Can we ID from CCTV: image quality in digital CCTV and face identification performance

Hina U. Keval*, M. Angela Sasse
Dept. of Computer Science, University College London, Gower St, London WC1E 6BT, UK

ABSTRACT

CCTV is used for an increasing number of purposes, and the new generation of digital systems can be tailored to serve a wide range of security requirements. However, configuration decisions are often made without considering specific task requirements, e.g. the video quality needed for reliable person identification. Our study investigated the relationship between video quality and the ability of untrained viewers to identify faces from digital CCTV images. The task required 80 participants to identify 64 faces belonging to 4 different ethnicities. Participants compared face images taken from a high quality photographs and low quality CCTV stills, which were recorded at 4 different video quality bit rates (32, 52, 72 and 92 Kbps). We found that the number of correct identifications decreased by 12 (~18%) as MPEG-4 quality decreased from 92 to 32 Kbps, and by 4 (~6%) as Wavelet video quality decreased from 92 to 32 Kbps. To achieve reliable and effective face identification, we recommend that MPEG-4 CCTV systems should be used over Wavelet, and video quality should not be lowered below 52 Kbps during video compression. We discuss the practical implications of these results for security, and contribute a contextual methodology for assessing CCTV video quality.

Keywords: Security, digital CCTV, MPEG-4, Wavelet, face identification, task performance, video quality, video compression.

1. INTRODUCTION

Closed Circuit Television Video (CCTV) is seen as an important tool for crime prevention and investigation. Currently, CCTV is undergoing immense changes: (1) the technology platform is moving from analogue to digital; (2) the number and variety of purposes for which it is installed is increasing; (3) the number of users interacting with CCTV is increasing and (4) the range of CCTV users are becoming increasingly diverse.

The first generation of CCTV systems was based on analogue technology, where video is recorded from a number of surveillance cameras directly onto a Video Cassette Recorder (VCR), and physically stored onto tapes. At the time, tape-based CCTV systems were perceived as easy-to-use and affordable, but practice over the past twenty years has revealed many shortcomings:

1. Recording on tape is inefficient. For instance, it can be very time consuming to search, copy and store analogue video compared with digital video. Inefficient access and retrieval to surveillance footage can slow down criminal investigations substantially.
2. VCR systems require constant human intervention. For instance tapes need to be constantly replaced to allow for continuous recording. This process is time consuming and susceptible to human error (i.e. tapes can be unintentionally over-written or misplaced).
3. Analogue video has limited resolution (240 to 340 TV Lines). In contrast, digital CCTV video is capable of recording up to 400 TV Lines. Digital video is also flexible as it can be recorded under different quality rates, by changing the recording and streaming quality (video resolution, frame rate and compression). The choice of quality depends on the user's system and budget requirements.
4. Analogue CCTV systems do not offer remote video access, whereas digital systems can be linked to the Internet to transmit video footage for remote monitoring and surveillance purposes.

*h.keval@cs.ucl.ac.uk.

In the past, CCTV video and images were mainly used by expert CCTV users: CCTV operators, police investigators and forensic experts. Today, an increasing number of systems are deployed by public and commercial sector organizations to serve an increasing number of security tasks, and there is now a wider range of CCTV users, with different levels of skills and experience. In fact, there are two main groups of CCTV users who directly interact with the system: (1) CCTV *owners*: who purchase, own and are therefore responsible for the operational performance of the system and (2) CCTV *users*: who perform security observation tasks with CCTV video. These users typically include the police, forensic experts, CCTV operators and now even untrained CCTV users. These untrained CCTV users have emerged as a result of the Internet widening access to CCTV video. It has been found that some CCTV owners (in the US and UK) are now aiming to recruit untrained members of the public to assist with security tasks:

1. In the US, web users can access real-time CCTV video via their web browser to monitor the Texas-Mexico border for illegal crossings. This system was set up to alert the authorities on illegal activities [3].
2. In the UK, east London residents of a housing project can view digital CCTV video from their television sets by subscribing to a community safety channel [2]. If a viewer observes crime being committed, or identifies a person breaking an anti-social behavior order (ASBO), they can immediately alert the police by phone. Viewers can access images of known criminals and trouble-makers by viewing a ‘rogue gallery’ on their TV.

These two projects illustrate the widely-held belief in the effectiveness of wide-spread CCTV monitoring, memorably summarised by Marge Simpson as: *“… the courts may not be working anymore, but as long as everyone is videotaping everyone else, justice will be done.”* The question we raise in this paper is, whether untrained CCTV users are capable of carrying out a face identification task with digital CCTV video effectively.

In the past, face identification was a task typically carried out by trained CCTV users with real-time and recorded CCTV video for investigative and evidential tasks. In this paper, we present a study which focuses examines the relationship between video quality and face identification with untrained CCTV. Furthermore, we investigate how this task is carried out by these users (as per the scenario 2 described above).

Previous research on face identification carried out by human observers [7][8][9][16] had established that both trained and untrained CCTV users are poor at recognizing people (face and body) from low-resolution analogue CCTV video. No work to date has systematically examined the impact of excessive digital video compression when used for security tasks by human observers. The closest area of work that attempts to quantitatively establish the limits of human face recognition performance was carried out by Bachmann [5]. Bachmann found that the identification of faces was affected very little when the spatial quantization (pixilation) of the face image was reduced from 74 pixels to 18 pixels. Participants were asked to match six different non-quantized face images with a collection of computer-quantized face images. The quantization conditions used were: 15, 18, 21, 24, 32, 44, and 74 pixels per face. It was concluded that, a face image at a total pixel count of 18 was considered unacceptable. Aside from this study, there has been no other research since which has investigated the impact of *digital* video compression on a face identification task performed by human observers – in particular untrained CCTV users.

