunication Standardization Sector
NTIA	National Telecommunications and Information Administration
OIC	Office for Interoperability and Compatibility
PSCR	Public Safety Communications Research program
PSVQ	Public Safety Video Quality
QP	Quantization Parameter
SRC	Source video
TRV	Target Recognition Video
VGA	Video Graphics Array (640 lines x 480 pixels)
VQEG	Video Quality Experts Group
VQiPS	Video Quality in Public Safety

This page intentionally left blank.

Abstract

This report describes a laboratory study to investigate how the interaction of various scene parameters and network conditions affect a viewer's ability to recognize a given target or object in a recorded video. The scene parameters under study were target size, scene motion, and scene lighting, while the network conditions were resolution reduction and H.264 compression. The task-based subjective tests this report describes follow the test methods described in ITU-T Recommendation P.912 [1]. Recognition rates are given as percentage of objects correctly identified. This study was undertaken to provide scientific measurements for recommendations that the Video Quality in Public Safety (VQiPS) working group will publish in a user guide. Combining the results of this study with the results in our previous report [2], we have addressed all of the Generalized Use Class (GUC) parameters except for discrimination level. Ultimately, this study will allow us to give important guidance to end users of video equipment in public safety applications.

Key words: object recognition, video quality, subjective test methods, target recognition video

1 Introduction

Video is used for a wide variety of public safety applications. These applications are task-based, and the tasks can frequently be generalized as identifying objects of interest, or targets, in the video. While some applications rely on computer vision to perform these tasks automatically, many others are intended for a human observer. These include surveillance, telemedicine, urban search and rescue, and remote command and control, among others. Video quality for these video systems can thus be evaluated in terms of the success rates of accomplishing such object-recognition tasks by test subjects in a controlled environment. The study described in this report follows this approach.

Generalizing the factors which will affect object-recognition success rates regardless of application makes the results applicable to a wide variety of potential public safety video system customers. The success rate of object-recognition tasks can be affected by many issues common to most applications. We can break these issues into two areas: 1) the parameters affecting the scene and 2) whether the scene video is live or recorded.

First, parameters of the scene itself may greatly affect the ability to recognize objects. For this report's study, scene parameters of particular interest are:

- Lighting
- Motion
- Size of the target of interest

Second, whether the scene video is live (viewed in real-time) or previously recorded (viewed later) may also affect the difficulty of the target recognition task. For this report, a recorded video scenario was studied. A study very similar to this one, but focused on the live video scenario, was performed in 2010 [2].

Defining use cases for video applications, and then categorizing use cases that share certain important aspects is the concept behind the generalized use classes (GUCs) that are included in the user guide developed by the Video Quality in Public Safety (VQiPS) working group [3]. This guide helps public safety video system end users identify their use cases and GUCs.

Figure 1 provides a graphical representation of a GUC. The set of all use cases that share the same checked attributes form a GUC.

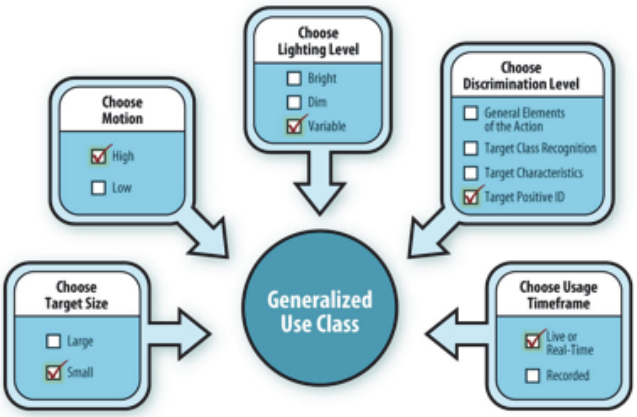


Figure 1: A representative example of determining a GUC.

The Recommendations Tool for Video Requirements [4] provides requirement considerations or suggestions for a single use case or for one or more GUCs in an effort to guide users through important video system considerations. Further, this study was conducted in such a way that its results could also be incorporated into the user guide [3] and the recommendations tool [4].

A myriad of parameters for the video system itself will contribute to target-recognition video (TRV) quality. OIC has supported the Public Safety Communications Research program in providing content for the user guide and a GUC framework regarding some of these parameters. Most importantly, OIC has also aided in the process of conducting laboratory experiments to explore the effects of network conditions upon video quality for public safety applications.

This study, therefore, is focused on network conditions. Specifically, the parameters under study are H.264 compression combined with resolution reduction to Video Graphics Array (VGA)—or 640x480 pixels—and Common Intermediate Format (CIF)—or 352x288 pixels. These types of conditions are representative of video that has been transmitted to mobile devices.

This study followed the test method illustrated in ITU-T P.912 [1]. As a result, subjective assessment methods for video that is to be used for TRV were addressed. A subjective test was also performed through the use of expert viewers. In addition, the multiple-choice method was used, in which subjects were asked to identify an object in the video given several choices. Test scenes included a number of combinations of the scene parameters named in the above paragraph. For each of the two resolutions, five bit rates were imposed on the H.264 encoder and tested. The results are reported as recognition rates, i.e., the percentage of objects correctly identified.

2 Targets and Scenario Groups

The test scenes used for this test also followed ITU-T P.912 [1], which introduces the concept of scenario groups. Scenario groups are collections of scenes with the same basic scenario, but with slight variations in each scene. For example, a scenario group could include a person walking by holding an object. Each scene within the scenario group would be nearly identical, with the exception of the object that is being carried. By using scenario groups, scene memorization or other visual clues should be minimized, and the test subject's ability to recognize the object itself can be more accurately ascertained.

Each scenario group used in this test included a collection of the same scene, with the variation being the object of interest, or target, in the scene. The targets were:

- Gun
- Electroshock weapon
- Hand-held land mobile radio
- Mug
- Soda
- Flashlight
- Cell phone

The objects were filmed while both lying on a pedestal and being carried at walking speed. The walking scenes included both a walking-left scenario and a walking-right scenario, with the object being carried in the carrier's left hand. This was done for the purpose of testing two views of the object as it was held in hand. The two levels of motion contained in the test scenes were stationary and walking speed.

Test scenes were filmed at two locations. The first set of scenes involved a sunny, cloudless, outdoor, mid-day, rural setting. The second set of scenes involved an underground law enforcement shooting range with various lighting options. Between the two locations, lighting levels filmed represented outdoor daylight, indoor bright light, indoor dim light, and indoor nearly dark conditions. For the indoor scenes, a constantly flashing law enforcement light bar was part of the lighting scheme as well.

Target size variations were created by filming each scene at two camera distances for the daylight scenario groups. Therefore, test scenes representing both a "small" target and a "large" target were created. The indoor test scenes employed one camera distance and were also designated as "large" targets.

For the sake of reducing the size of the test, not all possible combinations of target size, lighting level, and motion were used. A stationary scenario group for bright indoor lighting was omitted, as well as scenario groups with any type of indoor lighting in combination with a small target. Overall, fourteen scenario groups were created for use in the test. With two exceptions, each scenario group contained scenes with each of the seven test objects. The exceptions were the scenario groups with dark lighting and walking speed—which did not use the flashlight as a target. The total number of test scenes was thus 96: 14 scenario groups with seven objects each, minus the two that were omitted. The clip lengths ranged from four to nine seconds, with most having a length of five seconds. Table 1 provides a summary of all of the scenario groups. Further information can be found in Appendix A.

