A case study in identifying acceptable bitrates for human face recognition tasks

Anastasia Tsifouti [1, 2]
Sophie Triantaphillidou [2]
Mohamed-Chaker Larabi[3]
Efthimia Bilissi[2]
Aleka Psarrou[2]

[1]Home Office Centre for Applied Science and Technology, Sandridge, UK
[2]Faculty of Science & Technology, University of Westminster, UK
[3]University of Poitiers, France

A case study in identifying acceptable bitrates for human face recognition tasks

A. Tsifouti[a,b], S. Triantaphillidou[b], M.-C. Larabi[c], E. Bilissi[b], A Psarrou[b]

[a] Home Office Centre for Applied Science and Technology, Sandridge, UK; [b] University of Westminster, London, UK; [c] University of Poitiers, France.

Face recognition from images or video footage requires a certain level of recorded image quality. This paper derives acceptable bitrates (relating to levels of compression and consequently quality) of footage with human faces, using an industry implementation of the standard H.264/MPEG-4 AVC and the Closed-Circuit Television (CCTV) recording systems on London buses. The London buses application is utilized as a case study for setting up a methodology and implementing suitable data analysis for face recognition from recorded footage, which has been degraded by compression. The majority of CCTV recorders on buses use a proprietary format based on the H.264/MPEG-4 AVC video coding standard, exploiting both spatial and temporal redundancy. Low bitrates are favored in the CCTV industry for saving storage and transmission bandwidth, but they compromise the *image usefulness* of the recorded imagery. In this context, usefulness is determined by the presence of enough facial information remaining in the compressed image to allow a specialist to identify a person. The investigation includes four steps: 1) Development of a video dataset representative of typical CCTV bus scenarios. 2) Selection and grouping of video scenes based on local (facial) and global (entire scene) content properties. 3) Psychophysical investigations to identify the key scenes, which are most affected by compression, using an industry implementation of H.264/MPEG-4 AVC. 4) Testing of CCTV recording systems on buses with the key scenes and further psychophysical investigations. The results showed a dependency upon scene content properties. Very dark scenes and scenes with high levels of spatial-temporal busyness were the most challenging to compress, requiring higher bitrates to maintain useful information.

***Index Terms*—CCTV recording systems, human face recognition, H.264/MPEG-4 AVC, image usefulness, visual psychophysics, scene characterization.**

1 Introduction

The police use both subjective and objective methods for the completion of face recognition tasks. Objective methods employ automated face recognition systems (FR). Subjective methods involve visual examinations of recorded/transmitted imagery, carried out by operatives, such as police staff and external specialists. Police staff carries out recognition tasks and the court decides whether, or not to convict, based on the evidence provided. So, Closed – Circuit Television (CCTV) imagery is used in UK courts as documentary evidence [1]. Furthermore, CCTV imagery has been found to have an effective impact on conviction of crimes [2].

There are many surveillance applications where the recording of facial information is used, such as in trains, buses, underground, transport stations and open streets. When a person is further away from a camera than less facial information will be visible and other means of person recognition can be applied, such as gait analysis [3]. In this investigation, the London buses application was selected as a case study. London public buses make use of CCTV systems to prevent crime, identify offenders/actions and for insurance purposes [4, 5].

Factors, such as illumination conditions, angle of the face with respect to the camera plane, and camera to subject distance influence the accuracy of face recognition tasks [6-12]. These factors affect the image quality of the reproduced imagery thus the reproducibility of useful facial information. Illumination poses an important problem in CCTV imagery, since such systems often operate under totally uncontrollable, or semi-controllable illumination conditions (e.g. open street CCTV cameras, bus CCTV systems). Useful facial information in the reproduced imagery is further compromised by compression, implemented to satisfy limited storage capacity of CCTV recorded systems, or transmission bandwidths. Low bitrates are favoured in the CCTV industry for lowering data costs.

Subjective image quality for security systems has been defined as the image usefulness, or image suitability of the visual material to satisfy a specific task [13-15]. In this context, the specific task requires enough useful facial information to remain in the compressed image in order to allow a

specialist to recognize a person from the video footage. Quality of experience (QoE) related to such applications goes beyond the standard quality factors [16]. Therefore, image usefulness does not have the exact same meaning as image fidelity [17]. For example, an image with visible compression artefacts is considered distorted, but if the artefacts do not hide, or alter any facial information the image maintains its usefulness. A parallel of this concept is fingerprint identification. It has been shown that visible compression artefacts do not decrease the usefulness of the compressed fingerprint image compared to uncompressed original, as long as the artefacts have not affected important fingerprint ridges that are used in recognition [18]. The main users of security footage (police staff) are those who examine in detail the relevant information within the footage. So, image usefulness for CCTV footage is based on the visibility of information that could lead to recognition (facial, garments, gait).

Lossy compression is used in the security industry to enable efficient storage and video transmission [19-24]. Lossy compression is a distorting process that can eventually have an effect on the visible information in video [25, 26] and thus the usefulness of a facial image which is used in a recognition task. The performance of lossy compression is influenced by a) the content of the scene [27-31] and b) the compression algorithm, its properties and settings [29, 31, 32]. Scenes with different motion properties (temporal differences), regions (spatial differences) and combinations of different spatial-temporal properties will require different acceptable bitrates, leading to different levels of compression. Figure 1, provides an example of the H.264/MPEG-4 AVC lossy encoder performance under different illumination conditions.