CCTV video systems are often poorly set-up. Consequently poor quality images are produced, making the process of identifying unfamiliar faces from surveillance footage extremely difficult. The effects of poor video quality have already created problems for several criminal investigations see [24] for an overview. Firstly, digital CCTV requires high processing, storage and data transmission capabilities which come at a cost. If the cost is too high, owners will compromise video quality to accommodate for costs in running the system. Most digital CCTV systems can be configured to suit the needs and financial resources of system owners, however there is anecdotal evidence suggesting that CCTV owners are tempted to reduce costs by reducing video quality to accommodate low budgets available for CCTV [28]. CCTV owners are often not aware of the consequences of insufficient storage space and bandwidth unfortunately, producing potentially unusable video for key security observation tasks such as monitoring, detection, recognition and identification [4].

Digital video quality can be reduced through three main data reduction techniques: (1) lowering the frame rate; (2) recording at a low resolution and through (3) video compression. For Phase Alternating Line (PAL) video systems, frame rates range from 1-24 frames per second (fps). Rates between 1-5 fps are considered as low frame rates and appear ‘jerky,’ making it difficult for the observer to establish what is happening within the scene. High frame rate video is considered to be around 15 fps and above. Video which is recorded at a low resolution will show less detail than video recorded at a high resolution. Most digital CCTV systems by default record video at Common Image Format (CIF) resolution (352x288). However, many can be set to lower rates i.e. Quarter CIF (QCIF) resolution (176x144). Like frame rates, CCTV video resolution can be altered on the recorder interface when it is recorded or delivered across the network. There are dangers with recording video at such a low resolution, as the video will contain fewer pixels and therefore will hold less detail, making it difficult for an observer to identify faces and number plates.

Video compression exploits the spatial and temporal redundancy of moving images by the use of a software/hardware algorithm (also known as a video CODEC). Excessive video compression can create distortions and artefacts to the video, which becomes increasingly noticeable to the human eye as compression is further increased. Excessive video compression will produce unusable CCTV, which may result in misidentifications. In this paper, we present a face identification study, in which we chose to investigate the impact of excessive video compression of CCTV video on a face identification task carried out by untrained CCTV users.

In the remainder of this section, we provide additional background on CCTV where we review the relevant literature on the use of CCTV and the existing guidelines on digital CCTV video. We also make a review on some previous studies on video quality and face identification. In sect. 2, we describe the methodology for our study and in Sect. 3 the details on the data created for the study are given. The results and discussions are then detailed in Sect. 4. In the final section (Sect. 5), we make substantive and methodological conclusions, and provide recommendations for both CCTV manufacturers and owners on the recommended video compression rates for MPEG-4 and Wavelet digital CCTV systems. We also discuss the practical implications for digital CCTV systems as well as the methodological issues when evaluating video quality for security and surveillance applications.

1.1 Use of CCTV by human observers

CCTV video systems *"… are developed to be used either in real-time to prevent disasters or crimes, and/or to extract knowledge for a-posteriori investigation,"* [11]. Recorded video footage that shows a perpetrator committing a crime provides two types of evidence for trained and untrained CCTV users: (1) scientific forensic evidence and (2) eye-witness testimonies. The use of recorded CCTV has proved to be very successful in solving crime. For example, the suicide bombers involved in the terrorist attacks in central London (July 2005) were all identified from the public transport CCTV systems. Although the terrorists were identified in this case, prosecutors in general are still facing difficulties in establishing the identities of offenders from CCTV footage [6]. Eye-witnesses who are called by the police following an incident are typically untrained CCTV users. Their role in a criminal investigation is to help the police identify a suspect from a CCTV image/s when they usually have witnessed a crime or knows the suspect in question. The process of identification requires the witness to identify *a known person from memory*. This type of task has been evaluated by previous researchers (see Sect. 1.5); however no research has been done to examine how these users are able to identify unknown faces from CCTV images typically produced from a low-cost digital CCTV system.

1.2 Video quality guidelines for digital CCTV systems

There are currently no tested guidelines available for digital CCTV users on video quality and storage, to ensure that they are set up effectively for the intended tasks and CCTV users. The Operational Requirements manual, a CCTV manual published by the UK Home Office [4] provides some guidance on the video quality issues but only for older, analogue systems. The guidelines specify the minimum height a person should appear on a video monitor for five main security observation tasks typically carried out in real-time (monitor, control, detect, identify and recognize), but for tasks performed with recorded or streamed digital CCTV video. These guidelines were prepared when only analogue CCTV systems existed; therefore they do not address the storage and transmission issues associated with digital CCTV video. CCTV users need concrete guidelines on what video quality is acceptable for different CCTV security tasks. For example, how much should digital video be compressed before it becomes unusable for identifying unknown faces?

1.3 Video quality for automated surveillance applications

Most of the empirical research in CCTV has concentrated on building robust computer vision and classification algorithms for video surveillance applications (i.e. [17][18]). There is one particular study that examines the impact of video compression and quality on automated face detection and tracking for a surveillance system used for large-scale distribution purposes [21]. In order to conserve network and power resources, Korshunov proposed that bandwidth can be conservatively reduced up to 29 times for identifying faces and 16 times for tracking faces automatically. There is no corresponding research that has identified how much video data can be saved for face identification tasks performed manually by human observers, particularly untrained CCTV users.

1.4 Video quality for multimedia applications

A number of research studies have been carried out in the past to establish the impact of degraded digital video for applications such as multimedia conferencing (MMC) systems, e-learning, mobile and TV applications. Researchers have predominantly measured the user's perceptions to degraded video using subjective rating scales and questionnaires. The ITU Mean Opinion Score (MOS) subjective test [1] is commonly used by multimedia researchers to establish the threshold limits for video and image degradations. MOS are popular because the method is a standardized subjective test and is quick and easy to use. The problem however, with using MOS for video quality assessments is that, they *"rely merely on subjective ratings rather than on more objective performance in relation to a particular task or application of interest"* [19], therefore the results are not very meaningful. We considered for our study that, for evaluating the usability of CCTV video - task performance measures were suitable as a test method as 1) they take into account the context of the application and the user of the system and 2) the data measured is objective, making it easier to recommend a threshold video quality rate for achieving effective task performance.

