To Catch a Thief – you need at least 8 frames per second: The impact of frame rates on user performance in a CCTV detection task

Hina Keval
University College London
Department of Computer Science
Gower Street, London, WC1E 6BT, UK
h.keval@cs.ucl.ac.uk

M. Angela Sasse
University College London
Department of Computer Science
Gower Street, London, WC1E 6BT, UK
a.sasse@cs.ucl.ac.uk

ABSTRACT
The new generation of digital CCTV systems can be tailored to serve a wide range of security requirements. However, many digital CCTV systems produce video which is insufficient in quality to support specific security tasks, such as crime detection. We report a study investigating the impact of lowering frame rates on an observer's ability to distinguish between *crime* and *no crime* events from post-event recorded video. 80 participants viewed 32 video scenes at 1, 5, 8, and 12 frames per second (fps). The task required observers to determine if one of three possible events had occurred. Results showed that when the frame rate was lowered from 8fps, the number of correct detections and task confidence decreased significantly. Our results provide CCTV practitioners with a minimum frame rate level (8 fps) for event detection, a task performed by CCTV users of varying skill and experience.

Author Keywords
Multimedia, CCTV video, security, video quality, frame rate, crime detection, task performance.

ACM Classification Keywords
H5.m. Information interfaces and presentation: Multimedia Information Systems.

INTRODUCTION
Closed Circuit Television Video (CCTV) is seen as an important tool in the prevention and investigation of crime, and in the delivery of public and commercial security. Recent advances in technology have opened up a large number of uses for CCTV. High crime rates and recent terrorism events in the US and London have led to large investments being put into CCTV around the world.

Currently, CCTV is undergoing three major changes: 1) the technology platform has moved from analog to digital; 2) a rise in the number and variety of purposes for CCTV; and 3) an increasing number of users involved with CCTV systems, many of which have little or no training (see [1] and [2]).

First-generation CCTV was based on analog technology, where video is recorded from a number of surveillance cameras directly onto a tape recorder. At the time, tape-based (analog) CCTV systems were perceived as easy-to-use and affordable. However, the practice over the past 20 years has revealed many shortcomings (analog video is inefficient to record and replay, requires constant human intervention as tapes need to be changed for continuous recording resulting in low video quality).

In theory, moving from analog to digital CCTV should remove all these problems. But digital CCTV systems can be configured by system owners: they can choose the resolution and frame rate at which video is recorded, and how much bandwidth to assign for the transmission of video from CCTV cameras. There is an inevitable tradeoff between video quality and cost. Uncompressed, high-resolution, and high-frame rate video requires high bandwidth and disk space. The trend has been to spend money on increasing numbers of cameras, rather than increasing network bandwidth and data storage space.

Digital video quality depends on three factors: 1) image resolution; 2) video compression and 3) frame rate. High resolution, high frame rate and a low video compression ratio preserve the spatial and temporal details of a scene. There is very limited guidance on the video quality requirements for CCTV used by human observers for security tasks. The Home Office Operational Requirements Manual [3] provides guidelines on the minimum height a person should appear on a CCTV monitor for five observation tasks: 1) monitor; 2) control; 3) detection; 4) identification and 5) recognition. Due to the lack of research in this area, these guidelines focus mainly on analog CCTV systems – therefore they do not consider the requirements for digital and networked CCTV systems.

In the next section, a critical review of the previous research on video frame rate and task performance is provided.

Background

Video recorded at low frame rates (i.e. 5 fps and below) can result in important frames being discarded from a video scene. A common rule of thumb is implied that, *"[f]rame rates must be greater than 8 to 10 frames/s,"* [4]. CCTV video recorded in time-lapse mode is commonly recorded at frame rates between 5-8 fps [3]. Despite this, many CCTV owners are recording at very low frame rates some as low as ¼ fps [5].

