EFFECTS OF STREAMING VIDEO QUALITY OF SERVICE ON SPATIAL COMPREHENSION IN A RECONNAISSANCE TASK

Rudolph P. Darken, MAJ Kurt Kempster, & Barry Peterson
MOVES Institute, Naval Postgraduate School
Monterey, CA 93943

ABSTRACT

It has been proposed that if we could configure individual personnel with micro-video cameras and wireless communications such that they could transmit a video stream of what they were seeing to a remote observer, this would be an enormous improvement in reconnaissance and battlefield command and control. We looked ahead, based on current video and wireless communications technologies and trends to what we can expect to have available in terms of streaming video quality of service (QOS) and we used those predictions to conduct an experiment to determine if this assertion of improvement is true. Participants viewed a digital video with a data rate associated with a given transmission technology. They were asked to maintain their orientation by tracking the position of the camera on a paper floor plan diagram. They were also asked to identify a number of objects and place them in the correct room on the floor plan. The results show that participants found all conditions except the live walkthrough control condition to be extremely difficult with poor performance on both the spatial orientation task and the object identification task. Bandwidth does affect error as increased data rate improves performance. Rapid head rotations seem to be the largest contributor to disorientation, especially with low data rate video. Our results suggest that simply supplying video feedback to a remote observer may be useless at best or possibly damaging at worst. What is needed is not necessarily more bandwidth, but better interfaces and tools to help observers to remain oriented such that they can extract what is needed from the video stream.

Authors' Biographies

RUDOLPH P. DARKEN is an Assistant Professor of Computer Science at the Naval Postgraduate School in Monterey, California. He joined the department in July of 1996 and also holds a joint appointment to the Modeling, Virtual Environments, and Simulation (MOVES) Academic Group as Academic Associate, Academic Council Representative, and chair of the Human-Computer Interaction academic track. He is the technical director for Human-Computer Interaction and Technologies for Immersion for the MOVES Institute. He is a Senior Editor of PRESENCE Journal, the MIT Press journal of teleoperators and virtual environments. He is also the creator of PRESENCE-Lite, the on-line companion to PRESENCE Journal. He received his B.S. in Computer Science Engineering from the University of Illinois at Chicago in 1990 and his M.S. and D.Sc. degrees in Computer Science from The George Washington University in 1993 and 1995, respectively.

MAJOR KURT KEMPSTER graduated from Rollins College, Winter Park, Florida in 1985 and was commissioned in 1989. He has served with the 7th Engineer Support Battalion, 1st Force Service Support Group, Camp Pendleton, California, the 1st Recruit Training Battalion, Marine Corps Recruit Depot, San Diego, California, the Amphibious Warfare School, Quantico, Virginia, and the 1st Combat Engineer Battalion, 1st Marine Division, Camp Pendleton, California. He earned a Masters of Science in Information Technology Management from the Naval Postgraduate School in 2000. His current assignment is as Information Architecture Branch Head, Marine Corps Internet Technology and Network Operations Center, Quantico Virginia. His personal awards include the Navy Commendation Medal with "V" for valor with two gold stars, the Combat Action Ribbon, Sea Service Deployment Ribbon, Drill Instructor Ribbon, Humanitarian Service Medal and the National Defense Medal. His campaign medals include the Southwest Asia Service Medal, the Kuwait Liberation Medal and the Kuwait Liberation Medal.

BARRY PETERSON is a Research Assistant in the Modeling, Virtual Environments, and Simulation Academic Group at the Naval Postgraduate School in Monterey, California. Since 1995, he has been involved with the human factors of computer interaction and specifically conducting research in virtual environment training, completing his M.S. in Engineering at the University of Washington in 1998. Previously, upon graduation from the United States Military Academy, he was commissioned an infantry officer in the Army, where he gained experience in both performing and training infantry navigation.

EFFECTS OF STREAMING VIDEO QUALITY OF SERVICE ON SPATIAL COMPREHENSION IN A RECONNAISSANCE TASK

Rudolph P. Darken, MAJ Kurt Kempster, & Barry Peterson
MOVES Institute, Naval Postgraduate School
Monterey, CA

ABSTRACT

It has been proposed that if we could configure individual personnel with micro-video cameras and wireless communications such that they could transmit a video stream of what they were seeing to a remote observer, this would be an enormous improvement in reconnaissance and battlefield command and control. We looked ahead, based on current video and wireless communications technologies and trends to what we can expect to have available in terms of streaming video quality of service (QOS) and we used those predictions to conduct an experiment to determine if this assertion of improvement is true. Participants viewed a digital video with a data rate associated with a given transmission technology. They were asked to maintain their orientation by tracking the position of the camera on a paper floor plan diagram. They were also asked to identify a number of objects and place them in the correct room on the floor plan. The results show that participants found all conditions except the live walkthrough control condition to be extremely difficult with poor performance on both the spatial orientation task and the object identification task. Bandwidth does affect error as increased data rate improves performance. Rapid head rotations seem to be the largest contributor to disorientation, especially with low data rate video. Our results suggest that simply supplying video feedback to a remote observer may be useless at best or possibly damaging at worst. What is needed is not necessarily more bandwidth, but better interfaces and tools to help observers to remain oriented such that they can extract what is needed from the video stream.