1.5 Face recognition research in psychology

Psychology research has demonstrated that both trained and untrained CCTV users are generally poor at identifying unfamiliar faces from poor quality analogue CCTV video and images [7][8][9][16]. These experiments mainly assessed task performance with observers who were *familiar* with faces from CCTV images when high-quality photographs were available for direct comparison. It was thought that high performance for the *familiar* group was due to the sufficient low-spatial frequency information from faces contained within the high quality photographs [8]. Although these experiments demonstrated that low-quality analogue CCTV video significantly affected face recognition performance with unknown faces, the specific video quality parameters were not measured as the aim of these studies were to examine the task performance differences when the observer was familiar and unfamiliar with the faces shown in the CCTV images.

To understand whether digital video compression affects face identification performance and what impact excessive video compression has on the identification of unknown faces, we consider one key question: how much should the user compress digital CCTV video before it starts to affect the user's ability to identify faces from CCTV images? In addition to this, how do untrained go about identifying unknown faces from CCTV images? What cues are used? What makes the task difficult?

2. METHODOLOGY

2.1 Face identification study

We decided to take the human face recognition research further by measuring face identification performance with images taken from video encoded using two common CCTV CODECs (MPEG-4 and Wavelet) [26] at four different encoding bit rate levels: 32, 52, 72, and 92 Kbps. These two CODECs were used as no previous comparisons have been made in the past. Video quality was reduced through compression by altering the encoding bit rate of video, as this was a meaningful measure of CCTV quality for network surveillance users. Furthermore, it was a replicable and accurate method for altering video quality for experimental purposes. We chose 32 Kbps for the lowest quality condition, as this was realistic - many digital video recorders (DVR) offer this encoding rate as the minimum quality (for e.g. Indigo Vision's Video Bridge Networked video recorder allows users to encode from 32 Kbps – 1 Mbps). Other multimedia researchers have used this rate too in experiments. For instance, in a mobile TV video quality acceptability study, 32 Kbps was the lowest rate chosen by Knoche et al. [20] Participants in this study were asked to say whether the video quality was acceptable or not rather than choose a correct response, as the task was passive. For the highest video quality rate, we chose 92 Kbps. Although this is still a very low rate for streaming video, it is a rate very often chosen by CCTV owners to stream video.

To assess task performance, we applied a yes-no procedure taken from the signal detection theory [15]. This theory assumes that the decision-maker does not passively receive information, but actively makes difficult perceptual judgments under conditions of uncertainty. The yes-no paradigm allowed us to gather binary responses from participants to measure the hit rates (the average correctly identified faces) and the false alarm rates (the average false positives: participant said yes it matches when the answer was no). As well as measuring performance, we were also interested in how observers were identifying faces: their strategies and difficulties. This type of data was gathered qualitatively, gathered through a post-experiment questionnaire which was administered after the task.

In this study, we hypothesized the following:

- Hypothesis 1 (H1): As the quality of CCTV images changes as a result of video compression, face identification performance will also change.

- Hypothesis 2 (H2): There will a difference in task performance when faces are identified from MPEG-4 and Wavelet encoded CCTV images.

3. DATA

3.1 Preparation of CCTV images for identification task

The CCTV video stills used in this experiment were prepared by recording 64 video clips of a person walking towards a digital video camera holding a face mask in front of their own face. The masks were prepared by taking photographs of 64 different faces using a 2-mega pixel digital camera (image resolution: 828x1143). The faces were printed to an equivalent size of an average sized human head. The lighting in the space (a squash court) was uniformly distributed to eliminate the lighting variable.

The walking mask method was adopted rather than taking video recordings of 64 different individuals walking towards a video camera purposely to control variables such as target gait, expression, clothing, stature and physique. These variables can potentially affect the way in which the encoder handles the video and affects the observer's judgments, thus they were controlled. Previous research found that the removal of information about the person's gait and physique does not have a huge impact on face recognition performance [9]. Participants were photographed in a full-face pose under controlled lighting conditions. In order to gain a representative set of faces for the CCTV video stills, 16 individuals were photographed belonging to four different ethnic category groups: (1) Afro-Caribbean; (2) Indian-Asian; (3) Oriental and (4) White-Caucasian (equal gender split). The same person was used for the mask walking recordings, wearing plain clothes (black trousers, navy pullover and running shoes). An example of an encoded CCTV still used in the experiment is shown in Fig. 1 with its corresponding look-alike photograph (MPEG-4, 52 Kbps condition).

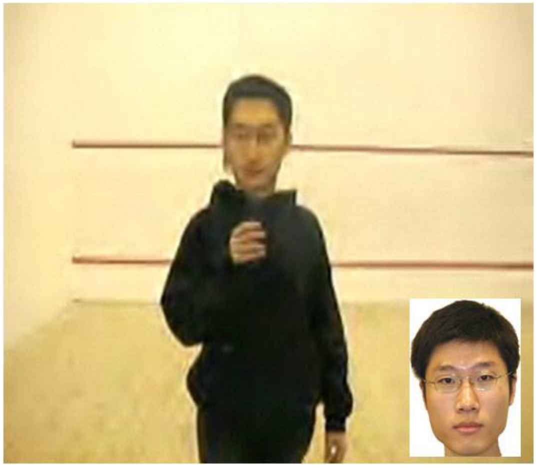

Fig. 1. An example of a degraded CCTV image and the look-alike photograph

Each video clip was reduced in video resolution from 720x576 to 352x288 (CIF resolution). This reduction was made to match the image resolution typically produced by digital CCTV systems. Half of the video clips were encoded using an MPEG-4 and the other half using a Wavelet video CODEC and then reduced into their bit rate conditions. As the recordings were very short in duration (8-seconds), no video frames were discarded from the video clips during the encoding process. One frame from each video clip was selected which showed the target occupying 120% of the video screen. This frame size was chosen to match the recommended target size for identification following the UK Home Office guidelines [4]. The video was shot at a horizontal field of view, showing faces at an average pixel size of 17x27. We evaluated a real-time security observation task typically performed by untrained CCTV users. This type of task involved the observer making a comparison of two images, to identify the individual - a face from the degraded CCTV with the correct or look-alike face from a high-quality photograph.