Table 1: Summary of scenario groups.

Scenario Group #	Lighting Condition	Motion	Target Size
1	daylight	stationary	large
2	daylight	walking speed, right	large
3	daylight	walking speed, left	large
4	daylight	stationary	small
5	daylight	walking speed, right	small

Table 1: Summary of scenario groups. (Continued)

Scenario Group #	Lighting Condition	Motion	Target Size
6	daylight	walking speed, left	small
7	bright/flash	walking speed, right	large
8	bright/flash	walking speed, left	large
9	dim/flash	stationary	large
10	dim/flash	walking speed, right	large
11	dim/flash	walking speed, left	large
12	dark/flash	stationary	large
13	dark/flash	walking speed, right	large
14	dark/flash	walking speed, left	large

3 Processed Scenes

All HD clips were down-converted to two display resolutions: VGA (640x480 pixels) and CIF (352x288 pixels). The frame rate was kept constant at 29.97 frames per second (fps).

The clips were then compressed via H.264 encoding at various bit rates. Five bit rates were chosen for each resolution. The bit rates (listed in Table 2) were chosen to represent a wide range of resultant video quality. Each combination of resolution and bit rate is what is referred to as a Hypothetical Reference Circuit, or HRC. (This term is used by the Video Quality Experts Group, or VQEG, a part of the ITU-T and refers to the distortion to the video signal that is being tested: in this particular case, combinations of compression and resolution reduction.) This test followed recommendations issued by the ITU-T.

Table 2: Encoder bit rates.

Resolution	Bit rates [kbps]
CIF	64, 128, 256, 512, 1024
VGA	128, 256, 512, 1024, 1536

4 Test Design

4.1 Test Size

As described in the previous sections, there were 96 source scenes and ten HRCs that underwent testing. The total number of clips—when all source scenes were processed with all HRCs—was 960. In order to reduce test length and viewer fatigue, each viewer did not see each clip, but instead saw three clips for each scenario group/HRC combination. Therefore, with 14 scenario groups and 10 HRCs present in this test, each viewer watched 420 clips. Additionally, at the beginning of the test, each viewer saw a training sequence showing each target with a label, then took a practice test consisting of four additional clips for

familiarization with the test software (scores were not kept). Appendix A provides frames from the training sequence.

4.2 Viewers

38 viewers participated in the test. In accordance to ITU-T P.912 [1], expert viewers were recruited. Each viewer had practitioner experience in law enforcement, fire service, or emergency medical services. Years of experience ranged from 3 to 40.

Viewers were screened for visual acuity and color vision, by way of Snellen and Ishihara tests, respectively. Nearly all viewers had normal vision. One viewer was red-green colorblind, and three others had minor acuity deficiencies. Viewers were not automatically excluded from the test if they demonstrated impaired acuity or color vision; however, additional data analysis was conducted to determine whether their test scores were consistent with viewers who had not demonstrated such impairments. This analysis revealed no significant difference between impaired viewers and those with normal vision

4.3 Test Environment and Software

Viewing conditions generally followed the recommendations in ITU-T P.910 [5]. One exception was that the viewers could choose their viewing distance, and it was not recorded. However, it is reasonable to assume that viewers' chosen viewing distances most likely fell into the 1-8 picture heights recommendation, given an approximate picture height of 5 inches

Figure 2: Test software user interface.

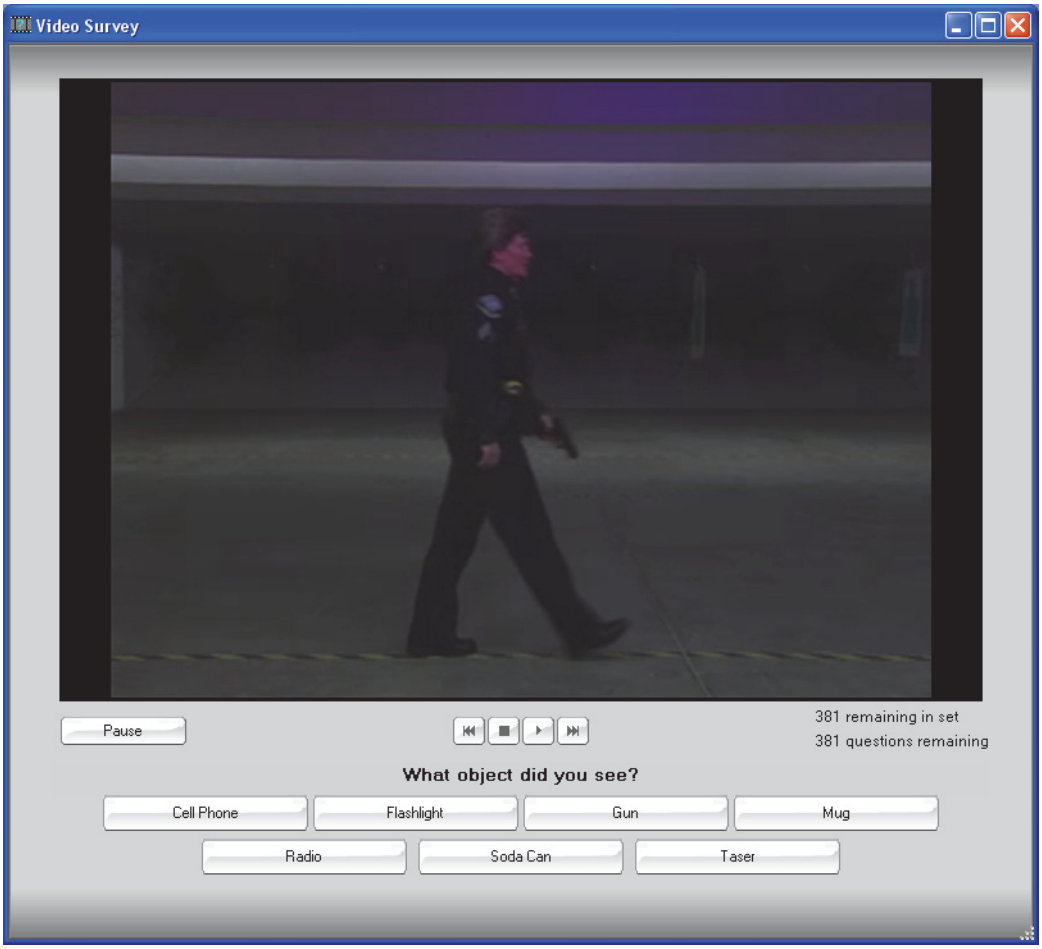


Figure 2 illustrates the user interface for the test. Viewers were shown processed video clips and were asked to select which target was present in the clip from among seven choices. The choices given were the same as the list of target objects used in the scenario groups. Viewers were allowed to view the clip as many times as they chose, and had the option to rewind and fast-forward, stepping through and freezing frame as much as desired. The user interface recorded all such interactions with the software for future analysis, although this information is not part of this report. The test software also recorded the specific frame the viewer was looking at when answering the question.

The test was expected to involve approximately 60 to 90 minutes of time spent actively viewing videos and providing answers to the object-recognition questions. Since the viewers controlled their own interaction with the video, the total time for the test could vary widely among viewers. Subjects were free to take breaks as needed.