INTRODUCTION

It has been proposed that if we could configure individual personnel with micro-video cameras and wireless communications such that they could transmit a video stream of what they were seeing to a remote observer, this would be an enormous improvement in reconnaissance and battlefield command and control. It is suggested that allowing a commander to “jack in” to a soldier’s live camera would allow him to virtually “be there”, to see what the soldier sees and be able to extract necessary information as needed. This is one of the many technologies utilized in the Army’s Land Warrior concept (Graves, 2001). The soldier is outfitted with a helmet mounted camera and can stream either digital video or thermal imagery as needed to any other member on the wireless network (see Figure 1). What is the expected operational improvement that this capability will facilitate?

We know that rapid decision making is a key factor in successful modern warfare. Boyd’s cycle, commonly known as the OODA loop (Observation, Orientation, Decision, and Action), states that when two opposing forces threaten one another, the one that dictates the pace of the cycle will have the advantage. The intention of digital video reconnaissance fits within this framework. It is suggested that if we can get the right information to the right people at the right time so that decisions can be made at an accelerated pace, an advantage will result. Marine Corps doctrine specifically calls for a shortening of the time required to make decisions, plan, coordinate, and communicate (United States Marine Corps, 1996).

Figure 1. The Land Warrior concept demonstration (taken from Graves, 2001). The helmet is configured with a monocle display and cameras with the capability to stream video and thermal imagery during reconnaissance operations.

The assumption here is that streaming video of this type will facilitate an increase in the pace of decision making by allowing a remote observer to obtain spatial and temporal information in a similar fashion to “being there”. While not explicitly stated, the requirement must be for *spatial and temporal information* in order for real-time video to be relevant. The capability already exists, in a much simpler form, for transmitting still imagery for reconnaissance tasks (e.g. Dalziel, 1998). But still imagery is discontinuous in terms of time and space unless additional information is also given (such as a time stamp, latitude/longitude reading, etc.). Clearly, the motivation for streaming digital video for reconnaissance is well founded. The question now raised is whether or not digital streaming video, in and of itself, provides the benefit it is assumed to provide, or if it will require additional information to be useful.

Our approach in this investigation began with the identification of current and burgeoning technologies that will have an impact on the quality of service (QOS) of streaming digital video over a wireless network. Based on the expected bandwidth along with data rates associated with current video compression algorithms, we were able to simulate video quality of a number of alternative technologies. We compressed the same video segment to data rates of 1.5 Mbps, 256 Kbps, 78 Kbps, and 20 Kbps, all encoded at 30 fps. We wanted to know how well an observer could track the movement of a remote camera moving through a building based on only the streaming digital video. The task required the participant to view the video segment while maintaining their orientation on a paper floor plan diagram. They were also given a secondary task of locating a set of objects in the video. They were shown images of these beforehand. The participant was asked to mark the floor plan every fifteen seconds as to where they thought they were at that time. They then were given a set of markers and were asked to place the objects they saw in their respective positions in the environment.

The results of this experiment will illustrate the critical quality levels and augmentations needed in streaming digital video for reconnaissance based on the task requirements of the observer. We define these requirements in terms of *spatial comprehension*, which is the ability to acquire spatial knowledge of a remote space.* We make two primary specific assumptions with regard to the task requirements of the observer; (1) There is a need to know the location of the remote camera, and (2) There is a need to extract spatial information about specific objects and locations from the video stream. (e.g. Where is the emergency exit?, Where are other members of the squad?, Where is the object I'm looking for?) These requirements help to constrain the tasks we used in our experiment while also maintaining relevancy to actual reconnaissance tasks.

The paper begins with a theoretical discussion of spatial comprehension, specifically the acquisition and subsequent representation of spatial knowledge. We then discuss video compression capabilities and current and impending wireless communication technologies in order to determine what video qualities to expect now and in the near future. Lastly, we describe the experiment in detail and conclude with recommendations based on our results on the use of streaming digital video for reconnaissance.

SPATIAL PERCEPTION AND KNOWLEDGE ACQUISITION

We begin our discussion with a survey of what is known about spatial knowledge acquisition in general, with a focus on passive sources such as video and still imagery.