3.2 Measurement and experimental design

In order to assess task performance, we measured the number of 'hits,' - the total number of times the participant correctly identified a face and the false alarms (false positives) – the total number of times the participant says "yes." The average hits and false alarms were then calculated to give the hit and false alarm rates for data analysis purposes. After the experiment, participants were given a post-experiment questionnaire which asked a number of open-ended questions about the difficulty of the task.

A 2x4 within-subjects experimental design was adopted for this study. In order to avoid practice effects, the video quality conditions were presented in a random order. As there were 64 different faces in the presentation test, the effect of face bias was eliminated by encoding each of the video clips into every possible condition and then counterbalancing the stimuli conditions into eight experiment tests, showing 64 CCTV stills in each of the tests. Counterbalancing was necessary to control for the different faces shown per test. The gender and the ethnicity of the target's faces in the CCTV stills were also counterbalanced; however the participant's gender and ethnicity were not counterbalanced as these factors were not examined in this experiment.

3.3 Participants and procedure

80 paid university participants were recruited from an opportunity sample. There were 47 females and 33 males (mean age 26) consisting of six Afro-Caribbeans, ten Orientals, 19 Indian-Asians and 45 White-Caucasians. Participants who appeared in the CCTV images were not recruited in the experiment. Participants were briefed about the identification experiment and were asked to view each of the 64 CCTV images one-by-one on a 15 inch video monitor. The task required participants to state whether each face matched its corresponding face from a book containing 64 faces from high-quality photographs. For each face, participants were asked to say 'yes,' if they thought the two faces matched, or 'no' if they thought they did not match. The participant was asked to complete the test in 30 minutes and no more. Once the task was complete, participants were asked to say if they found the task easy or difficult to complete. They were then given a post-experiment questionnaire to complete. The first question asked participants to order the ethnic groups they found most difficult and then to explain their choices. Question two asked whether they found male or female faces difficult to identify, and question three asked what face features they used to identify the faces.

4. RESULTS

4.1 Task performance

Fig. 2 shows the average task performance results as CCTV image quality decreased for MPEG-4 and Wavelet images. For MPEG-4 encoded CCTV images, the average number of correctly identified faces decreased by 12, (~18%) as the quality decreased. As Wavelet quality decreased, task performance remained steady from 92 to 52 Kbps (at around 69-71% performance). The average number of correctly identified faces decreased from 52 to 32 Kbps by 5 (~8% decrease in performance). An analysis of variance (ANOVA) with video bit rate as a within subjects factor showed a significant effect on face identification performance between the bit rate levels [F (3, 237) =16.50, $p<0.001$]. However, we found no overall effect of the type of video CODEC on task performance. There was however, a significant interaction effect between video bit rate and video CODEC with task performance, [F (3,237) =5.37, $p<0.001$].

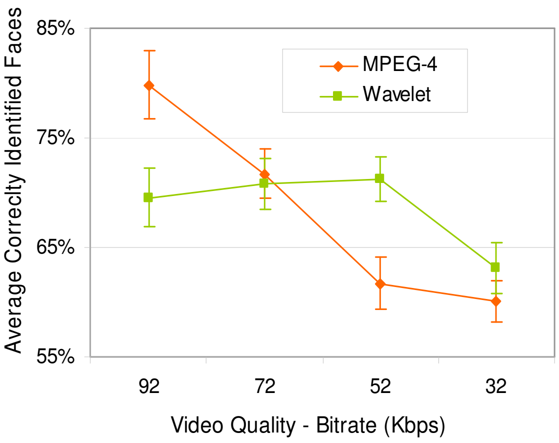

Fig. 2. Identification performance as CCTV video quality increased

To examine the significant ANOVA results further, we ran a pairwise comparison between the video bit rate conditions. Paired t-tests were calculated between video bit rate levels. In order to maintain the overall Type I error rate (α) across all comparisons at 0.05, a Bonferroni-corrected significance level for a two-tailed test of 0.008 was used. For MPEG-4 encoded video, there was no significant decrease in identification performance as video quality decreased from 32 to 52 Kbps. There was, however, a significant increase between the MPEG-4 video quality conditions: 52 and 72 Kbps [t (79) = 3.33, p<0.008] and 72 and 92 Kbps [t (79) = 2.81, p<0.008]. The differences in identification performance between the Wavelet encoded conditions were *not* significant.

The number of times participants said "yes," that is the same face when the actual match was incorrect ("no") was recorded giving the total number of false alarms (type one error). Fig. 3 shows that, as MPEG-4 video quality decreased, the number of false alarms decreased from ~31% to ~13%. Contrary to our expectations, as Wavelet video quality decreased, the number of false alarms raised from ~11% to ~21%.

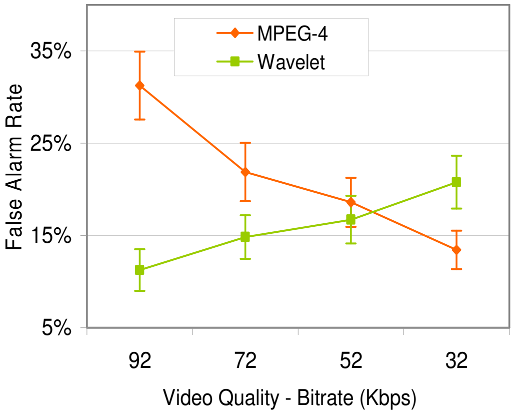

Fig. 3. False alarm rate as CCTV video quality for MPEG-4 and Wavelet CODECs

An ANOVA with video bit rate as a within subjects factor showed that there was a significant difference with false alarms between the bit rate levels [F (3, 237) = 3.97, p<0.001], but there was no significant difference with the false alarms when we compared the results between the two video CODECs. There was however, a significant interaction effect between video bit rate and video CODEC, [F (3,237) = 6.59, p<0.001] (see Fig. 3).