5 Results

Recognition rates (i.e., percentages of correct answers) were calculated for scenario group/HRC combinations. Each viewer was considered a replicate and each answer a unique data point. Guessing was expected, since the options of "I don't know," or "unsure" were unavailable to test subjects. The recognition rates were adjusted for guessing. 95-percent confidence intervals were calculated.

For the scenario groups involving a carried object, the data from the walking-left and walking-right scenario groups were combined to provide an overall estimate of the effect of motion on object recognition. Therefore, recognition rates for moving objects are based on twice as many samples as results for stationary objects. Further information regarding data analysis can be found in Appendix C.

5.1 Best-Case Recognition Rates

Figure 3: Recognition rates for large stationary targets in daylight.

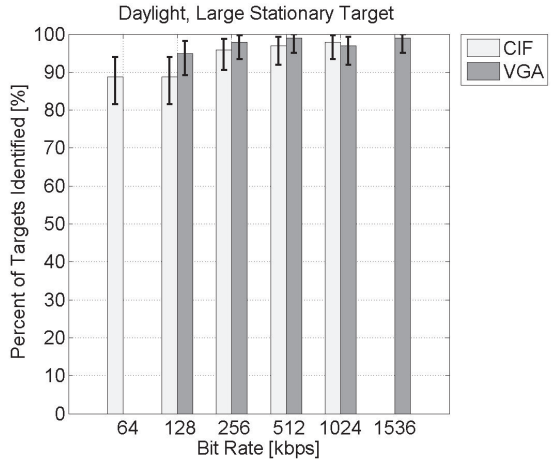


Figure 4: Recognition rates for large targets carried at walking speed in bright light.

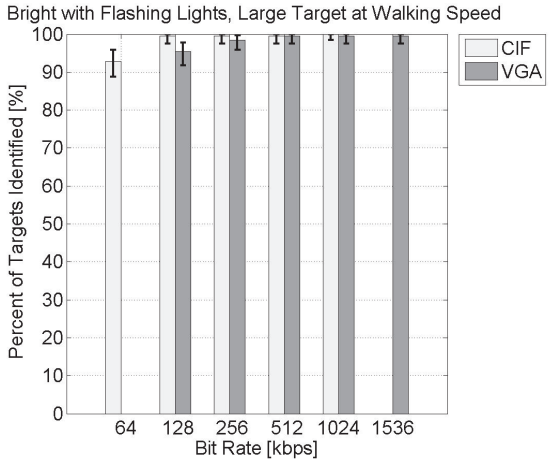
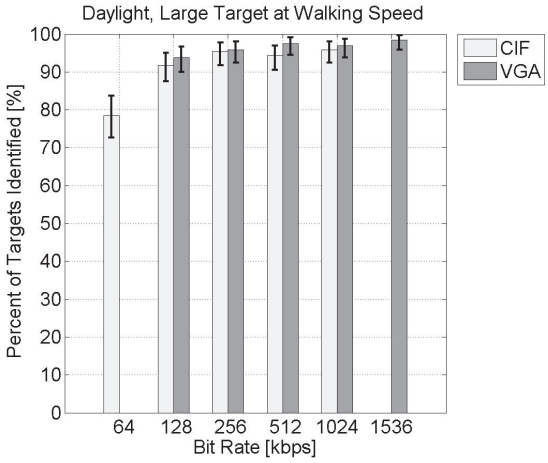


Figure 5: Recognition rates for large targets carried at walking speed in daylight.



5.2 Recognition Rates and Lighting

As lighting conditions were degraded, recognition levels no longer reached 90-percent for any HRCs. Figure 6 shows that for large stationary targets, recognition rates were reduced to slightly below 90-percent as lighting became dim, and Figure 7 shows that the dark lighting condition drastically reduced recognition rates further. For three of the HRCs, even 50-percent recognition was not achieved. It should also be noted that a saturation effect similar to the one described in [2] is present in these data. For example, Figure 6

shows that recognition rates near 90-percent are achieved at 256 kbps, and there is no significant improvement in recognition performance with higher bit rates. This implies that poor lighting in a scene impairs the usefulness of that video in a way that cannot be compensated for with any amount of bandwidth.

Figure 6: Recognition rates for large stationary targets in dim lighting.

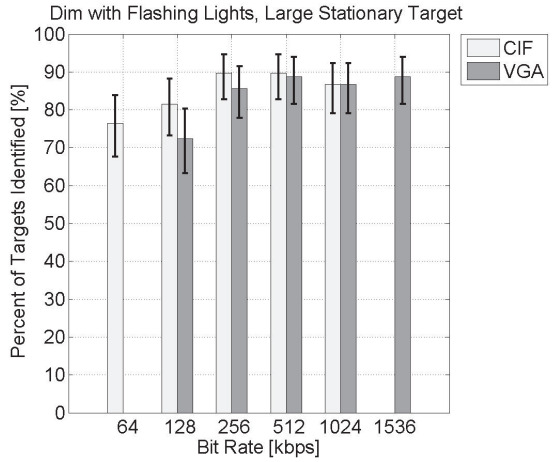
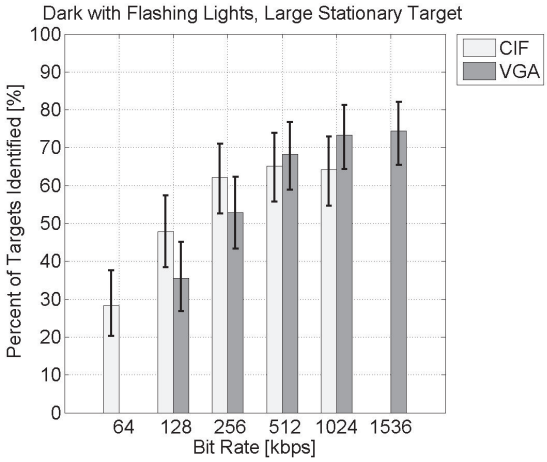


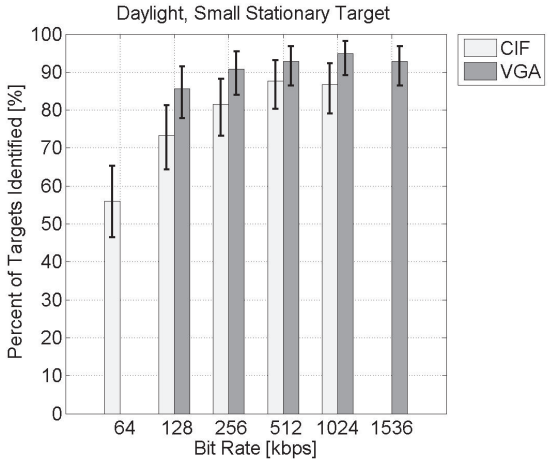
Figure 7: Recognition rates for large stationary targets in dark lighting.



5.3 Recognition Rates and Target Size

As the target size was reduced from large to small, a significant decrease in recognition rates resulted. Figure 8, when compared to Figure 3, shows that recognition rates no longer exceed 90-percent for most of the HRCs. The impact is most significant for the two lowest bit rates in combination with the CIF resolution. Figure 8 reveals a saturation effect with a small target similar to the one observed for poor lighting. As expected, the higher resolution VGA HRCs show better performance than CIF for a small target, but surprisingly, none of the differences are statistically significant.

Figure 8: Recognition rates for small stationary targets in daylight.