Fig. 1. As the bitrate decreases the useful information decreases. Defining acceptable compression ratios depends on the original image and observers' acceptability standards. For instance, the bright (at the bottom) and dark (at the top) scenes are more susceptible to compression. They require lighter compression to achieve "acceptable" responses from observers, in comparison to the well-illuminated scene (in the middle).

The majority of the CCTV recorders on London buses use a proprietary format based on the H.264/MPEG-4 AVC encoder. The latter is a hybrid video encoder, exploiting both spatial and temporal redundancy and uses a 4×4 integer adaptive transform, which is an approximation of the 4×4 DCT. H.264/MPEG-4 AVC produces blocking artefacts that become more visible at low bitrates [33, 34]. The ITU-T and ISO/IEC JTC 1 video standards, such as H.264/MPEG-4 AVC, specify only the decoding part in order to ensure interoperability and syntax capability between different technologies implementing the standard. Image quality is not specified in the standards. As a result, these different implementations of the same encoder will produce different compressed qualities.

Image usefulness, is under the category of visuo-cognitive attributes of image quality and is only evaluated using subjective investigations and psychophysics [32, 35]. These types of investigations are complicated and require detailed planning, because multiple variables can affect the observer's impression and the QoE [36]. The ITU provides guidance in relation to assessments of video based imagery [37]. When image fidelity (i.e. the visibility of image distortion from reference imagery) is assessed, the ITU recommends that the reference video is provided and runs simultaneously on a single monitor, along with the reduced quality version. This arrangement helps the observers to make a direct judgment of what they see, and does not rely on memory. A similar methodology to the fidelity assessment is used for the assessment of image usefulness in the present investigation.

The main resource of information on the subject of image quality for CCTV systems comes from the extensive work of Klima and his co-authors. Although, Klima et al [27, 28, 38] have tested many different compression techniques using subjective testing, they have not used an extensive set of scenes, and have not included different scene properties (e.g. difference in illumination or distances of the subject to the camera). Their work is related mainly to a few close up faces and number plates. Additionally, they concluded that the subjective results were not dependent just on compression rate, but on the initial information content of the scene and its purpose [28]. For this reason, in the present investigation a more extensive set of scenes, with different properties, is employed.

In this work we propose a reproducible methodology and tools to derive acceptable bitrates of scenes with human faces using an industry standard implementation of H.264/MPEG-4 AVC, and the CCTV recording systems on London buses. Scenes of 20-second duration were grouped based on the following inherent properties: scene brightness, camera to subject distance, angle of the face to the camera plane and level of busyness (based on spatial and temporal information). The majority of the CCTV recorders on London buses use proprietary formats based on the H.264/MPEG-4 AVC video coding standard [4, 39]. Two psychophysical investigations were conducted with the help of experts from the Metropolitan Police Service (MPS) and bus analysts. The first was used to identify the *key scenes* (i.e. scenes affected most by compression), from an initial selected set of 25 scenes, using an implementation of H.264/MPEG-4 AVC. The second was used to identify acceptable bitrates of the pre-selected key scenes (resulted in a set of 6 scenes), using five of the most commonly used CCTV recording systems on London buses. The former psychophysical investigation acted as a filter in order to reduce the viewing experimental time in the latter psychophysical investigation. In both investigations, the expert observers had to answer with a *yes* or *no* to the question "Is the compressed version of the scene as useful as the reference original in terms of facial information?".

Findings are aimed to contribute in optimizing the conditions around facial identification tasks undertaken by specialists, by tuning the compression to a just acceptable level. The developed video bus dataset can be obtained from the Home Office Centre for Applied Science and Technology in UK [40], to assist those wishing to investigate solutions in relation to the bus video recordings (e.g. automated detection of actions, such as pick pocketing). Scene content characterization allows the use of the video dataset to be valid for other CCTV applications. Furthermore, such characterization provides identification of scene properties with which systems (in this case compression algorithms) have been assessed and thus their influence on the systems' performance. For example, low brightness scenes are more affected by compression than medium brightness scenes, and this is valid not solely for CCTV but for any human face recognition application.

This paper is an extension of our initial investigation on the subject [26]. It contains further work on scene grouping, expanded analysis of results and, further recommendations and conclusions. More specifically, it includes an expanded introductory section, combining information from five different subject areas (i.e. human face recognition, police tasks/procedures, compression, image/video quality, and visual psychophysics). Further work has been put in the characterization and grouping of video scenes, which was based on the quantification and grouping of individual scene properties. The results are analyzed using a systematic approach that resulted to new findings. Finally a comparison is made between the results from the industry standard coding H.264//MPEG-4 AVC and from the CCTV recording systems.

The rest of the paper is organized as follows: Section 2 contains the description of the experimental methodology. Data analysis of the results and discussion, of the two psychophysical investigations, is described in Sections 3, 4 and 5. In Section 6 conclusions are drawn along with suggestions for future work.

2 Methodology

Four main steps were carried out in order to derive the acceptable bitrates: 1) Development of a representative video dataset, 2) Selection, characterization and grouping of video scenes. 3) Identification of key scenes using an industry standard implementation of H.264/MPEG-4 AVC. 4) Testing of five CCTV recording systems using the identified key scenes.