While there are alternative theories, much of the spatial knowledge acquisition and representation literature is founded on the concept of a three-tiered structure of spatial knowledge (Siegel & White, 1975). At its lowest level is *landmark knowledge*, which involves recognition of distinctive features of an object in an environment. When landmarks are connected within the environment, *route knowledge* develops. Here, one or more routes between landmarks are known but there is not a global sense of the space. It is difficult to point between locations with only route knowledge. Route knowledge also tends to be egocentric in nature. Lastly, *configuration (or survey) knowledge* develops when a holistic, well connected representation is present. Configuration knowledge tends to be map-like, geocentric in nature, and allows for the development of new routes between locations because a consistent global coordinate frame has been defined.

Still imagery is capable of developing landmark knowledge but cannot reasonably represent route or survey knowledge due to its static nature. Based on our assumptions of the tasks of the observer, still imagery could possibly designate where the remote camera was located (particularly if coupled with a GPS readout), but it cannot possibly give spatial information concerning the location of other objects relative to the camera in a dynamic situation. Streaming video should be capable of developing route knowledge because it is a representation of a single route through an environment. There may be concerns about the development of configuration knowledge (see discussion below) due to the passive nature of video,

* Spatial comprehension is actually the ability to acquire spatial knowledge of *any* space, virtual or real, remote or local and can be described in terms of levels of spatial knowledge (see next section).

but the necessary information for the tasks of interest here are represented in streaming video.

Active Versus Passive Viewing

Goldin and Thorndyke (1982) showed that spatial knowledge differs based on its source. In particular, they showed that there is a significant difference between active and passive navigation. They studied a variety of alternative sources based on this concept to include active navigation (e.g. driving a car) versus passive navigation (e.g. being a passenger in a car). While it is clear that route knowledge does develop from both active and passive navigation, it develops faster in active navigation and it tends towards configuration knowledge at a more rapid pace.

Video is an extension of passive navigation but is even more restrictive because the observer no longer has control over the view direction. It has not been shown that viewing passive video alone is sufficient to obtain configuration knowledge, but we do know that route knowledge is possible. However, configuration knowledge may not be entirely necessary for typical reconnaissance tasks. Route knowledge may suffice in most cases. We make this assumption based on our analysis of reconnaissance tasks but there may be cases where a full understanding of the entire space is warranted.

Field of View and Frame-Rate

It has been shown that limited field-of-view has a negative effect on a variety of tasks such as target detection (Osgood & Wells, 1991) and self location in an environment (Alfano & Michel, 1990). In general, video formats concentrate their resolution in the center, on the fovea, eliminating peripheral vision entirely. With limited visual feedback, a simple task such as detecting a single body rotation can be difficult under these conditions. However, even if we could increase the field-of-view (e.g using an HDTV video format), this would vastly increase the bandwidth requirements, thus significantly lowering the frame rate available.

Frame rate is intimately tied to optical flow. At very low update rates, streaming imagery loses continuity and also its spatial quality. With large spatio-temporal gaps between frames, the observer is no longer capable of making accurate judgments concerning distances traversed or the relative distance between objects of interest.

We would like to maximize our ability to acquire route knowledge, and preferably survey knowledge if possible, from a video stream. Yet field-of-view and frame rate limitations work against this goal. For the purposes of this study, we assume that field-of-view is fixed and that frame rate is the variable of concern. This is dictated by the compression algorithm and the wireless communication technology used.

STREAMING VIDEO AND WIRELESS COMMUNICATIONS

It is beyond the scope of this paper to delve into the details of video compression algorithms and the strengths and limitations of current technologies. As an example of current standards, Table 1 shows a set of technologies now in use with their associated data rates. EDTV/MPEG-2 is designed for digital broadcast television while HDTV/MPEG-2 is an extension for high-definition television. MPEG-4 is the standard typically used today for streaming digital video.

FORMAT	PARAMETERS	BIT RATE
EDTV/MPEG-2	960x486@30Hz	7-15 Mb/sec
HDTV/MPEG-2	1920x1080@30Hz	20-40 Mb/sec
Multimedia/MPEG-4	160x120@30Hz	9-64 Mb/sec

Table 1: Streaming video formats and associated bandwidth requirements.

The primary bottleneck in streaming digital video is network bandwidth. This is exacerbated by wireless technologies which have lower bandwidth capabilities than conventional wired networks.

The AN/PSC-5 (V) Shadowfire is a military man portable radio capable of supporting wireless bandwidth for this application. The Shadowfire utilizes the AN/PSC-5 Spitfire architecture and can achieve data rates of 76.8 Kbps line of sight, and 56 Kbps using satellite communications.

Looking into the near future, MILSTAR II will be the next generation military satellite communications network. It is expected to be capable of 4.8 Kbps to 1.54 Mbps throughput. Given the wide variance in performance of MILSTAR II, it will be important in the future to account for significant changes in frame rate of streaming video, an issue not addressed by this experiment. Lastly, Teledesic™ is a commercial venture to construct a broadband “Internet in the Sky” network. Teledesic is expected to be able to support millions of simultaneous users with 2 Mbps uplink and 64 Mbps downlink two-way communication.