5.4 Recognition Rates and Motion

As depicted in Figure 5, walking speed motion with large targets in daylight still achieved good recognition rates. With the small target, as motion was added to the scene, the recognition rates dropped so that none of the HRCs, including the highest bit rates with VGA resolution, yielded recognition rates above 90-percent. Figure 9, in contrast with Figure 8, illustrates this effect.

Poorly lit large targets suffered an even greater degradation in recognition rates as motion was introduced. Figure 10 and Figure 11 show this degradation. For the dark condition with motion, most HRCs did not achieve even 50-percent recognition.

Figure 9: Recognition rates for small targets carried at walking speed in daylight.

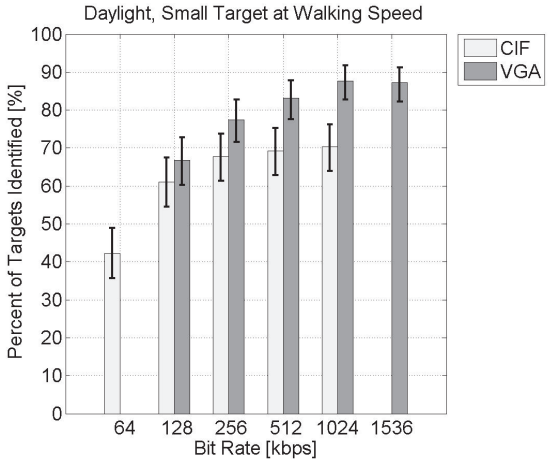


Figure 10: Recognition rates for large targets carried at walking speed in dim lighting.

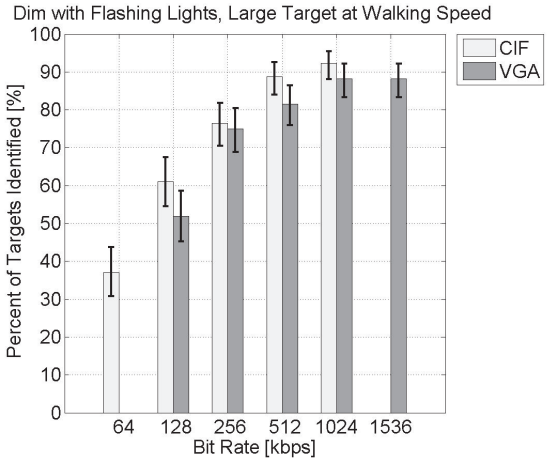
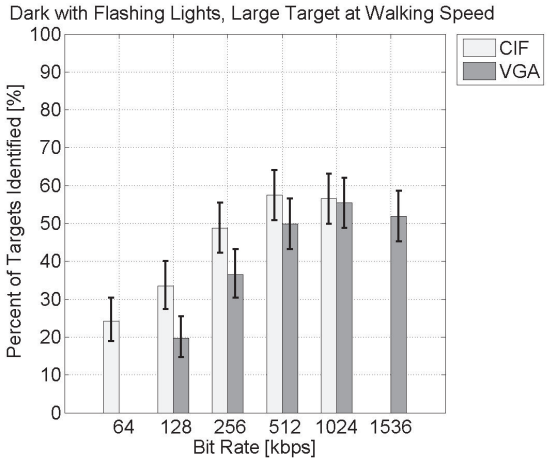


Figure 11: Recognition rates for large targets carried at walking speed in dark lighting.



5.5 Comparison Between Live and Recorded Results

The test described in this report is nearly identical in design to the test described in [2]. The test scenes, HRCs, test method, and viewing conditions were the same in both tests, however the important distinction of simulating a recorded video application by giving the viewer the ability to control playback with pause, replay, rewind, and fast-forward capabilities makes this test different from the live/real-time situation simulated in [2].

It is of particular interest, then, whether the recorded video capabilities improved upon or otherwise deviated from observations made about recognition rates that were achieved in the previous test. As a

summary, the following observations were made after analyzing the live/real-time recognition rate data (for complete results, see Appendix F):

- Given the large targets, viewers achieved 90- to 100-percent recognition with daylight lighting, as well as bright indoor lighting for almost all HRCs.
- Given large stationary targets and dim indoor lighting, recognition rates did not substantially exceed 90-percent for any HRC; with dark indoor lighting, that number dropped to 80-percent.
- The lower CIF resolution appeared to have a possible slight advantage in poor lighting-level scenarios; this was an unexpected finding.
- Recognition rates substantially higher than 90-percent could not be achieved given the small target with any HRC if motion was present.
- Motion worsened the degradation of recognition rates caused by poor lighting or small target size; for the dark indoor lighting condition, recognition rates fell from a best case of 80-percent to a best case of 60-percent.

The data in this test suggest that the recorded video situation does not differ significantly from the live video situation. In fact, when the recognition rates for specific scenario groups and HRCs are compared, there are very few differences that are statistically significant. Most of these differences occur in lowest quality HRCs. For example, the 64 kbps CIF and 128 kbps VGA clips for a large moving target in bright lighting showed better performance in the recorded setting as compared to live. Similarly, 64 kbps CIF for a large stationary target in dim lighting recognition rates are higher in the recorded setting. For this reason, it is hypothesized that the viewers' ability to interact with the video clips in the recorded setting allowed them to compensate for some of the quality lost due to compression of the video. In other words, compression introduces impairments to the recognition task that can be partially overcome by viewers in a recorded setting. The exact strategies used by viewers to improve their recognition have not yet been determined, but data were gathered in this test that may allow such an analysis in the near future.

Video processed at128 kbps CIF for a large stationary target in daylight actually showed significantly worse performance in the recorded case than in the live case. Also, recognition rates are significantly higher in a live video setting for a large stationary target in a dark room with flashing lights processed at 256 kbps and 512 kbps for CIF resolution. It is not clear why this is the case. Since the viewers have the option of watching the video once without pausing, they should be able to match the performance of the live test. This suggests that, given the freedom to interact with the video, the viewers may be choosing to behave in a way that ultimately hurts their performance. Further study is necessary to determine exactly what this behavior might be. On the other hand, it was noted in [2] that it was unclear why CIF so significantly outperformed VGA in the dark lighting case. These differences may be explained by some unknown effect in the previous test rather than self-defeating behavior on the part of the viewers.

For a small moving target in daylight, significantly improved performance for video processed at 1024 kbps VGA was seen. It is believed that this reflects an aberration in the data from the live test. The live data show 512 kbps and 1536 kbps each have better performance than 1024 kbps.

6 Recommendations

Although the test was limited in scope and further research should be conducted into the scene content and network parameters under study, tentative recommendations can be suggested based on the results of the test. These recommended minimum bit rates are shown in Table 3.

Table 3: Recommended minimum bit rates for H.264 encoding with recorded video analysis.

Scenario	Bit rate for 90-percent recognition [kpbs]		Bit rate for 50-percent recognition [kbps]	
	VGA	CIF	VGA	CIF
Daylight, stationary, large target	128*	256	128*	64*
Daylight, stationary, small target	256	NA	128*	64
Daylight, moving, large target	128*	128	128*	64
Daylight, moving, small target	N/A	N/A	128*	128
Bright and flashing lighting, moving large target	128*	64	128*	64*
Dim and flashing lighting, stationary, large target	1024	256	128*	64*
Dim and flashing lighting, moving, large target	N/A	1024	128	128
Dark with flashing lights, stationary, large target	N/A	N/A	256	256
Dark with flashing lights, moving, large target	N/A	N/A	1024	512

*A lower bit rate may also meet this criteria; the minimum bit rate tested sufficiently exceeded criteria.