2.1 Development of a representative video dataset

A sunny day presents challenges in terms of illumination for recording activities on buses. When the sun illuminates the one side of the bus, some areas in a scene are over-exposed, while others under-exposed. As the bus moves, the windows allow illumination from different directions, causing the areas of over- and under- exposure to vary rapidly. On the contrary, an overcast day will produce diffuse light and uniform illumination, which is not challenging enough for testing compression. When it is dark, bus illumination will dominate the scenes and will therefore produce a predictable and uniform illumination. The following conditions were used during data collection (footage recording).

- *Camera system*. A consumer quality mini digital video (DV) camcorder was used for the filming of all scenes. An automated exposure setting was chosen to replicate what happens with CCTV camera capture. Ten camcorders were set up according to Transport for London (TfL) recommendations, i.e. the camera views.
- *Illumination conditions*. Sunny day (during day time) and bus illumination together with some exterior illumination e.g. from shops (during night time).
- *Participants.* Twenty-six actors from various ethnicities, ages and gender acted as the bus passengers, according to given scenarios.

The DV camcorder was chosen for the recording of the bus dataset over a CCTV camera for various reasons, including accessibility, quality and cost. For example, expensive, specialised equipment is required in order to record the output of a CCTV camera in an uncompressed format. Also, there are numerous companies that provide CCTV systems to London buses, these have large variations in quality, which have not been studied and quantified. Figure 2 provides a comparison of the Spatial Frequency Response (SFR) [41], of a typical sample CCTV camera used on buses with that of the DV camcorder used for the collection of the dataset.

The DV camcorder SFR indicates image sharpening in the vertical camera orientation, in the low and mid frequencies. Further, the camcorder has a much greater optical resolution (i.e. the SFR falls to 0.1 at nearly 4 pixel^{-1}) and produces sharper images (i.e. 0.5 SFR corresponds to approximately 2.7 pixel^{-1}) than the CCTV (i.e. optical resolution limit at less than 3 pixel^{-1} and 0.5 SFR at less than 2 pixel^{-1}).

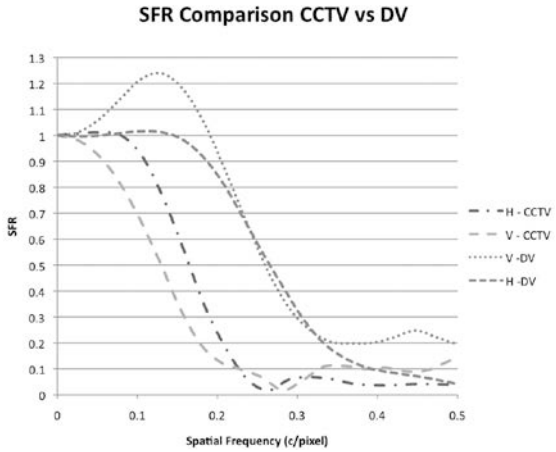


Fig. 2 Horizontal (H) and Vertical (V) Spatial Frequency Responses of a CCTV camera and a DV camcorder.

The consumer DV camcorder is shown to have produced overall higher quality output than the CCTV system. One option to compensate for this difference is to apply a frequency filter, aiming to visually match the frequency response of the DV recorder to that of the CCTV camera [42]. In this case, this option was omitted, since the current rapid development of CCTV system technology will result in CCTV systems producing comparable image quality to that of consumer video systems. The focus of the work was put on setting up an experimental paradigm and implementing a suitable analysis of results.

The footage dataset was recorded in a DV format, at 25 megabits per second (Mbits/s), 4:2:0 chroma subsampling, at full D1 PAL resolution (720 × 576) and with interlaced scanning at 25 frames per second (fps).

2.2 Selection, Characterization and Grouping of Video Scenes

In this investigation, various scenes were selected from the bus dataset and were further compressed using MPEG-2 coding standard at approximately 25Mbits/s (4:2:0 chroma subsampling). This compression has enabled the five suppliers of bus CCTV recording systems to have the key scenes on a DVD. The suppliers were asked to play out (with a DVD player) the key scenes into their recording system according to a pre-defined number of bitrates and to return their recordings for use in the experimental testing.

The main difference between DV and MPEG-2 compression is in the temporal domain; otherwise both encoders are based on the DCT transform [43] (i.e. MPEG-2 exploits both spatial and temporal redundancies whereas DV exploits only spatial redundancy). An initial experiment was conducted to appreciate empirically the visible differences between the two encoders. The experiment involved careful observation of a number of compressed scenes, with various scene properties. No visible differences were observed between the compressed scenes. Figure 3 illustrates an example comparison using both encoders. The compression bitrates used in the CCTV industry are typically much lower than 25 Mbits/s. Thus, the additional compression of the reference using the MPEG-2 encoder should not affect the results.

Fig. 3. Comparison of two images compressed at 700kbps, with MPEG-2 (right) and DV (left).

Due to the miscommunication between transmission and recording, it was observed that CCTV recording systems sometimes recorded the two fields as one frame causing the interlace effect (see Figure 4). In order to avoid this effect in the compressed scenes, one of the fields (the odd line numbers) was removed. Thus, the selected original reference for this present investigation consists of 25 fields per second (not 25 frames per second) and is compressed with MPEG-2.