To date, only one experiment has examined the impact of low frame rates on a crime detection task [6]: a number of crime events were staged in public by actors and police officers outside a railway station. Incidents included bag snatches, cell phone snatches, and threatening incidents. The test was carried out with two groups of expert CCTV users: 1) 22 police academy students and 2) 16 CCTV operators in a control room. Task performance was measured under the following frame rate levels: 25, 2 and ¼ fps. The results showed that, for both groups as video frame rate decreased, task performance also decreased. A review of the study design identified two major issues: 1) task performance was not measured appropriately, instead a software tool for training CCTV operators was used to determine detection and 2) the frame rate levels chosen were not appropriate as they do not allow for a threshold frame rate to be identified.

A number of studies have investigated the impact of using low-frame rate video on human performance in other application domains:

1. In a lie detection study [7], participants were asked to judge whether a person was lying or not in an interview observed on video. Performance in lie detection was measured using a 6-point Likert truthfulness scale when video was played under the following conditions: 320x240 and 29.97 fps; 106x80 and 10 fps; 106x80 and 5 fps; 53x40 and 10 fps; and 53x40 and 5 fps. It was found that a slight reduction in image resolution impaired lie detection accuracy; however performance suffered the most when video was played back at 5 fps. At this rate, behavioral actions such as shoulder shrugs, eyebrow movements, and eye blinks were difficult to interpret.

2. In a gaming study [8], participants were asked to fly a vehicle in a simulator that displayed animated video at different frame rates. Task performance deteriorated as frame rates were lowered from: 7.5, 4, and 2 fps – with 4 fps being enough for a gaming task.

These two frame rate studies investigate the relationship between video frame rate and the user's ability to perform tasks with video. Clearly, the frame rate required for a given task is dependent on the perceptual demands of the user's task as well as the content being shown.

It may seem that the ability to detect facial movements and body language requires higher temporal resolution, thus higher quality - but the indicators of some crimes can be subtle and hard to detect. Therefore, detecting these types of subtle actions from low frame rate video could prove challenging for users - particularly untrained participants. This research problem was investigated through an empirical study which is described in the following section

DETECTION STUDY

We designed a study to evaluate user performance and task confidence when detecting *crime* and *no crime* events from CCTV video at four video frame rate levels: 1 fps; 5 fps; 8 fps and 12 fps. The hypotheses were:

Hypothesis 1 (H1): As the frame rate of CCTV video is lowered, the number of correct detections will decrease.

Hypothesis 2 (H2): As the frame rate of CCTV video is lowered, the observer's confidence in their choices and their perceived ratings of video quality will decrease.

MATERIALS

Preparation of CCTV Video Stimuli for Detection Test

32 scenes (~10 seconds) were recorded with 8 actors. 50% of the scenes contained crime events and the other half contained no crime. All of the video scenes were filmed outdoors using a digital video camera fixed on a stable tripod (1.53m height). The camera height was chosen at eye level for practical reasons and this was approved by 2 video and image experts who work in the security and surveillance domain.

Each recording was edited in segment length and to typical spatial video quality (CIF resolution/MPEG-4/2000 Kbps). The video was then lowered into the frame rate conditions (1, 5, 8, and 12 fps). The lowest condition (1 fps) was chosen as this is what most systems offer as the lowest recording and streamed frame rate. 5 fps was chosen as this was the threshold frame rate identified in the lie detection study [7]. The frame rate levels 8 fps and 12 fps were chosen as these were the next two equal levels on the frame rate scale. The highest frame rate condition - 12 fps was chosen as this is the point at which video is generally perceived by humans as smooth in appearance.

Measures

Three measures were recorded from each participant: 1) task performance (average number of correct detections); 2) task confidence (using a 5-point Likert scale, 1 - not

very confident to 5 - very confident); and 3) perceived video quality (using a 5-point Likert scale, 1 - very bad to 5 - very good).

DESIGN

A 1x4 within-subjects design was used. The video frame rate conditions were counterbalanced appropriately to avoid order and practice effects. The ordering of the conditions was made using a 4x4 Graeco-Latin square design to create four test blocks. Twenty participants completed each test block. Figure 1 illustrates one of the CCTV scenes taken from the 'crime' video set. This clip was given with the following scenarios:

A: The person wearing a rucksack stops and asks a passer-by what time it is.