7 Limitation, Conclusions, and Future Work

There was no scenario group for bright indoor lighting with stationary objects in this test. Unfortunately this makes it difficult to make a comparison between indoor and outdoor lighting because it is preferable to isolate the effect of lighting by only considering stationary objects. However, the general performance of outdoor and bright indoor lighting is very good, so both of these lighting conditions could be considered very useful for recognition tasks in general.

In conclusion, the results of this test have allowed for bit rate recommendations related to recorded video to be made. The most significant finding in this test is that recognition rates are almost all statistically equivalent to recognition rates in the live video test. It is hypothesized that the particular recognition task under study is being performed at a cognitively low level—perhaps even at a subconscious level. Conscious efforts to improve performance at this task, armed with tools like the ability to pause and replay the video, mostly failed. The exception to this assertion seems to be found in the lowest quality HRCs. This implies that whatever strategies viewers are using for the recognition task in a recorded setting are somewhat successful at overcoming compression impairments, but are less successful or not successful at

overcoming impairments introduced by poor lighting, small targets, or high motion. If a viewer cannot recognize an object due to poor scene conditions, the ability to replay or pause the video offers no statistical advantage.

In the near future, the Public Safety Video Quality (PSVQ) project will focus on using data from this test and from the previous test to make recommendations relative to a variety of public safety video tasks. PSVQ will expand beyond the object-recognition task using measures of acuity. Acuity is intended to be a single number applied to a video system that describes its usefulness for public safety tasks. Additionally, PSVQ will conduct one test to measure a relationship between object-recognition rates and acuity. Then another test will be conducted to measure the acuity levels required for a variety of public safety tasks. In this way, the data gathered here can be applied to a wide variety of situations.

8 References

[1] ITU-T Recommendation P.912. "Subjective video quality assessment methods for recognition tasks. Recommendations of the ITU, Telecommunications Standardization Sector. Geneva, 2008 (available at http://www.itu.int).

[2] "Video Quality Tests for Object Recognition Applications," http://www.safecomprogram.gov/library/Lists/Library/Attachments/231/Video_Quality_Tests_for_Object_Recognition_Applications.pdf, September 2010. Cited September 2011.

[3] *Defining Video Quality Requirements: A Guide for Public Safety,* Volume 1.0, July 2010. http://www.safecomprogram.gov/SiteCollectionDocuments/3aVideoUserRequirementGuidedoc.pdf. Cited September 2011.

[4] "Recommendations Tool for Video Requirements," http://www.pscr.gov/outreach/vqips/vqips_guide/rec_tool_vid_reqs.php. Cited September 2011

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

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

[7] N. Johnson, S. Kotz, and A. Kemp. *Univariate Discrete Distributions*, p. 129, Wiley, New York, second edition, 1992.

Appendix A Source Scenes

The test scenes used for this test also followed ITU-T P.912 [1], which introduces the concept of scenario groups. Scenario groups are collections of scenes with the same basic scenario, but with slight variations in each scene. For example, a scenario group could include a person walking by holding an object. Each scene within the scenario group would be nearly identical, with the exception of the object that is being carried. By using scenario groups, scene memorization or other visual clues should be minimized, and the test subject's ability to recognize the object itself can be more accurately ascertained.

Each scenario group used in this test included a collection of the same scene, with the variation being the object of interest, or target, in the scene. The targets were:

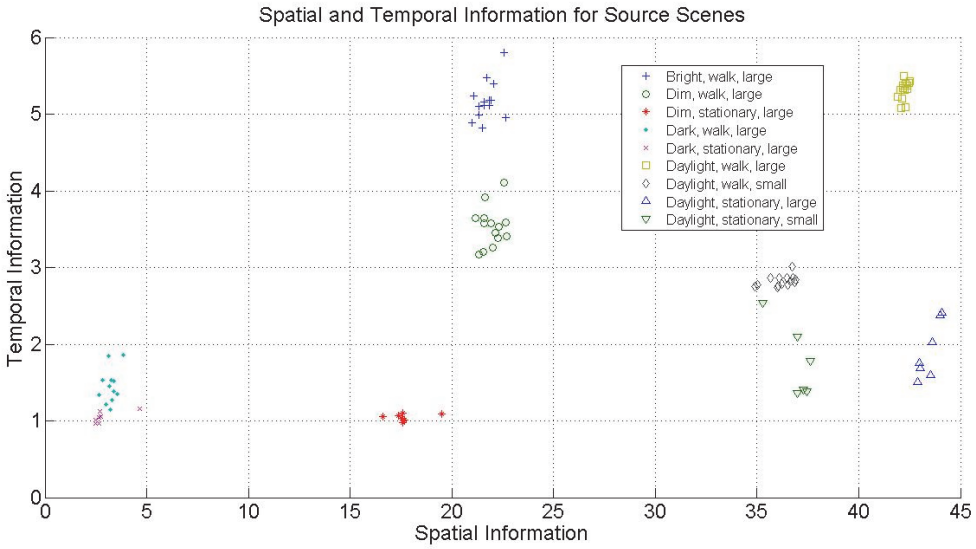
- Gun
- Electroshock weapon
- Hand-held land mobile radio
- Mug
- Soda
- Flashlight
- Cell phone

The objects were filmed while both lying on a pedestal and being carried at walking speed. For the walking speed scenes, the objects were carried in a way that was appropriate to that object, but not in a manner that might overtly indicate the use of the object. For example, the gun was not held out as if to shoot, and the cellular telephone was not held as if making a telephone call. Also, the walking scenes included both a walking-left scenario and a walking-right scenario, with the object being carried in the carrier's left hand. This was done for the purpose of testing two views of the object as it is held in hand.

Test scenes were filmed at two locations. The first set of scenes involved a sunny, cloudless, outdoor, mid-day, rural setting. The second set of scenes involved an underground law enforcement shooting range with various lighting options. Between the two locations, lighting levels filmed represented outdoor daylight, indoor bright light, indoor dim light, and indoor nearly dark conditions. For the indoor scenes, a constantly flashing law enforcement light bar was part of the lighting scheme as well. Light level measurements were made for the indoor dim and dark conditions using a handheld photometer manufactured by Quantum Instruments. The photometer was pointed in the same general direction as the camera. The dim with flashing lights condition registered 3.1 lux, while the dark with flashing lights condition registered 2.2 lux.

The two levels of motion contained in the test scenes were stationary and walking speed. ITU-T Recommendation P.910 [5], the Recommendation referred to in [2] for test scene selection, suggests that the full range of spatial and temporal perceptual information of interest be contained amongst the test scenes. In this study, only two levels motion were examined and two locales used to create all the scenario groups. The spatial and temporal perceptual information of interest is, therefore, limited; however, a plot of this information is included in Figure 12. The spatial and temporal information was calculated in accordance to ITU-T P.910 [5].

Figure 12: Spatial and temporal perceptual information.



Target size variations were created by filming each scene at two camera distances for the daylight scenario groups. Therefore, test scenes representing both a "small" target and a "large" target were created. The indoor test scenes employed one camera distance and were also designated as "large" targets. Information about the field of view for the various target sizes will be provided further in this section, as well as target sizes measured in pixels in the stationary scenario groups (after down-converting to the 640x480 display resolution).