Since compression performance is dependent on scene content, the various scenes selected from the bus dataset were characterized and grouped based on local and global scene properties. In total, 27 scenes were grouped, of which two were used for training the expert observers.

Fig. 4. Illustration of the interlace effect.

The training scenes were not included in the results. Scenes of 20 seconds duration were selected, to enable the temporal reduction processes of the video compression algorithms to adjust to the scene content. The local characterization techniques discussed below focused on only 8 fields in scenes of 20 seconds duration. In this duration, a face that appeared in 8 fields at an approximately consistent subject to camera distance, angle to the camera and under constant illumination was selected. The following paragraphs provide information on the scene characterization techniques.

1) *Camera to subject distance***.** This local property was derived objectively, by measuring manually the inter-pupillary distance, in pixels. The average value among the 8 fields of the face was used to classify the face into a selected camera to subject distance group. The scenes were classified empirically into two groups: close (44 pixels distance, +/-4.5 pixels) and far (25 pixels distance, +/- 3.5 pixels)

2) Scene brightness. This local property was derived objectively from measuring skin lightness using the CIELAB *L** metric. Scene illumination and the colour of the person's skin affected the derived lightness (*L*) value. Lightness (*L**) values ranged from 0 (no lightness – black) to 100 (maximum lightness– white). An average of four areas on the face was used. The areas were the forehead, the right cheek, the left cheek and the jaw. In case of facial hair the jaw area was not measured. The average value among the 8 fields of the face was used to classify the face into a selected brightness group. The scenes fell into 5 groups of brightness using two types of illumination (daylight and bus illumination):
1) Medium (bus illumination): L*≈ 42 (+/- 11). 2) Medium (daylight): L* ≈ 46 (+/- 6). 3) Low (daylight): L* ≈ 8.5 (+/- 2.5). 4) High (daylight): L*≈ 92 (+/- 4.5). 5) Mixed (daylight): L*≈ 97 (+/- 2.5) and L*≈ 49.5 (+/- 15.5), (i.e. approximately half of the face had L*≈ 97.5 and the other half L*≈ 49.5).

The medium skin brightness groups differ in terms of 'type' of illumination (i.e. bus illumination and daylight). It was observed that the camcorder produces noisier imagery under bus illumination at night than under daylight illumination. Daylight produces higher illumination levels than bus illumination. To compensate the exposure for decreased levels of illumination, when the bus lights are on, the camcorder increases the ISO settings resulting to increased noise levels. It was thus considered important to include bus illumination on its own in the investigation.

3) Angle of face to camera plane. This local property was deduced subjectively by visual inspection. Two groups were derived: tilted angle and frontal angle. Figure 5 illustrates examples of face angles. Images that include most of both cheeks (between -20 and + 20 degrees on the horizontal axes) and the very top of the head is not visible (between 0 and +10 degrees on the vertical axes) are classified as frontal. Images that include most of both cheeks (between -20 and + 20 degrees on the horizontal axes) and the very top of the head is visible ((e.g. +20 degrees and above on the vertical axes) are classified as tilted.

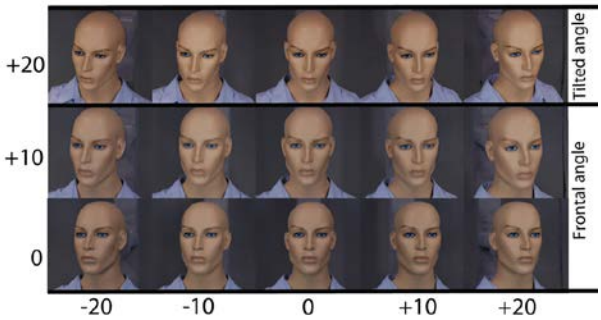


Fig. 5. Partial group of facial angles in degrees.

*4) Busyness***.** This global property was deduced objectively, by implementing an ITU specification [44] measure to derive the spatial and temporal properties of the scenes. The spatial information was extracted by using the standard deviation of Sobel filtered fields; the maximum value represented the spatial information for the scene. The temporal information was obtained by using the standard deviation of the field differences; the maximum value represented the temporal information for the scene. The grouping was made based on the measured spatial and temporal values only of the available 25 scenes. Their middle values were chosen as the limits. For example, the middle value for the spatial

measures is 14.58 and for the temporal is 27.16. The following four groups were created:
1) High Spatial (> 14.58) – High Temporal (> 27.16). 2) High Spatial (> 14.58) – Low Temporal (< 27.16). 3) Low Spatial (< 14.58) – High Temporal (> 27.16). And, 4) Low Spatial (< 14.58) – Low Temporal (> 27.16).

Figure 6 includes all 25 scenes used in the psychophysical investigations. Table I summarizes the grouping of the scenes. For example, scene 1 (S1) belongs to the following groups: medium scene brightness (bus illumination), close camera to subject distance, frontal angle to the camera plane and low spatial - low temporal busyness.