B: A person steals something from the passer-by who is wearing a rucksack. ** Correct scenario **

C: A person walking behind the rucksack carrier attempts to steal something from him but does not succeed.

Participants were asked specifically to pick a scenario instead of giving an open-ended response so that performance could be objectively measured.

Figure 1: Crime clip (theft from rucksack)

PARTICIPANTS

80 participants (31 males, mean age 29 years) who had no prior experience of CCTV tasks were recruited. Each participant was paid £5 Sterling (~$10). The actors who appeared in the CCTV scenes were not permitted to take part in the experiment.

EXPERIMENTAL PROCEDURE

Participants were seated in front of a 21 inch PC monitor in a computer laboratory. Each participant was shown two CCTV scenes with 3 scenarios for practice. Participants then proceeded with the detection test where they were shown 32 CCTV scenes and required to choose one scenario from three which best described the scene for each clip, rate their confidence in their decision and then rate the perceived video quality.

RESULTS

As CCTV frame rate increased, the average number of correct detections increased from 39.2% (at 1 fps) to 65% (at 12 fps) – see Figure 2. A Friedman test showed that that task performance significantly increased as video frame rate increased [$\chi 2$ (3) = 63.4, p<0.001].

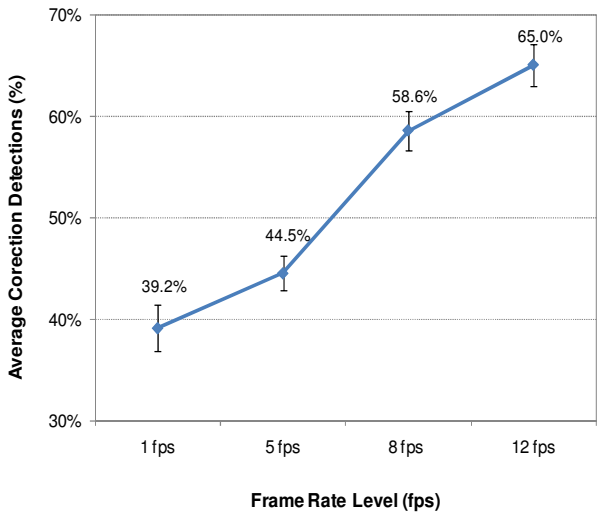

Figure 2: Average task performance across frame rate levels

Ratings: Perceived Video Quality and Confidence

Figure 5 shows that as CCTV frame rate increased, the average perceived video quality and confidence ratings also increased (1 fps to 12 fps).

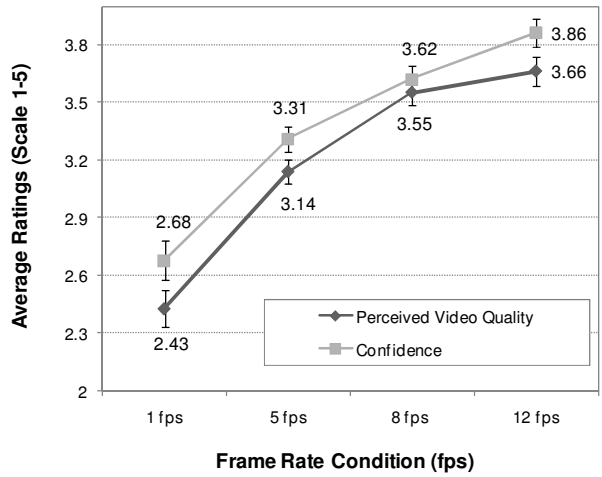

Figure 3: Perceived video quality and confidence ratings

A follow up test was carried out to determine the actual significant difference in task performance between the frame rate conditions. The results showed a significant difference in detection performance between frame rate

levels 1 and 5, 5 and 8 fps ($p<0.0001$), but not between 8 and 12 fps ($p = 0.294$).

A Friedman test showed that the confidence ratings were significant as CCTV frame rate increased [$\chi 2\ (3) = 63.4$, $p<0.001$] and comparisons also revealed significant differences between all the frame rate levels. The perceived video quality ratings were also found significant [$\chi 2\ (3) = 109.8$, $p <0.001$] however, individual comparisons revealed no significance between the perceived video quality ratings from 8 to 12 fps ($p = 0.818$).