Since we cannot use either MILSTAR II or Teledesic for this study, we chose surrogate wired technologies with comparable performance so that we could study the effects of quality of service on human performance. We chose a typical T-1 line with 1.5 Mbps data rate in place of MILSTAR II and Teledesic, and we also added

Video Teleconferencing (VTC) with 256 Kbps which falls in between T-1 and Shadowfire performance.

EFFECTS OF QUALITY OF SERVICE ON SPATIAL COMPREHENSION

We conducted an experiment to determine how video degradation due to compression and bandwidth constraints affects the spatial comprehension of the remote viewer. Our objective was to accurately represent the quality of service (QOS) associated with several wireless transmission technologies, both currently available and projected to be available in the near future. The primary task for each participant was to track the position of the camera on a given floor plan while viewing a video stream degraded to a level comparable to a specific wireless transmission technology.

Methods

We obtained access to a typical office building at the Defense Language Institute that would not have been previously seen by our sample population. We created a full resolution, no compression digital video of a walkthrough of the building, to include motions (e.g. pace, head rotations, etc.) typical of a military building clearing action. We then degraded the full resolution video based on specifications of the target wireless transmission media.

We approximated MILSTAR II and Teledesic with a conventional T-1 transmission rate. We also used a condition using the Shadowfire (SF) performance specifications. We added a condition for Video Teleconferencing (VTC) because of the large gap in performance between the SF and T-1 conditions. We also added a condition approximating the minimum (MIN) data rate at which the video could be transmitted in its original form. Lastly, as a control condition, we included a group that walked through the actual building along the same route taken in the video.

The data rate of the streaming video was used as the independent variable in our experiment. Table 2 describes each of the conditions with associated parameters. Data rate indicates the bits per second that are transmitted from the source to the user. These were simulated in our experiment using Windows Media Encoder™. The video streams for each group were exactly identical except for the data rate. An interesting artifact of the compression algorithm is that the VTC condition, at 256 Kbps actually has a higher actual frame rate than the T-1 condition, at 1.5 Mbps.

VIDEO STREAM	DATA RATE	FILE SIZE	ACTUAL FRAME RATE
Live (LIVE)	N/A	N/A	N/A
T-1	1.5 Mbps	17.24 MB	17.9 fps
VTC	256 Kbps	2.78 MB	21.37 fps
Shadowfire (SF)	78 Kbps	869 KB	6.71 fps
Minimum (MIN)	20 Kbps	230 KB	1.43 fps

Table 2. Video compression conditions used in the experiment.

As a secondary task, we wanted to know if participants could identify objects seen in the video stream. We captured five images directly from the video stream, each of which was shown to the participants before beginning the first trial. Figures 2 shows three of these images.

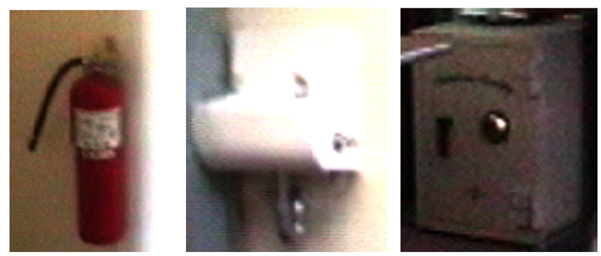

Figure 2. The fire extinguisher, drinking fountain, and floor safe, images taken from the video.

We screened our participants to assure that they had never been inside the target building before. We then randomly divided our participants into five experimental groups, one for each of the four data rates plus one control group that would do a live walkthrough of the building along the same path as the video.

Participants viewed the video stream at 320 by 240 pixel resolution using Windows Media Player™ on a typical PC with color monitor. Each was provided with the floor plan of the building before beginning the first trial. The starting position was identified on the floor plan. They were also shown the images they were to look for during each trial. Our intention was to try to replicate the media that a commander might have at his disposal while viewing live steaming video from a reconnaissance unit. Assuming very good intelligence had been gathered (an unlikely occurrence), the commander might have an accurate floor plan and possibly images or sketches of objects of interest. Our experiment can be considered a "best case scenario" for this type of mission.

Participants were told what they would be doing and that they needed to track their position in the building as well as look for the target objects along the way. The experiment was a repeated measure between groups design, with each participant viewing only one video stream but completing the entire task sequence twice. We did this because we wanted to know if there might be a learning curve or familiarity process that might take place. We knew the task was difficult but wanted to know how much performance would improve if participants were given a second attempt. There are also strategies involved in the task. A participant may try a poor strategy on the first attempt and thus might be expected to improve if given a second trial.

During the trial, participants were asked to mark the current position and view direction every fifteen seconds. The video was never stopped during a trial. This was done to simulate the conditions a commander would likely encounter in the proposed implementation of "real time" streaming video. After each trial, we asked participants to rate the difficulty of the task on a six point scale, with one being "very easy" and six being "extremely difficult".