To establish the actual significance of the changes with false alarms, we ran paired t-tests between the video bit rate conditions and video CODECs. Our analysis showed that as MPEG-4 video quality decreased from 52 to 32 Kbps, the false alarms significantly increased [t (79) = 3.41, p<0.008]. There was, however no significance change in the false alarms as video quality decreased from 92 to 52 Kbps (MPEG-4), and no significant change in the false alarms as Wavelet quality decreased.

4.2 Discrimination of faces conditions

d' prime is a statistical measure used in signal detection theory which is a measure of the difference between the hit and false alarm rate. In a single value, it signifies how well the observer discriminates between the signal and noise stimuli. For this study, d' values were calculated to illustrate how well participants were able to discriminate between a signal (correct target) and noise stimuli (look-alike face). The larger the d' value, the observer is better at discriminating between the correct face and the look-alike face.

The d' values in Fig. show that, as MPEG-4 video quality decreased, the level of sensitivity in identifying the actual face with a look-alike face also decreased. For Wavelet images, the sensitivity between actual and look-alike faces remained more or less the same across the video quality conditions. An ANOVA showed that there was no significant effect of d' between video CODECs, as video quality decreased. There was however, a significant difference in d' as MPEG-4 video quality decreased, [F (3,237) = 7.54, p<0.001] and no significant change in d' was found for the Wavelet conditions.

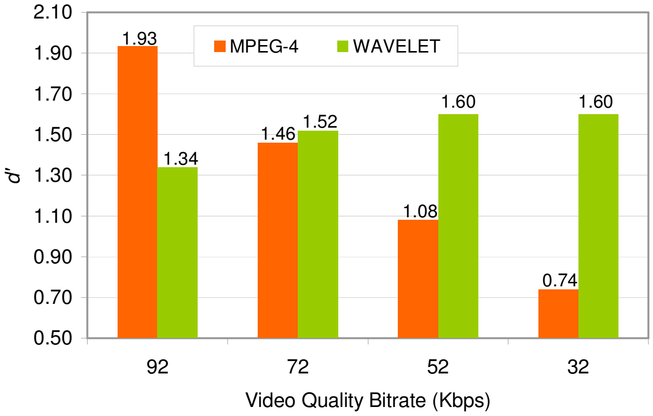

Fig. 4. *d'* prime as video quality increased for MPEG-4 and Wavelet CODEC

4.3 Post-study questionnaire results

After the test was complete, participants were asked questions about the task difficulty. They were asked whether they found the task easy or difficult. 94% of the participants rated the task difficult, and the other 6% said it was easy. Some participants commented in a separate section of the questionnaire about the task in terms of the level of difficulty:

"Made me aware how easy it is to be wrong when trying to identify someone from a CCTV. There is a lot of scope for error."

"I wasn't very confident about any of my decisions. It was a really difficult task for me."

"In the vast majority of pictures, if I was a police officer I would definitely hesitate to say for certain that I had correctly identified or even clearing somebody."

"The resolution of CCTV images should definitely be improved to catch the guilty."

4.4 Preference of Target Ethnicity

In this experiment, participants were recruited from an opportunity sample, thus we did not control for participant gender. 46 female and 34 male participants took part in the study which consisted of six Afro-Caribbeans, ten Orientals, 19 Indian-Asians and 45 White-Caucasians. Once the task was complete, participants were asked to complete the post-experiment questionnaire. Question 1 asked participants to rate the difficulty in identifying the four ethnic groups using a scale starting from easy (1) to the most difficult (4). The average ratings per target ethnicity group across participants are given in Table 1. Participants rated White-Caucasian faces as the easiest (38%), then the Orientals (30%) - giving a rating of 1. Afro-Caribbean targets were rated as the most difficult (66%) – giving a rating of 4. 49% rated Indian-Asians faces as the second most difficult - giving a rating of 3.

Table 1. A summary of the difficulty ratings for the different ethnicity of target faces.

Rating	Ethnicity of targets in CCTV			
	Afro-C	Oriental	Indian-A	White-C
1	5%	30%	8%	38%
2	4%	25%	17%	34%
3	6%	21%	49%	4%
4	66%	4%	6%	4%

A Friedman test illustrated that the rankings for target ethnicity were significantly different from each other [Fr (3) = 31.41, p<0.001]. We were interested in understanding a little further, *why* observers found it easier to identify faces belonging to a particular ethnic group, and so the second part of this question asked participants to explain their ratings. Not all participants had commented on the question, but those who did shared two common views:

1. Afro-Caribbean and Indian-Asian faces were very difficult to identify:

“In the blurred images, the Afro-Caribbean - their facial characteristics were difficult to distinguish.”

“Darker faces were impossible as no light reflected off their faces.”

“Afro-Caribbean faces often appeared as silhouettes with no obvious features; this was also true in some cases with the Indian-Asian faces.”

“The eyes and noses of the darker faces were all the same to me, those were the features I use anyway to identify faces.”

2. Oriental faces were easy to identify:

“The features of the Oriental faces seemed to be more distinct, especially the nose and length between eyes.”

“The oriental eyes seem to be more expressive, making it easier to read.”

“I probably found Oriental faces easier to spot because I come from an Oriental community!”

“I found I could recognize them better as they’re skin tone was just right to help me match the faces with the look-alikes – much better than the Whites Caucasians and Afro-Caribbeans.”

4.5 Performance on task: target ethnicity

Fig. shows the levels of sensitivity in the responses of participants identifying targets across the various ethnic groups. The d’ values in Fig. shows that participants were better at discriminating between the correct and look-alike faces in the following order of target ethnicity: Orientals, Indian-Asians, White-Caucasians and Afro-Caribbeans.