DISCUSSION & CONCLUSIONS

Our results support H1: as CCTV frame rate was lowered, the number of correct event detections decreased significantly. Participants' confidence in their ability to detect events, and their video quality ratings also decreased. However, detection performance and the perceived video quality ratings were not significant between 8 and 12 fps, this finding was unexpected. Although there was a significant difference found between 1-5 fps, this work furthers the work by Horn [7] – in that detection performance is actually worst at 1 fps and not 5 fps.

Our study goes beyond previous research [6], as it considers a new and emerging CCTV user scenario [1 and 2] and evaluated commonly applied frame rates within real-world settings [3]. Our results therefore provide CCTV practitioners and owners with an optimum frame rate (8 fps) to apply for real-time detection and post-event analysis of events from recorded CCTV video.

There are serious implications in using low temporal video quality for a detection task. For instance, at 5 fps, depending on the speed of the activity observed within a scene, potentially important frames may be discarded by the video encoder. Without having access to parts of a scene when observing CCTV in real-time, the observer's confidence will be poor which will reduce their vigilance levels and responses to incidents. Equally, when examining post-event CCTV video following an incident, the observer will not be able to confidently depict actions of targets, reducing the strength of CCTV evidence when needed to prosecute unknown criminals.

Future Work

Our future work will involve repeating the current study with a wider range of scenarios (crime and no crime) with the aim of yielding representative results for real-world application. The CCTV data can then be used by other multimedia researchers for evaluating CCTV performance with other variables such as resolution and video CODECS etc.

A further study is under planning which will evaluate the difference in detection performance between different groups of CCTV observers: 1) CCTV operators; 2) police staff; and 3) untrained and inexperienced participants. A direct comparison between these groups will determine whether task experience can overcome the problems with low temporal CCTV video quality.

ACKNOWLEDGEMENTS

This research was funded by an EPSRC studentship (grant no: EP1P50064811). We would like to thank Jay Gattuso and Ken-Brown MacLennan from the UK Home Office Scientific Development Branch for offering advice on CCTV video stimuli preparation. We would also like to thank all actors who volunteered for the CCTV filming.

REFERENCES

[1] Web users to 'patrol' US border, BBC News. http://news.bbc.co.uk/1/hi/world/americas/5040372.stm

[2] Rights group criticises 'Asbo TV', BBC News. http://news.bbc.co.uk/1/hi/england/london/4597990.stm

[3] Cohen, N., Gatusso, J., and MacLennan-Brown, K. CCTV Operational Requirements Manual - Is your CCTV system fit for purpose? Vers. 4, (55/06), Home Office Scientific Development Branch (HOSDB), 2006.

[4] Card, S., Moran, T., and Newell, A. Handbook of perception and human performance, In K. Boff, L. Kaufman, & J. Thomas (Eds.), Vol. 2, New York, NY: Wiley, 1986, 451-3.

[5] Gill, M., Spriggs, A., Allen, J., Jessiman, P., Swain, D., Hemming, M., Kara, D., and Little, R. Control room operation: findings from control room observations, Home Office Scientific Development Branch Publication, 17/05, 2005.

[6] van Voorthuijsen, G., van Hoof, H., Klima, M., Roubik, K., Bernas, M and Pata, P. CCTV Effectiveness Study. In Proc. of 39 IEEE ICCST, Piscataway, 2005.

[7] Horn, D. B., Karasik, L., and Olsen, J. S. The effects of spatial and temporal video distortion on lie detection performance. In Proc. CHI '02 Extended Abstracts on Human Factors in Computing Systems, ACM Press, New York, NY, 714-715, 2002.

[8] Swartz, M. and Wallace, D. Effects of Frame Rate and Resolution Reduction on Human Performance, Proc. IS&T's 46th Annual Conference, 1993.

[9] Claypool, M., Claypool, K., and Damaa, F. The Effects of Frame Rate and Resolution on Users Playing First Person Shooter Games, Proc. ACM/SPIE Multimedia Computing and Networking (MMCN), 2006.