We made several assumptions in this experimental design. The video quality and frame rate in each of our conditions remains constant over the duration of the trial. This is not realistic. In practice, quality can change significantly based on network traffic, or in the case of wireless communications, atmospheric or terrain interference. The degradation schemes we used were based on published bandwidth specifications for our chosen transmission technologies. Each stream was encoded using Windows Media Encoder in single data rate mode. This method allowed us to encode each stream at the expected data rate for the simulated network. Using this method optimized the video for the data rate selected, ensuring the highest quality possible; a best case scenario. This is not the only way to limit data rate. However, our secondary task required that objects have enough pixel resolution to be identifiable. Therefore, we chose to maintain a reasonable pixel resolution while varying temporal resolution (e.g. frame rate).

The sample population was 40 student officers from the Naval Postgraduate School with between one and nineteen years of military experience. All participants took part on a volunteer basis and no compensation was given.

Analysis

Figure 3 shows the building floor plan with the actual camera path identified. The check points (at fifteen second intervals) are marked one through six. The start point is marked with an S. There were five different types of objects that participants were asked to look for in the video. There were two floor heaters (A), a laser printer (B), two fire extinguishers (C), a floor safe (D), and a drinking fountain (E).

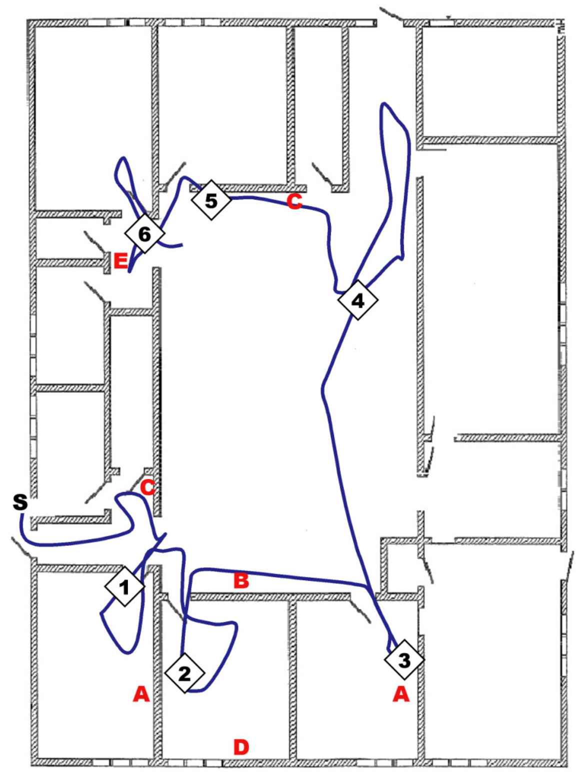

Figure 3. The building floor plan showing the actual route, check points (numbered), and object locations (lettered).

Figure 4 shows a typical floor plan marked by a participant. The solid line indicates the path they thought was taken by the camera and the circles indicate the positions where they thought they were at the fifteen second intervals. Note that our instructions do not require that the entire path be drawn on the floor plan during viewing. Participants were asked only to mark the location of the camera at the six chosen check points. Nevertheless, many participants chose to draw the entire path while they viewed the video (see discussion below on strategy).

We added a four foot error buffer around the marked position to allow us to differentiate between lack of precision and error. Each of the six points marked on the floor plan have a circle around them indicating this buffer. We added the actual positions of the check points for comparison here but these were not seen by the participant. The dashed line indicates the error associated with each location mark. The error is measured as the distance from the actual location (indicated by the diamond) to the nearest edge of the outside circle. Lastly, the object markers they were able to place are indicated by the letters. Most participants were not able to place all the objects on the floor plan with more precision than a guess. Our analysis accounts only for how many objects they were able to identify and place in the correct room. This was an extremely difficult part of the task.

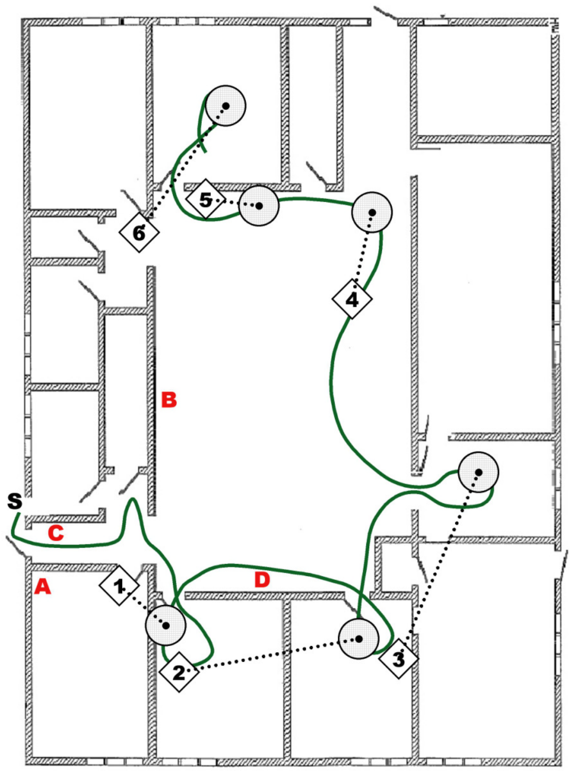

Figure 4. The building floor plan showing the path as indicated by a participant, the check points as indicated by the participant (circles), the actual check point locations (numbered), and the object locations as indicated by the participant (lettered).