An ANOVA, with target ethnicity as a within subjects factor showed a significant difference in the d' values: [F (3, 237) = 31.46, p<0.001].

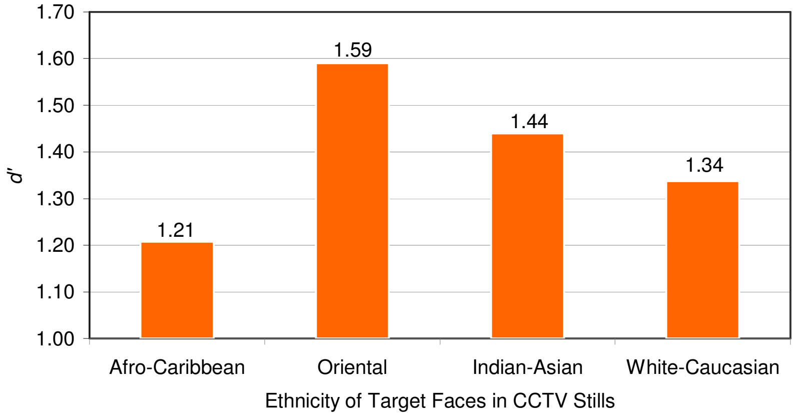

Fig. 5. *d'* prime: Ethnicity of Targets

To establish whether d' was significantly different between the target ethnic groups, six paired t-tests between target ethnic groups (using a 0.008 Bonferroni-corrected significance level for a two-tailed test). The paired tests showed that, d' was significantly different between Orientals and Afro-Caribbeans [t (79) = 8.60, p<0.001], Indian-Asians and Afro-Caribbeans [t (79) = 8.00, p<0.001] and Oriental and White-Caucasians [t (79) = 2.87, p<0.001]. There was no significant difference in the d' values between Orientals and Indian-Asians, and between Indian Asians and Orientals.

No analysis was carried out on the hit and false alarm rates with participant ethnicity. An equal mix of ethnicities were not recruited for the study as the impact of lowering video quality on a face identification task was the aim of the experiment, rather than the identification performance between participant ethnic groups.

4.6 Preference of target gender

In question 2, we asked participants which gender group they found the easiest to identify. 44% found females easy to identify, 31% found males easy to identify and 25% said neither. A 2x2 Chi-Square was conducted to determine whether there was a difference in preference. The Chi-Square revealed a significant relationship between target gender preference and participant gender [X^2(2, N=80) = 5.99, p< 0.05]. Some of the reasons given for their choice of gender included:

"Females have many differences compared to males."

"Such as changing hairstyles and make up introduces doubts in identification?"

"Female hairstyles and the shapes of their faces help me identify them better I reckon."

"It was sometimes possible to spot men based on the presence of a moustache, parting and length of their hair."

"Males were easier, if you went on the basis of hair."

"Men have short hair and it's basically easy to spot the differences between the fake faces and real targets."

Although the participant gender distribution was unequal (female n = 46 and male = 34), females on average preformed marginally better than males (40.78% vs. 40.29%). This result was, however, not significant as a Mann-Whitney U test showed that there was no significant difference in task performance between male and female participants. Performance between participant gender groups for female and male targets was more or less equal:

- Female participants correctly identified: 22.3% male faces and 21.7% female faces

- Male participants correctly identified 22.3% male faces and 21.47% female faces

4.6 Cues used for identification

The final question in the post-experiment questionnaire asked participants what type of face features: internal (nose, eyes, mouth etc.) and external (hair, facial hair, accessories) were being used to identify faces from the CCTV stills. In total, 12 different internal and external face features were used. These features in order of the highest frequency included: hair, nose, eyes, mouth, lips, eyebrows, face shape, facial hair, jaw line, hair line, glasses and jewelry.

4.7 Results summary

Our results partially supported hypothesis 1 ("As the quality of CCTV images changes as a result of video compression, face identification performance will also change."). We found that, as MPEG-4 CCTV video quality decreased, the number of correct detections decreased significantly by 12 (~ 18% performance) from 92 to 52 Kbps. There was, however, no significant change in task performance or false alarms as Wavelet video quality decreased. Although our results showed that Wavelet compression did not have a significant impact on task performance, we feel that further tests are required, particularly as Wavelet CCTV systems are becoming increasingly popular and are non-standard CODECs which come in many forms unlike MPEG-4.

Hypothesis 2 predicted that there would be a difference in identification performance when faces are identified from MPEG-4 and Wavelet encoded CCTV video stills. We chose to test this hypothesis objectively based on a subjective opinion formed from CCTV video and imaging experts [27] who believed that, *"…wavelet video encoded at the same bit rate as MPEG-4 video is perceived as better quality."* Our results showed that performance was in fact better with MPEG-4 encoded CCTV images than Wavelet, thus H2 was accepted as a change a difference was found. These results, particularly in relation to the wavelet conditions, imply that CCTV practitioners and owners should be cautious when configuring Wavelet systems at very low quality levels.

In addition to video quality, other factors affected face identification performance. Consistent with previous research [25], participants preferred identifying faces that belonged to the same ethnic group as their own. This finding is known as the 'other-race effect,' whereby people are better at recognizing faces of their own ethnicity as they are more familiar with them through every-day social interactions [12]. Afro-Caribbean faces were the most difficult to identify (based on subjective response and task performance). There were however, some inconsistencies between performance and the subjective responses for target ethnicity difficult ethnicity. For instance, White-Caucasians were reported as the easiest to identify, yet performance was the highest for Orientals. This could have been due to participants being over-confident when identifying lighter skinned faces from CCTV images perhaps because darker skinned faces were so difficult to identify. Consequently, this over-confidence led to participants committing a larger number of false alarm errors when identifying White-Caucasians from degraded CCTV images. The UK population is becoming increasingly ethnically diverse. The findings with regard to task performance and target ethnicity are important when considering the practical performance issues with CCTV systems. These factors need to be further assessed if the CCTV system will be used by untrained CCTV users and when used for tasks such as the one described in [2]. This is also a fundamental issue for those systems operating in dense cities where the population is diverse.