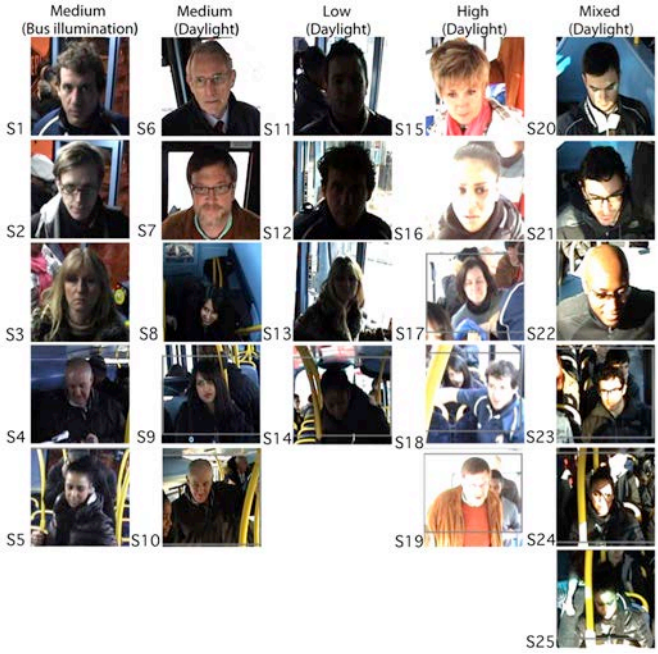


Fig. 6. The 25 scenes grouped in columns based on the scene brightness property.

TABLE I
SUMMARY OF SCENE GROUPING

GROUP NAME	*SCENE NAME*	*TOTALS*
	CAMERA TO SUBJECT DISTANCE	
Close	S1, S2, S3, S6, S7, S11, S12, S15, S16, S20, S21, S22	12
Far	S4, S5, S8, S9, S10, S13, S14, S17, S18, S19, S23, S24, S25	13
	SCENE BRIGHTNESS	
Medium (Bus)	S1, S2, S3, S4, S5	5
Medium (Day)	S6, S7, S8, S9, S10	5
Low (Day)	S11, S12, S13, S14	4
High (Day)	S15, S16, S17, S18, S19	5
Mixed (Day)	S20, S21, S22, S23, S24, S25	6
	ANGLE OF FACE TO THE CAMERA PLANE	
Frontal	S1, S2, S3, S6, S7, S9, S11, S12, S13, S16, S17, S23, S24	13
Tilted	S4, S5, S8, S10, S14, S15, S18, S19, S20, S21, S22, S25	12
	SCENE BUSYNESS	
Low Spa. Low Temp.	S1, S2, S3, S4, S7	5
Low Spa. High Temp.	S5, S6, S20, S21	4
High Spa. Low Temp.	S8, S9, S15, S18, S23, S24, S25	7
High Spa. High Temp.	S10, S11, S12, S13, S14, S16, S17, S19, S22	9

Each scene from figure 6 belongs to different groups. The totals indicate the total number of scenes in the specific group.

2.3 Identification of Key Scenes

The key scenes, those affected most by compression, were identified by carrying out a psychophysical investigation on the 25 grouped scenes. The MPEG Streamclip implementation encoder was employed to compress the scenes at selected target bitrates, using the video coding standard H.264/MPEG-4 AVC. Implementation encoders such as verification models used for compliance testing (e.g. Joint Model (JM) and FFpmeg) are often used by the scientific community; they allow the setting of over 50

parameters, such as quantization parameters, I, P and B frames and the target bitrate. These verification models, when tuned properly, tend to apply 'high quality' compression, whilst encoders in the consumer and CCTV industry apply 'lower quality' compression [45]. It was decided that the verification models were not appropriate for this work. Thus, an encoder from the consumer industry was selected (MPEG Streamclip) with only bitrate control (i.e. no GOP size or B frames were selected), which complies with the security recording systems on buses.

Most of the scenes were compressed at 9 different bitrates, whilst some 'difficult' ones at 12 different bitrates, all at 25 fields per second. The 'difficult' ones were perceived to require less compression to maintain useful information than the rest of the scenes. The levels and ranges of compression were selected empirically, after careful visual examination, to provide enough data for the derivation of an accurate psychometric function [46]. The compression bitrates used were approximately the following in kilobits per second (kbps):
- 9 bitrates: 300, 400, 600, 800, 1000, 1200, 1400, 1600, 1800;
- 12 bitrates: 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800.

A similar method to the fidelity assessment was implemented for the assessment of image usefulness of the compressed scenes in both psychophysical investigations. The observers were presented each time with the reference scene and its compressed version. Both versions were presented on a mid grey background, at 25 fields per second, as illustrated in figure 7. Although the compression was applied on a 20 second scene at 25 fields per second, the observers were only presented with 8 fields, in which the face was placed within a grey square. The observers could see the displayed compressed version field by field and as many times as they wished before making their judgment.

Fig. 7. Example of the test display used in the psychophysical investigations. The left image is the reference whereas the right image is the compressed version of the reference.

The monitor was calibrated to a white point D65 (6500K), at a luminance of 120 cd/m^2 using an sRGB ICC profile. Based on our current knowledge, there are no standards available on monitors used for CCTV viewing purposes. The experiment was conducted in dark conditions to minimize reflections and monitor flare. The specialist observers were asked to wear glasses, if they would normally do so in front of a monitor.