Results & Discussion

In the following discussion and figures, we refer to the conditions as live (LIVE), T-1, VTC, Shadowfire (SF), and minimum (MIN).

Combining the errors across each of the six check points in the video gives an aggregate measure of total error. This allows for a compounding of errors which we will address in another analysis, but as a coarse measure of performance, total error does reflect the participants' ability to remain oriented in the environment. An analysis of variance (ANOVA) of total error per trial shows a significant difference across conditions ($F(4,35)=8.66$), $\underline{P}\leq.0001$, see Figure 5).

Other than the LIVE condition, the best performance was seen in the T-1 condition over the VTC condition in spite of the fact that the VTC condition had a better frame rate. The T-1 condition may have had slightly superior clarity but we had no way to assess this other than subjectively. We conclude that since the difference seen between VTC and T-1 is not statistically significant, there is likely no difference between them in terms of performance.

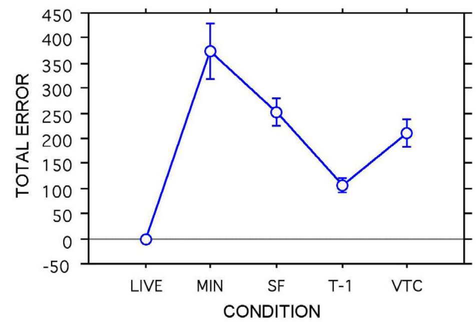

Figure 5. Between group results of total errors.

While performance does improve from the first to the second trial, this difference was not shown to be significant (see Figure 6). We conclude that any learning effect is minimal. If we had considered a second trial with a new video stream, we could expect performance to be even lower than shown here, likely eliminating any improvement between trials entirely.

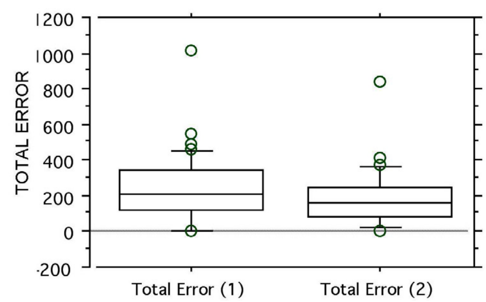

Figure 6. Box plot showing total error results of both trials.

This same analysis was done on the number of objects participants were able to identify and place on the floor plan at the conclusion of each trial. We see similar results ($F(4,35)=38.33$), $\underline{P}\leq.0001$, see Figure 7).

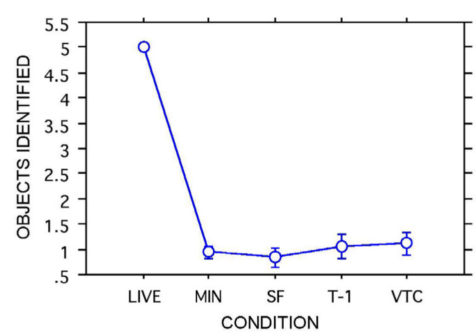

Figure 7. Between group results of objects identified.

Interactions between all the video conditions are insignificant. Participants in the LIVE condition accurately identified all objects without error while very few participants in any other group were able to complete this task. This is a difficult comparison to make since the LIVE group had information none of the video groups had, such as kinesthetic feedback from walking through the building and the ability to turn their head to actively search the space for objects. We conclude from this analysis only that the object identification portion of the task was nearly impossible to do given the demands of the primary route tracking task.

Figure 8 shows that while there was again a trend towards improvement from the first to the second trial, it is not significant suggesting that there is little if any learning effect.

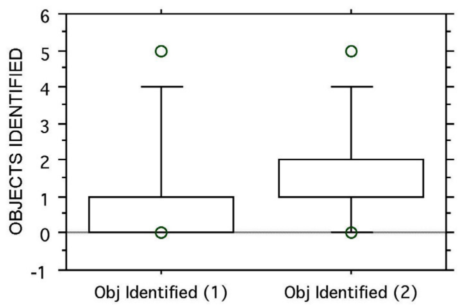

Figure 8. Box plot showing total objects identified results of both trials.