A higher number of female participants (n = 46) took part in the experiment than male participants (n = 34). In our study we found that across participants, females were considered easier to identify. This finding was consistent with previous research [22]. Despite this, we found no significant difference in task performance between male and female participants. Also, there was no significant difference in task performance when participants identified male and female targets from the CCTV stills. The reason for comparing performance between gender groups was so that we could better understand the effect of face and observer gender bias. If there was a strong effect in performance for faces from a particular gender group, our results could have been used to better inform CCTV owners on the practical issues concerning face identification from poor quality CCTV – particularly when untrained CCTV users are being employed.

5. CONCLUSIONS

5.2 Substantive conclusions

We examined in this study how well untrained CCTV users could identify faces from degraded digital CCTV video stills. Our main finding was face identification performance deteriorated as MPEG-4 video quality was lowered. From the post-study questionnaire, the qualitative responses revealed that as CCTV quality reduced, the internal face features was less clear for identification particularly when with dark skinned faces. Consequently, participants used external face features to help them identify the person. Based on these responses, we believe that the internal face features are poorly represented from images taken from excessively compressed CCTV video (i.e. when compressed to as low as 32 Kbps), and more so when video is encoded with a Wavelet CODEC.

The two most popular external face features used for identification were hair and jewelry, which is perhaps likely as these features were the only visual cues available from severely compressed CCTV video images. Hair and jewelry are features that are easily changeable and can be used as accessories to intentionally disguise a person's physical identity. In the past, internal face features have always been difficult for witnesses to translate when identifying targets [10][13][14]. Physical disguises can also be used to disguise the identity of a person (e.g. wigs, glasses and head gear) when attempting to commit crime such as shoplifting, robbing a bank, committing fraud etc. If these props are heavily used for the identification of people, and the face image has been highly distorted due to excessive CCTV compression, the task will be error prone, reducing the strength of investigations and criminal prosecution cases.

This study explores the practical implications of using digital CCTV video that has been excessively compressed. One instance where an excessive reduction in video quality would pose a problem for security would be for remote video security door entry systems. For such systems, CCTV video can be viewed at a remote location, and displayed on a video screen for person verification and entry into a building premise. If an attacker is viewed on the system by the user (typically an untrained CCTV user) and imitates certain features of the usual delivery person using a disguise, the person viewing the screen may mistake them for the known and trusted delivery person. The quality of CCTV data is the single most important factor in achieving high performance and low false alarm errors for security tasks. The higher the quality of the CCTV video, the better the observer's chance will be in correctly identifying an unknown individual from CCTV. Digital video for security applications is becoming increasingly pervasive and are increasingly being used by untrained CCTV users.

5.3 Methodological conclusions

A number of variables were deliberately controlled to create the CCTV images in this study so that performance could be measured objectively and accurately. Although the walking mask method does not hold extremely high ecological validity, we considered this method as a good compromise, given the difficulties with controlling variables such as the gait, pose, expression, stature and physique. These variables hugely vary from one person to another and have the potential to influence the observer's perceptions, thus affecting their task performance and confidence. To completely validate the walking method, further research is needed to compare whether the walking mask method would give similar results to images produced from real-time CCTV video.

Gaining access to ground truth CCTV can be very difficult, particularly if high resolution uncompressed video is required for experiments where video quality is under study. To ensure that the data holds high ecological validity, targets recorded in video scenes must realistically represent real world CCTV. Rather than selecting an opportunity sample typically consisting of White-Caucasian middle aged males – the participant sample for CCTV targets should contain an equal number of male and female targets from a wide age and ethnicity range. For those researchers who wish to assess the effectiveness of CCTV video quality, when used for security observation tasks, it is recommended that the data video and images used for the test are high in quality and the content is representative of real world CCTV imagery. Also, a reliable video CODEC should be used for altering video quality, something that can be easily manipulated and replicated by other researchers. Variables such as environmental lighting, recording height and angle should also be controlled as best as possible, without factoring in motion blur which is a very common issue when recording video using a digital camera recorder.

5.4 Recommendations for CCTV owners

We chose to evaluate the impact of excessive video for the new and emerging scenario in which CCTV is used by everyday users (untrained) for identifying unknown targets from CCTV images [2]. We also considered this user group, as getting access to a large sample of trained CCTV users (i.e. CCTV operators, police and forensic staff) was impossible for this research study - therefore untrained CCTV users took part.. This was considered beneficial, as our study results can be applied for systems used by both trained and untrained CCTV users. We therefore provide CCTV security owners and practitioners with the bare minimum for encoding digital CCTV video for data streaming and recording:

- MPEG-4 systems should be configured to record CCTV video to an equivalent video quality level of 52 Kbps (minimum) for effective face identification and above.
- Although our results showed no significant improvement for Wavelet CCTV video on task performance, to avoid errors in identifications - Wavelet CCTV systems should be configured *cautiously* by CCTV owners. For Wavelet CCTV systems, video quality should be configured to 92 Kbps or more.
- Where darker-skinned targets are likely to be captured on surveillance video, we recommend that sufficient lighting is given to the CCTV camera capture area, so that the observer can capture as much facial detail as possible from the CCTV image - rather than relying on cues that can be easily disguised such as hair, jewelry and other accessories.