For the sake of reducing the size of the test, not all possible combinations of target size, lighting level, and motion were used. A stationary scenario group for bright indoor lighting was omitted, as well as scenario groups with any type of indoor lighting in combination with a small target. Overall, fourteen scenario groups were created for use in the test. With two exceptions, each scenario group contained scenes with each of the seven test objects. The exceptions were the scenario groups with dark lighting and walking speed, which did not use the flashlight as a target. The total number of test scenes was thus 96.

The clips lasted for five seconds, with the exception of the daylight/walking-speed scenario groups and the indoor dim/stationary group. The indoor dim/stationary scenario group scenes lasted for four seconds. Among the daylight/walking-speed scenes, the scenario groups with a large target were six seconds long, while the scenario groups with the small target were nine seconds long. The difference in length is due to the increased time needed to walk across the increased field of view.

Test scenes were shot in 1080p HD with a Panasonic AJ-HPX3700, a broadcast-quality camera. The recording formats supported by this camera are AVC-Intra 100/50 and DVCPRO HD.

Table 4 provides a summary of all of the scenario groups.

Table 4: Summary of scenario groups.

Scenario group #	Lighting Condition	Motion	Target Size
1	daylight	stationary	large
2	daylight	walking speed, right	large

Table 4: Summary of scenario groups. (Continued)

Scenario group #	Lighting Condition	Motion	Target Size
3	daylight	walking speed, left	large
4	daylight	stationary	small
5	daylight	walking speed, right	small
6	daylight	walking speed, left	small
7	bright/flash	walking speed, right	large
8	bright/flash	walking speed, left	large
9	dim/flash (lighting: 3.1 lux)	stationary	large
10	dim/flash (lighting: 3.1 lux)	walking speed, right	large
11	dim/flash (lighting: 3.1 lux)	walking speed, left	large
12	dark/flash (lighting: 2.2 lux)	stationary	large
13	dark/flash (lighting: 2.2 lux)	walking speed, right	large
14	dark/flash (lighting: 2.2 lux)	walking speed, left	large

The horizontal field-of-view as seen by the camera at the distance of the target was measured for each scenario group. Table 5 shows these field-of-view measurements, as well as the distance from the camera to the target.

Table 5: Field-of-view and camera distance measurements.

Scenario group #	Field-of-View	Distance from camera
1	23’ 6”	35’ 9”
2-3	32’ 7”	35’ 9”
4	48’ 11”	48’
5-6	58’ 8”	48’
7-14	12’ 8”	17’ 2”

The sizes in pixels of the objects were measured in still frames that were taken from clips already down-converted to the 640x480 display resolution; these sizes are shown in Table 6 along with the color of the objects.

Table 6: Target sizes in pixels.

Object	Color	Size [pixels]: Scenario Group 9	Size [pixels]: Scenario Group 1	Size [pixels]: Scenario Group 4
Cell	Red	282	100	23
Flashlight	Black	466	105	44
Soda	White and Red	355	111	36
Mug	White	335	145	42
Taser	Yellow	439	158	38
Gun	Black	568	204	78
Radio	Black	677	283	77

The targets are shown in Figure 13 as they appear in still frames taken from the test software training sequence.

Figure 13: Test targets, as seen in training sequence.

A representative sample of still frames from the fourteen scenario groups is shown in the next several figures.

Figure 14: Frame from scenario group 1: daylight, stationary, large target. Target is flashlight.

Figure 15: Frame from scenario group 7: daylight, walking to the left, small target. Target is electroshock weapon.

Figure 16: Frame from scenario group 8: bright indoor light with flash, walking left, large target. Target is soda can.

Figure 17: Frame from scenario group 10: dim indoor light with flash, walking right, large target. Target is radio.

Figure 18: Frame from scenario group 12: indoor, dark lighting with flash, stationary, large target. Target is a mug. Note that this frame was taken during the flash from the law enforcement light bar.

Appendix B Processed Scenes

All HD clips were down-converted to two display resolutions: VGA (640x480 pixels) and CIF (352x288 pixels). The CIF resolution clips were enlarged to cover the same area on the screen as the VGA clips. The interpolation was done with a Lanczos filter. The frame rate was kept constant at 29.97 fps.

The clips were then compressed via H.264 encoding at various bit rates. Five bit rates were chosen for each resolution. The bit rates were chosen to represent a wide range of resultant video quality. Table 7 lists bit rates.

Table 7: Encoder bit rates.

Resolution	Bit rates [kbps]
CIF	64, 128, 256, 512, 1024
VGA	128, 256, 512, 1024, 1536

Encoding was done with TMPGEnc Xpress 4.0 software, which employed the MainConcept H.264 encoder.

Table 8: Software settings for H.264 encoding.

Parameter	Setting
Profile	Baseline
Level	Automatic
Frame Rate	29.97 fps
Bit rate mode	One-pass CBR
Motion Search Range	63
Detection of Scene Changes	Yes
GOP Length	33
B-Frame Count	0
Quantization Parameters	I Picture: 24 P Picture: 25
Entropy Coding Mode	CAVLC
Motion Estimation Sub-pixel Mode	Quarter-pixel

Appendix C Notes on Experimental Design

C.1 Randomization

The total number of clips when all source scenes were processed with all HRCs was 960. In order to reduce test length and viewer fatigue, each viewer did not see each clip, but instead saw three clips for each scenario group/HRC combination.

The clips to be viewed were selected in advance and distributed uniformly among the scenario groups and viewers. Each viewer saw a different pre-selected order of clips. The use of scenario groups counteracts memorization of clips—therefore, the only noticeable change between clips within a scenario group would be the object itself. As a result, each HRC can be used for each clip while still testing only the subject's ability to recognize the object.

C.2 Data Analysis

Recognition rates, (i.e., percentages of correct answers) were calculated for all scenario group/HRC combinations. Guessing was expected, since the options of "I don't know," or "unsure" were unavailable to test subjects. The recognition rates were adjusted to account for guessing using the following formula:

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

Where R_A represents the adjusted number of right answers, R represents the number of right answers, W represents the number of wrong answers, and n represents the number of answer choices [6]. 95-percent confidence intervals were calculated using the Clopper-Pearson method [7].

For the scenario groups involving a carried object, the data from the walking-left and walking-right scenario groups were combined to provide an overall estimate of the effect of motion on object recognition. Therefore, data representing the results with motion present are based on twice as many individual answers as results for stationary objects.

Appendix D Viewer Instructions

Viewers received the following text as instruction.

Public Safety VIDEO TEST
Overview for the Subjects

Thanks for coming in today to participate in our study. This study concerns the quality of video images for use in Public Safety applications. As a likely user of next-generation devices for Public Safety applications, we are interested in whether the videos to be presented are of sufficient quality to be used by you to perform several different potential tasks.