The observers consisted of 7 Metropolitan Police Service (MPS) police officers, 10 MPS surveillance officers, and 10 bus analysts. Table II provides a summary of the observers' average years of experience and purpose of use of security imagery.

TABLE II
OBSERVERS' BACKGROUND

	BUS ANALYSTS	***MPS POLICE***	***MPS SURVEILLANCE***
Average years of experience in assessing security imagery	5 years	9 years	18 years
Use of security imagery	To identify for security purposes and bus issues, gathering evidence for the police.	To identify and provide evidence to court mainly for volume crime (e.g. antisocial behavior, assaults)	To monitor activities and behaviors, identify and provide evidence to court mainly for major crime (e.g. murder)

Instructions to the observers were given via a demonstration of a selected scene from the training set. The training scenes were excluded from the results. The instructions were: "The reference represents the maximum facial information that can be captured under the available illumination conditions and

should be considered to have acceptable image usefulness. The aim is to find how much degradation (compression) from the reference is acceptable. You are required to respond with a *yes* or *no* to the question: *Is the compressed version as useful as the reference in terms of facial information?* You are judging only the face within the grey square, not the clothes or the surrounding area. Everything else that surrounds the face is irrelevant and should not influence your judgment. This experiment will help to identify the maximum acceptable degradation (compression) from an uncompressed reference. If you are paired while doing the experiment, you are allowed to discuss your thoughts with your partner, but your final answer should be independent of your partner's answer. Be aware of peer pressure. If you get bored or tired during the experiment, please inform the experimenter". In most cases, the observers were paired during the experiment. This is usual practice during police examination of CCTV footage.

The *yes*/*no* tasks have the possession of being 'criteria dependent' [47]. For example, the observer might adopt his/her own criteria on the strength of the signal (facial information) before a *yes* response is obtained. If the criterion is loose, then a weak signal might be sufficient, whereas if a strict criterion is adopted then a relative strong signal might be required to obtain a positive response. The observers in this investigation have possibly used criteria that have been derived from their individual work experience. Results are presented in section 3.

2.4 Testing of CCTV Systems

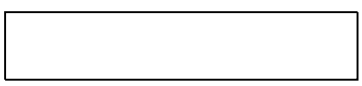
The identified key scenes from section 2.3 were given to five suppliers of CCTV recording systems together with instructions on amount and ranges of compression. The key scenes were compressed at 4 fields per second (which was the requirement by TfL) and the compressed bitrates, in kbps were: 10, 160, 352, 544, 736, 928, 1120, 1312, 1504. The amount and ranges of compression were selected empirically, after careful visual examinations and observation of results obtained from the first psychophysical investigation. Each second consists of 25 fields. Reducing the fields from 25 to 4 per second has resulted, in the majority of cases, to an output from the most CCTV recorders of 1 field from the 8 fields with the face.

In this second psychophysical experiment, the methodology detailed in section 2.3 was followed aside from the number of observers. The number of observers was reduced to 2 MPS police officers and 9 bus analysts. All observers were trained on the task by participating in the first psychophysical investigation. The number of observers is still acceptable for fitting psychometric functions to their responses (see section 3.1).

The observers had to judge the output of each CCTV recorder (1 field) against the reference (8 fields) for each key scene. The performance of the five CCTV recording systems using the key scenes is presented in section 4.

3 Results from the Identification of Key Scenes

The obtained data from the first psychophysical investigation were modelled by fitting psychometric functions (PF), as instructed in [48], for each scene. The PF describes the response of the observers' sensory mechanism to the different stimulus levels (i.e. compression levels). The sigmoid logistic PF was fitted to the obtained psychophysical data points (i.e. proportion of *yes* responses) at each different level of compression in kbps. The logistic function is given as:



(1)

The shape of the curve is established from parameters α, β, and λ. a corresponds to the absolute threshold (*i.e* it is at the point of 50% *yes* responses); β to the gradient of the curve; λ is the stimulus independent lapse rate and was fixed for most fittings at 0.01 except for scene number 24 were the value was fixed at 0.02 (*i.e*. produced a more acceptable fit to the data points). The lapse rate parameter determines the upper bound of the curve given by $1 - \lambda$ (*see* eq. 1). Observer lapses need to be taken

into consideration as they can introduce biases to the estimated *a* and *b* parameters. The effect of lapses can be minimized by setting it to a small but no zero value, such as 0.01 or 0.02 [49, 50]. The maximum-likelihood estimation technique was used to estimate the curve parameters α and β [51, 52].

Figures 8 and 9 present the obtained psychometric functions from the first psychophysical experiment. Tables III and IV include measures of the obtained PFs: a) the estimated function parameters α and β, b) the α and β estimated standard errors (SE), c) the value of goodness of fit (pDev), and d) the value, in kbps, that corresponds to the 75% proportion of observers *yes* responses.