These results suggest that any degradation beyond live interaction causes large negative effects in performance in terms of spatial orientation. Data rates at T-1 and VTC levels may result in tolerable performance on the spatial orientation part of the task, but when we consider the ability of the viewer to identify objects from the video, performance across all conditions except the live condition was extremely poor. It was evident that the primary task of maintaining spatial orientation during the video viewing was an extremely demanding task, resulting in poor performance on the secondary task. Had we adjusted the primary task to make it easier, such as by using a less cluttered or less complex space, we expect that performance on the secondary task would have improved.

Participants' subjective evaluation of the difficulty of the task correlates well with their performance. Significant differences were seen between the live condition and all others ($F(4,35)=31.54$), $\underline{P}\leq.0001$, see Figure 9).

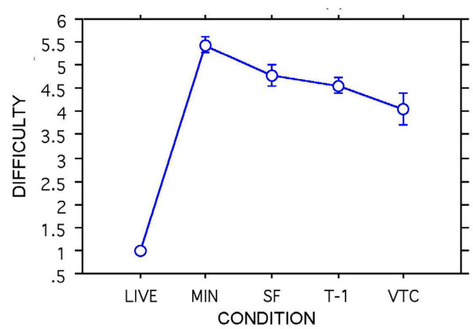

Figure 9. Between group results of task difficulty.

Differences between the video conditions were not significant. This also correlates well with performance in terms of total errors ($F(1,78)=29.01$, $\underline{P}\leq000.1$, see Figure 10). Even though participants had received no feedback as far as how well they did on the task before doing the subjective assessment, they anticipated the results in terms of how confident they were in tracking the route and placing the objects. As total error increase, so also does their subjective assessment of task difficulty.

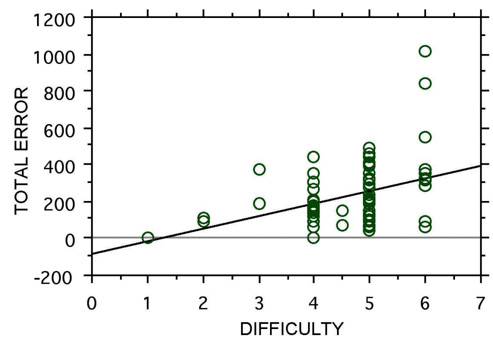

Figure 10. Correlation between total error and difficulty.

To determine the possible cause of disorientation in the video stream groups, we looked at errors by individual check points rather than in an aggregate measure as we did previously. Here we see that there is an increase in errors somewhere between the third and fourth check points. This coincides with a portion of the video where the camera turns down toward the floor and pans very quickly as it exits the room where the third check point occurred. This combination of simultaneous rapid changes in pitch, yaw, and linear movement likely caused disorientation in the viewers that was very difficult to recover from. The higher bandwidth groups were able to compensate better but the low bandwidth groups often never caught up and consequently were lost for the duration of the video stream. This is illustrated in Figure 11.

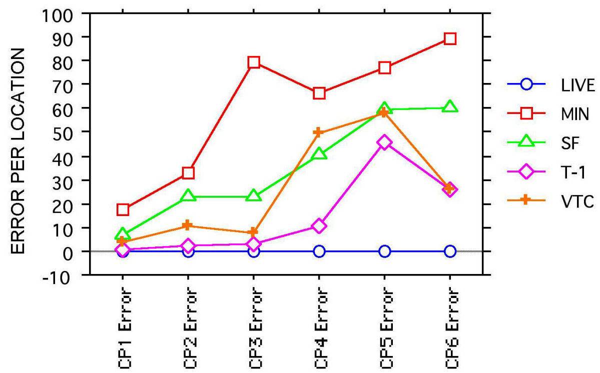

Figure 11. Error viewed by location between groups.

There were two primary strategies used to maintain orientation during the video stream. The first, and most obvious, was to draw the route while viewing the video. In this way, when a checkpoint was reached, the participant would merely mark the current location on the floor plan for that specific check point. The other strategy was to follow along with their finger and then mark the location at the appropriate time. While these strategies may seem to be almost identical, we found that participants who actually drew the route had greater difficulty in maintaining orientation. This was determined mainly by post trial interviews. It seems that the task of maintaining orientation while viewing a digital video stream is so demanding that any distraction will likely cause significant negative effects.

CONCLUSIONS

We begin with recommendations for the use of streaming digital video for reconnaissance tasks based on the results of this experiment and conclude with proposed future work in this area.

Recommendations

The results of this study clearly show that even with impending improvements in bandwidth capabilities we expect to see in the very near future, simply streaming video off of a man in the field to a remote observer and expecting him to extract useful information may be useless at best and possibly damaging at worst. If we presume that people are capable of extracting information from a video stream that they are actually incapable of doing, there will likely be a shortfall in overall execution of the mission. Even under the absolute best of conditions, maintaining spatial comprehension from a video stream is an intensely difficult task requiring complete concentration; conditions that typically do not occur in reality.

Should we conclude that streaming video is operationally ill-advised, or could it be that it requires augmentation to be truly useful? We believe the latter is the case.