The study examines video used in a ***recorded****, real-time situation, and the ability to use this video to make real-time decisions on how to respond to an incident. This study does not apply to video which has been recorded for later examination. The application currently being focused on is object recognition. You will be asked to answer specific questions regarding content in the video. The scenes you will be shown, and the response requested, are from the following categories:*

Scene Description	*Response*
Person walking by, holding an object *Lighting scenario* ■ *Indoor flashing lights* ■ *Indoor, dark, flashing lights* ■ *Outdoor, daytime*	*Multiple choice: Identify the object from a list*
Stationary objects *Lighting scenario* ■ *Indoor flashing lights* ■ *Indoor, dark, flashing lights* ■ *Outdoor, daytime*	*Multiple choice: Identify the object from a list*

Each scene will be approximately 7 seconds long. While the clip is playing, you may pause or step backward or forward frame by frame. You may replay each clip as many times as you wish. You will then be asked to answer the question relating to the scene as described in the table above. The test software will record your answers, as well as when you paused, replayed, or stepped through frames of the clip and the total time you spent on each clip. ****Please wait for the video clip to finish playing before answering the question, and please do not close the media player window at any time during the test.***

Multiple Choice Instructions

Please choose the answer that most matches what you saw in the video. For this study there is no "other" or "I don't know" option. Therefore, please select the answer you believe to be most likely.

You will be asked to participate in one viewing session which is approximately 90 minutes long. A practice session will be presented to help you get familiar with the scene material and rating process, as well as a clip showing the objects you might see in the videos. You may take a break at any time during the session.

Appendix E Data Tables

Note: In Table 9, targets detected and percent correct have been adjusted to account for guesses made by participants in the study.

Table 9: Recorded video data.

	Bit rate [kbps]	64	128		256		512		1024		1536
	Resolution	CIF	CIF	VGA	CIF	VGA	CIF	VGA	CIF	VGA	VGA
Daylight Stationary Large	Targets Present	114	114	114	114	114	114	114	114	114	114
	Targets Identified	101.16	101.16	108.16	109.33	111.66	110.50	112.83	111.66	110.50	112.83
	Percent Correct	88.74	88.74	94.88	95.90	97.95	96.92	98.97	97.95	96.92	98.97
Bright Walking Large	Targets Present	228	228	228	228	228	228	228	228	228	228
	Targets Identified	211.66	226.83	217.50	226.83	224.50	226.83	226.83	228	226.83	226.83
	Percent Correct	92.83	99.48	95.39	99.48	98.46	99.48	99.48	100	99.48	99.48
Daylight Walking Large	Targets Present	228	228	228	228	228	228	228	228	228	228
	Targets Identified	179	209.33	214	217.50	218.66	215.16	222.16	218.66	221	224.50
	Percent Correct	78.50	91.81	93.85	95.39	95.90	94.37	97.44	95.90	96.92	98.46
Dim Stationary Large	Targets Present	114	114	114	114	114	114	114	114	114	114
	Targets Identified	87.16	93	82.50	102.33	97.66	102.33	101.16	98.83	98.83	101.16
	Percent Correct	76.46	81.57	72.36	89.76	85.67	89.76	88.74	86.69	86.69	88.74
Dark Stationary Large	Targets Present	114	114	114	114	114	114	114	114	114	114
	Targets Identified	32.33	54.50	40.50	70.83	60.33	74.33	77.83	73.16	83.66	84.83
	Percent Correct	28.36	47.80	35.52	62.13	52.92	65.20	68.27	64.18	73.39	74.41

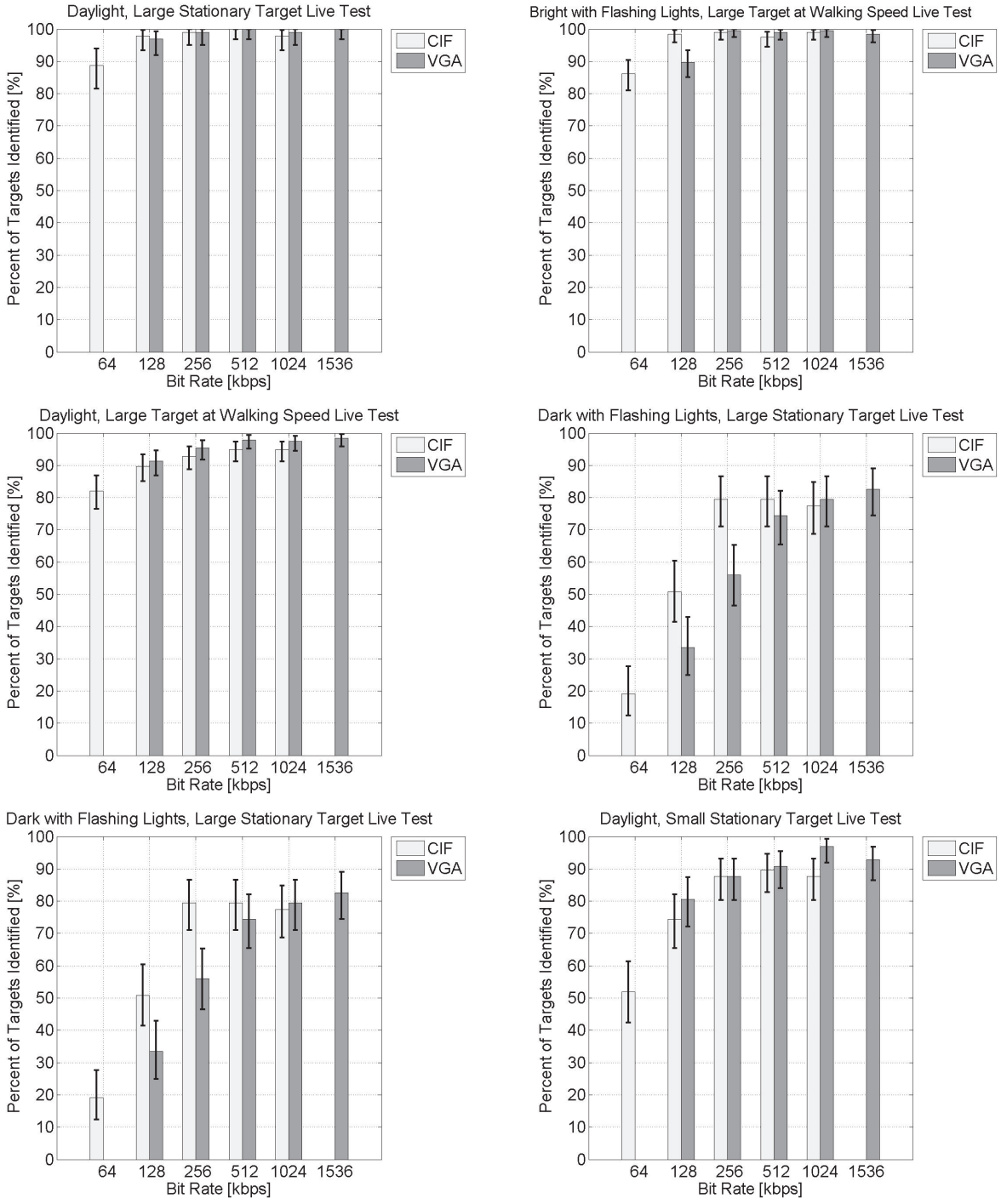
Table 9: Recorded video data. (Continued)