The α and β parameters are just estimates of the true parameters of the sensory mechanism. The errors on the estimated parameters were derived by implementing a non-parametric bootstrap analysis, which is a Monte Carlo re-sampling technique. Bootstrap methods produce simulated repetitions using the data from the original experiment [53, 54]. The standard deviation among the obtained values from the simulated experiments is used as the measure for standard error. In this investigation the recommended 400 converged simulated experiments were used in order to obtain the errors [50]. Stimulations that did not converge were excluded. A parametric method is most frequently suggested when the PF is a good fit to the data points [50]. A non-parametric method was employed in order to sustain a harmonized analysis among all the fitted PFs, the good ones and the less good ones. Additionally, there is a controversy on which of the two methods produces better error estimates [55].

The goodness of fit is a measure that describes how well the curve fits the data. The measure derives the pDev value (i.e. is the statistical p-value) that ranges between 0 (a bad fit) to 1 (the best fit). When the pDev value is less than 0.05 then the fit is considered unacceptably poor. When the curve falls precisely on the points then this indicates a good fit. The goodness of fit measure was calculated using 400 bootstrap simulated experiments and the method is illustrated in [49]. Both α and β were set as free parameters and λ as a fixed parameter during the process of estimating the errors and the pDev values. The 75% of *yes* responses was taken as the just noticeable difference (JND) point on the psychometric curve to identify the acceptable bitrates for the scenes under test. It is typically the value used in qualitative work relating to imaging science [46, 56]. The 50% is defined as the absolute threshold. This is the point where the observers are starting to seeing, in this case, the compressed version to be equal in terms of usefulness with the reference [57]. The subsequent sections provide the analysis of the results from the first psychophysical investigation.

3.1 Psychometric curve fitting

Figure 8 and table III present the results derived from fitting psychometric curves to the data, for each observer group. Figure 9 and table IV present the results from the 25 scenes under investigation. The errors on the obtained β parameters were greater than for the α parameters. Error values could be reduced by increasing the number of observers and, also, by having a better distribution of stimulus intensities. In image related investigations, it is recommended to use between ten to thirty observers. The use of more observers will increase the precision of the estimated values (decrease the error) and not their accuracy [36]. The error estimates in this work are included only for references. The fitting results have shown the pDev values to score above 0.05 for all the scenes and thus all the fitted curves are acceptable (see tables III and IV).

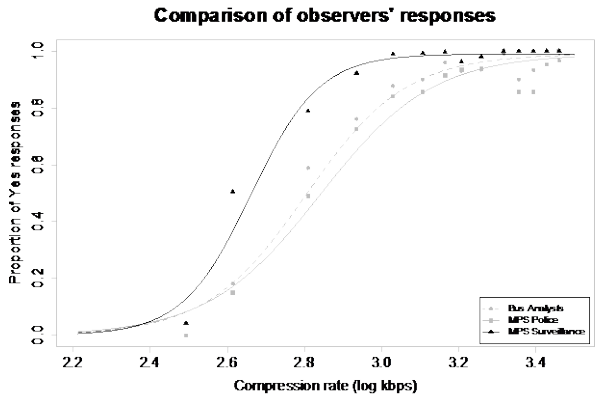


Fig. 8. Psychometric curves for each observer group. The average value of all the tested scenes was used for each observer group.

TABLE III
DATA FROM CURVE FITTING FOR EACH OBSERVER GROUP

	α	SE	β	SE	pDEV	75% (KBPS)
BUS ANALYSTS	2.804	0.042	7.736	2.429	0.977	894
MPS POLICE	2.840	0.048	6.888	2.516	0.950	1014
MPS SURVEILLANCE	2.661	0.038	11.308	3.400	0.923	578

Illustrates the parameter estimates along with their estimated standard error (SE) for the different groups of observers. The goodness of fit is given by the pDev value. The 75% of *yes* responses, in kbps, for each curve is also provided.

The police officers have tolerated less compression for maintaining usefulness, from the original reference, than the bus analysts and surveillance officers (*see* figure 8). The point at 75% of *yes* responses for the police officers were at 1014kbps, for the bus analysts at 894kbps, and for the surveillance officers at 578kbps (see table III). The bus analysts are considered as having the highest technical understanding of video compression and video systems, followed up by the surveillance officers and last the police officers. The surveillance officers have started as police officers and their work, in most cases, involves monitoring (such as following and recording) targeted individuals and gathering evidence to present in court or to help with a case. Their experience and in general the use of different sources of information of the targeted individual (e.g. knowing where the individual has been helps to identify the correct CCTV system to extract supportive imagery) make even a highly compressed CCTV imagery useable for the completion of their task. This is not the case for the police officers as the individuals are often unknown and thus making an identification task from facial imagery more difficult.