At the very least, viewing streaming digital video in real time is inherently problematic for several reasons.

- In real time, the observer may not be afforded the opportunity to review the video, going back to catch something that was missed.
- It is also not possible to slow down the video to compensate for fast movements or complex (or cluttered) environments.
- In real time the frame rate will fluctuate, probably dramatically. Although we do not know what effect this might have on performance from this study, we believe it will degrade performance further.

Consequently, if no augmentation is possible or available, we suggest that streaming video be cached off-line and viewed asynchronously with its capture to address these issues. But even so, our data suggests that the level of cognitive effort exerted in maintaining spatial orientation is so severe, that some augmentation will be needed before we can realize an actual operational benefit to the use of streaming digital video for reconnaissance.

If the video stream were to be augmented with spatial tracking information, such as a moving marker on a floor plan or map, this would greatly simplify the spatial orientation task allowing more attention to be focused on extracting objects from the imagery. This, of course, would require remote sensing technologies, some of which are not yet mature enough to support this requirement. In an outdoor environment, streaming GPS data along with the video may accomplish this end.

Another possibility is to augment the performance of the observer by extracting symbolic information for him. Object recognition algorithms are maturing. It may be timely to think about head-worn cameras and streaming video technologies as more than merely an image transfer mechanism. Consider these technologies packaged as a mission support tool that is capable of identifying objects (e.g. threats, targets, people) in the environment directly from the video and is able to call attention to those objects for the observer, thus lessening the cognitive workload and raising overall performance. This is, in part, the focus of *augmented cognition* (Schmorrow, Worcester, & Patrey, 2001).

Future Research

There are a number of areas that will need to be addressed before we will fully realize the impact of streaming digital video on Boyd's cycle. A number of aspects were not considered in this experiment but will likely have an impact on this issue.

- We did not consider field-of-view in any way, nor did we address the likelihood of changing frame rates during transmission.
- Even though there is no intent to achieve any level of "immersion" via streaming video for the observer, it may be useful to display the video in a head-mounted display or other enclosure thus eliminating distractions from the local area. This could raise performance by focusing attention on the video stream alone.
- We do not yet understand the importance of environmental complexity and clutter on performance. It follows that it will be harder to identify an object in a space among many other objects as opposed to if it is isolated. Is it possible to identify clutter and assist the observer based on task complexity?
- How much would performance improve if we had added a dynamic floor plan that tracked the position of the camera? We would assume that performance on the spatial orientation task would improve because of the direct match to the information given on the floor plan, but how much better would object recognition be?
- What other types of augmentation make sense for this type of application?

We believe that rapid improvements in wireless communications and digital video will continue to make streaming digital video for reconnaissance a more viable operational tool. The experiment described in this paper should serve as a warning, however, against rushing the technology into practice. There will be an extraordinary impact on Boyd's cycle as a result of these technologies, but only when we have determined how best to present information to the observer that directly amplifies his ability to accomplish his task.

ACKNOWLEDGEMENTS

The authors would like to thank the participants who volunteered their time to assist us. Also special thanks to the Defense Language Institute for allowing us to use one of their buildings for our study. Lastly, thanks to Roger Dalziel, A.J. Pasagian, and Jim Munroe for their earlier work supported by the National Reconnaissance Office, which set a solid foundation for this study.

REFERENCES

Alfano, P. & Michel, G. (1990). *Restricting the Field of View: Perceptual and Performance Effects*. Perceptual and Motor Skills, 70, pp. 35-45.

Dalziel, R. (1998). *Improving the Engineer Reconnaissance Reporting Process Through the Use of Digital Imagery and Handheld Computers*. Masters Thesis, Naval Postgraduate School, Monterey, CA.

Goldin, S. E., & Thorndyke, P. W. (1982). *Simulating Navigation for Spatial Knowledge Acquisition*. Human Factors, 24(4), 457-471.

Graves, W.A. (2001). *Land Warrior Live Fire Sheds Light on Army of Future*. http://www-tradoc.army.mil/pao/awe/landwarlfire.htm

Osgood, R. & Wells, M. (1991). *The Effect of Field-of-View Size on Performance of a Simulated Air-to-Ground Night Attack.* Proceedings of Helmet Mounted Displays and Night Vision Goggles, Neuilly sur Seine, France.

Schmorrow, D., Worcester, L, & Patrey, J. (2001) *Augmented Cognition: New Design Principles for Human-Computer Symbiosis*, Proceedings of the International Applied Military Psychology Symposium.

Siegel, A. W., & White, S. H. (1975). *The Development of Spatial Representations of Large-Scale Environments*. In H. Reese (Ed.), Advances in Child Development and Behavior (Vol. 10,). New York: Academic Press.

United States Marine Corps, (1996). *Marine Corps Doctrinal Publication 6, Command and Control*, Washington, D.C.