CCTV security owners, practitioners and designers must be aware that digital CCTV systems differ from one another in terms of functionality. For example, digital and network video recorders may encode video using different video CODECs, CODEC versions, resolutions and frame rates. Different CCTV systems also operate in different environments, and can be used for many purposes, not solely for the identification of people. For each particular application, the owner of the system should explicitly define the security goals and user tasks, and – until guidelines based on scientific data become available – test if their chosen configuration is fit for purpose. For now, the recommendations on video quality should be followed to consider where the biggest gains in performance are, and weigh the cost of higher quality against the cost of missed hits and false alarms. At the very minimum, system owners should test whether the chosen level of video compression results in video quality that results in sufficient performance for their specific tasks with the intended CCTV users.

Our future research will involve researching the limits of temporal video quality with a different type of CCTV task: the detection of events from CCTV video. The final research aims to provide digital CCTV video guidance to ensure that systems are producing usable and reliable data for real-time crime detection and post-even investigation tasks – performed under a range of contexts and by users of varying skills and experience.

ACKNOWLEDGEMENTS

This research was funded by an EPSRC studentship (grant no: EP1P50064811).

We would like to thank Simon Walker, Neil Cohen, Ken Brown MacLennan, Jay Gattuso (UK Home Office Scientific Development Branch (HOSDB) and Jim Aldrige (Effective Pictures) for their help, support and providing equipment to process the CCTV video material for the experiment

We would also like to thank Rae Harbird for volunteering to be filmed, all those photographed and the participants who took part in the face identification experimental study.

REFERENCES

[1] ITU-R Recommendation BT.500-1, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union - Geneva: Switzerland (2002).

[2] BBC News, "Rights group criticises ‘Asbo TV’," http://news.bbc.co.uk/1/hi/england/london/4597990.stm (2006).

[3] BBC News, "Web users to 'patrol' US border," http://news.bbc.co.uk/1/hi/world/americas/5040372.stm (2006).

[4] Aldridge, J. "CCTV Operational Requirements Manual," Police Scientific Development Branch (PSDB), version 3 (1994).

[5] Bachmann, T. "Identification of spatially quantised tachistoscopic images of faces: how many pixels does it take to carry identity?" European Journal of Cognitive Psychology, 3, 87-103 (1999).

[6] Bromby, M. "To Be Taken At Face Value? Computerised Identification," Information Communication Technology Law Journal, 11(1), 63-73 (2002).

[7] Bruce, V., Henderson, Z., Greenwood, K., Hanwood, P., Burton, A. M., and Miller, P. "Verification of face identities from images captured on video," Journal of Experimental Psychology: Applied, 5, 339-360 (2001).

[8] Bruce, V., Henderson, Z., and Burton, A. "Factors affecting accuracy of verifying identities from CCTV images," Journal of Experimental Psychology: Applied, 7, 201-208 (2001).

[9] Burton, A. M., Wilson, S., Cowan, M., and Bruce, V. "Face recognition in poor quality video: evidence from security surveillance, " Psychological Science, 10, 243-248 (1999).

[10] Cambell, R., Coleman, M., Walker, J., Benson, P. J., Wallace, S., Michelotti, J., and Baron-Cohen, S. "When does the inner-face advantage in familiar face recognition arise and why?" Cognition, 6, 209-218 (1997).

[11] Cucchiara, R. "Multimedia surveillance systems," Proc. of Third ACM International Workshop on Video Surveillance & Amp; Sensor Networks (Hilton, Singapore, ACM Press, New York, NY, 3-10 (2005).

[12] Elliot, E. S., Wills, E. J., and Goldstein, A. G. "The effects of discrimination training on the recognition of white and oriental faces," Bulletin of the Psychonomic Society, 2, 71-73 (1973).

[13] Ellis, H., Shepherd, J. W., and Davies, G. M. "Identification of familiar faces and unfamiliar faces from internal and external features: some implications for theories of face recognition, " Perception, 8, 431-439 (1979).

[14] Gibling, F. and Bennett, P. "Artistic enhancement in the production of photofit likeness and examination of its effectiveness in leading to suspect identification," Psychology Crime and Law, 1, 93-100 (1994).

[15] Green, D. and Swets, J. [Signal detection: theory and psychophysics], Wiley, New York (1966).

[16] Henderson, Z., Bruce, V., and Burton, A. M. "Matching the faces of robbers captured on video," Applied Cognitive Psychology, 15, 445- 464 (2001).

[17] Javed, O. and Shah, M. "Tracking and object classification for automated surveillance," Proc. 7th European Conference on Computer Vision, 343-357 (2002).

[18] Ketnakker, V. and Zabih, R. "Bayesian multi-camera surveillance," Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 253-259 (1999).

[19] Knoche, H. and de Meer, H. "Utility Curves: Mean Opinion Scores considered biased," Proc. of IWQoS, 12-14 (1999).

[20] Knoche, H., McCarthy, J. D., and Sasse, M. A. "Can small be beautiful?: assessing image resolution requirements for mobile TV," Proc. 13th Annual ACM International Conference on Multimedia, ACM Press, New York, 829-838 (2005).

[21] Korshunov, P. and Ooi, W. T. "Critical video quality for distributed automated video surveillance," Proc. of 13th Annual ACM International Conference on Multimedia, ACM Press, New York, 151-160 (2005).

[22] Lewin, C. and Herlitz, A. "Sex differences in face recognition- women's faces make the difference," Brain & Cognition, 50, 121-128 (2002).

[23] Malpass, R. S. and Kravitz, J. "Recognition for faces of own and other race faces," Journal of Personality and Social Psychology, 13, 330-334 (1969).

[24] Mead L. The Changing Jurisdiction: usage of video recordings in surveillance, the value of such as evidence and potential problems which can arise. 13th Annual BILETA Conference, Trinity College, Dublin, 1998.

[25] Rehnman, J. and Herlitz, A. "Higher face recognition ability in girls - Magnified by own-sex and own ethnicity bias," Memory, 14, 289-296 (2006).

[26] Robertson, I. L. and Monro, D. M. "Video Surveillance using low bandwidth high compression systems," Proc. European Conference on Security and Detection, 31-35 (1997).

[27] S. Walker and N. Cohen (private communication).

[28] N. Cohen (private communication).