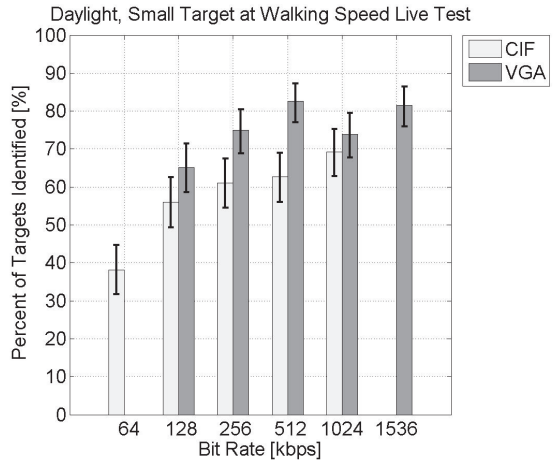
	Bit rate [kbps]	64	128		256		512		1024		1536
	Resolution	CIF	CIF	VGA	CIF	VGA	CIF	VGA	CIF	VGA	VGA
Daylight Stationary Small	Targets Present	114	114	114	114	114	114	114	114	114	114
	Targets Identified	63.83	83.66	97.66	93	103.50	100	105.83	98.83	108.16	105.83
	Percent Correct	55.99	73.39	85.67	81.57	90.78	87.71	92.83	86.69	94.88	92.83
Daylight Walking Small	Targets Present	228	228	228	228	228	228	228	228	228	228
	Targets Identified	96.16	139.33	152.16	154.50	176.66	158	189.50	160.33	200	198.83
	Percent Correct	42.17	61.11	66.73	67.76	77.48	69.29	83.11	70.32	87.71	87.20
Dim Walking Large	Targets Present	228	228	228	228	228	228	228	228	228	228
	Targets Identified	84.50	139.33	118.33	174.33	170.83	202.33	186	210.50	201.16	201.16
	Percent Correct	37.06	61.11	51.90	76.46	74.92	88.74	81.57	92.32	88.23	88.23
Dark Walking Large	Targets Present	228	228	228	228	228	228	228	228	228	228
	Targets Identified	55.33	76.33	44.83	111.33	83.33	131.16	113.66	128.83	126.50	118.33
	Percent Correct	24.26	33.47	19.66	48.83	36.54	57.52	49.85	56.50	55.48	51.90

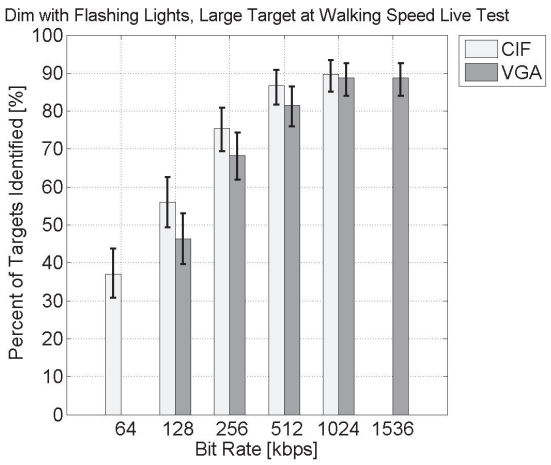
Appendix F Live Test Results

Note: In Figure 19 and Table 10, targets detected and percent correct have been adjusted to account for guesses made by participants in the study.

Figure 19: Live video data (9 charts).







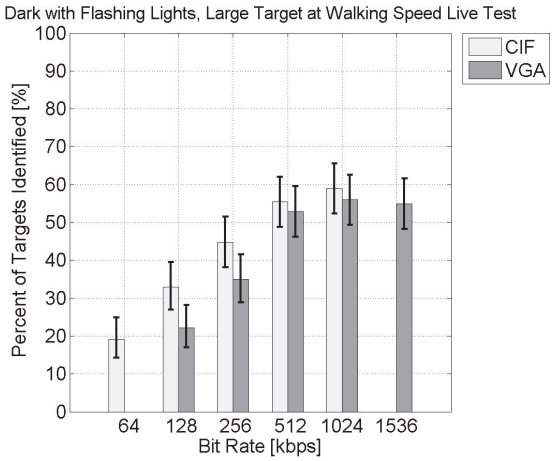


Table 10: Live video data.

	Bit rate [kbps]	64	128		256		512		1024		1536
	Resolution	CIF	CIF	VGA	CIF	VGA	CIF	VGA	CIF	VGA	VGA
Daylight Stationary Large	Targets Present	114	114	114	114	114	114	114	114	114	114
	Targets Identified	101.17	111.16	110.50	112.83	112.83	114.00	114.00	111.67	112.83	114.00
	Percent Correct	88.74	97.95	96.93	98.98	98.98	100.00	100.00	97.95	98.98	100.00
Bright Walking Large	Targets Present	228	228	228	228	228	228	228	228	228	228
	Targets Identified	196.50	224.50	204.67	225.67	226.83	222.17	225.67	225.67	226.83	224.50
	Percent Correct	86.18	98.46	89.77	98.98	99.49	97.44	98.98	98.98	99.49	98.46

Table 10: Live video data. (Continued)

	Bit rate [kbps]	64	128		256		512		1024		1536
	Resolution	CIF	CIF	VGA	CIF	VGA	CIF	VGA	CIF	VGA	VGA
Daylight Walking Large	Targets Present	228	228	228	228	228	228	228	228	228	228
	Targets Identified	187.17	204.67	208.17	211.67	217.50	216.33	223.33	216.33	222.17	224.50
	Percent Correct	82.09	89.77	91.30	92.84	95.39	94.88	97.95	94.88	97.44	98.46
Dim Stationary Large	Targets Present	114	114	114	114	114	114	114	114	114	114
	Targets Identified	87.33	101.17	87.17	103.50	102.33	103.50	104.67	103.50	104.67	104.67
	Percent Correct	71.35	88.74	76.46	90.79	89.77	90.79	91.81	90.79	91.81	91.81
Dark Stationary Large	Targets Present	114	114	114	114	114	114	114	114	114	114
	Targets Identified	21.83	58.00	38.17	90.67	63.83	90.67	84.83	88.33	90.67	94.17
	Percent Correct	19.15	50.88	33.48	79.53	55.99	79.53	74.42	77.49	79.53	82.60
Daylight Stationary Small	Targets Present	114	114	114	114	114	114	114	114	114	114
	Targets Identified	59.17	84.83	91.83	100.00	100.00	102.33	103.50	100.00	110.50	105.83
	Percent Correct	51.90	74.42	80.56	87.72	87.72	89.77	90.79	87.72	96.93	92.84
Daylight Walking Small	Targets Present	228	228	228	228	228	228	228	228	228	228
	Targets Identified	86.83	127.67	148.67	139.33	170.83	142.83	188.33	158.00	168.50	186.00
	Percent Correct	38.08	55.99	65.20	61.11	74.93	62.65	82.60	69.30	73.90	81.58
Dim Walking Large	Targets Present	228	228	228	228	228	228	228	228	228	228
	Targets Identified	84.50	127.67	105.50	172.00	155.67	197.67	186.00	204.67	202.33	202.33
	Percent Correct	37.06	55.99	46.27	75.44	68.27	86.70	81.58	89.77	88.74	88.74

Table 10: Live video data. (Continued)

	Bit rate [kbps]	64	128		256		512		1024		1536
	Resolution	CIF	CIF	VGA	CIF	VGA	CIF	VGA	CIF	VGA	VGA
Dark Walking Large	Targets Present	228	228	228	228	228	228	228	228	228	228
	Targets Identified	43.67	75.17	50.67	102.00	79.83	126.50	120.67	134.67	127.67	125.33
	Percent Correct	19.15	32.97	22.22	44.74	35.01	55.48	52.92	59.06	55.99	54.97