TABLE IV
CURVE FITTING DATA FOR EACH OF THE 25 SCENES

SCENE NAME	α	SE	β	SE	pDEV	75% (KBPS)	SCENE NAME	α	SE	β	SE	pDEV	75% (KBPS)
S1	2.611	0.016	26.917	5.699	0.930	450	S14	2.891	0.018	14.359	1.917	0.058	934
S2	2.640	0.020	13.553	2.234	0.135	529	S15	2.677	0.018	22.214	5.147	0.925	534
S3	2.721	0.022	16.374	2.904	0.730	617	S16	2.659	0.017	25.929	5.302	0.388	505
S4	2.683	0.020	17.017	2.663	0.223	562	S17	3.079	0.014	14.522	2.241	0.305	1437
S5	2.795	0.019	13.901	2.959	0.868	753	S18	2.709	0.021	15.07	1.897	0.145	609
S6	2.659	0.021	16.199	3.127	0.150	536	S19	2.828	0.016	17.881	3.744	0.153	780
S7	2.530	0.000	524.084	0.000	0.845	340	S20	2.634	0.022	15.346	12.334	0.103	510
S8	2.531	0.000	531.635	0.000	0.198	341	S21	2.728	0.021	17.246	4.836	0.978	623
S9	2.747	0.021	19.426	5.724	0.673	639	S22	2.709	0.028	13.1	1.925	0.423	625
S10	2.856	0.015	21.501	5.318	0.798	811	S23	2.662	0.020	17.228	3.590	0.068	535
S11	2.953	0.018	15.565	2.421	0.715	1063	S24	2.665	0.018	19.923	12.208	0.063	530
S12	3.057	0.018	10.441	1.300	0.863	1467	S25	2.825	0.017	15.877	2.476	0.553	788
S13	3.044	0.019	9.252	1.063	0.090	1469							

The parameter estimates along with their estimated standard error (SE) for each of the 25 scenes under investigation. The goodness of fit is given by the pDev value. The 75% of *yes* responses, in kbps, for each curve is also provided.

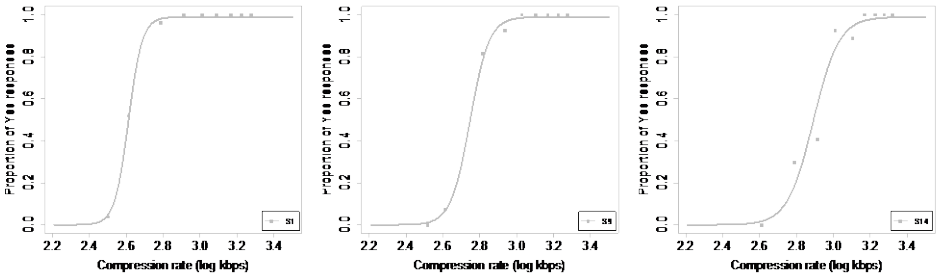


Fig. 9. Three example psychometric fitted curves, for scenes S1, S9 and S14. With high, medium and low derived pDev values respectively.

Few of the scenes have fallen under the exact combination in terms of camera to subject distance, scene brightness and so on (see table I). For example, scenes S1 (75% =450), S2 (75%=529), and S3 (75%=617) represent one exact combination, where 75% is the proportion of *yes* responses (see table IV). Another exact combination is for scenes S11 (75%=1063) and S12 (75%=1467). One more is for scenes S20 (75%=510) and S21 (75%=623). Lastly, another one is for scenes S23 (75%=535) and S24 (75%=530). Most of the exact combinations have produced similar results except for S11 and S12, where the difference is more than 300kbps. Furthermore, in the high brightness group of scenes (S15, S16, S17, S18, and S19) only S17 was affected to a greater degree by compression than the rest. The presented scene characterization methods might not be enough in order to describe the properties of the scenes/faces. For example, the results could be influenced by the degree of distinctiveness/uniqueness or overall appearance of the actual faces in the scenes. Distinctive faces are more memorable [58]. Bruce et al have found that distinctiveness correlates with how much a face deviates from an 'average face' [59]. Perhaps, distinctive faces (e.g. Arnold Schwarzenegger) can take more compression than typical faces (e.g. Leonardo DiCaprio) [58]. Furthermore, Penry provides guidance on how facial features/shapes can be classified [60].

3.2 Comparison of the characterized groups

The point at 75% (in kbps) of *yes* responses for each of the 25 scenes was chosen for further analysis. This analysis investigates the differences and similarities between the characterized groups. Table V illustrates the descriptive statistics for each group. Mostly, the statistics describe the variability of the obtained values of the scenes in each group. The values of the mean and the median for each group are similar, a result indicating near normal distributions. Although, parametric statistics are used with normal distribution, in the following analysis a non-parametric method was used due to the small number of scenes in each group.

TABLE V
DESCRIPTIVE STATISTICS AT 75% OF *YES* RESPONSES

GROUP NAME	***N***	***RANGE***	***MIN***	***MAX***	***MEAN***	***MEDIAN***	***STD***
SCENE BRIGHTNESS							
MED (BUS)	5	303	451	753	582	562	113
MED (DAY)	5	471	340	811	533	536	202
LOW (DAY)	4	534	934	1469	1233	1265	276
HIGH (DAY)	5	933	505	1437	773	609	386
MIXED (DAY)	6	278	510	788	602	579	104
CAMERA TO SUBJECT DISTANCE							
CLOSE	12	1127	340	1467	652	534	325
FAR	13	1127	341	1469	784	753	335
ANGLE OF FACE TO THE CAMERA							
FRONTAL	13	1128	340	1469	778	536	421
TILTED	12	593	341	934	656	624	163
SCENE BUSYNESS							
HIGH SPA. HIGH TEMP.	9	964	505	1469	1010	934	372
LOW SPA. LOW TEMP.	5	277	340	617	500	529	108
HIGH SPA. LOW TEMP.	7	447	341	788	568	535	135
LOW SPA. HIGH TEMP.	4	243	510	753	606	579	110

Where N is the number of scenes in the group. Range is the difference between the minimum (MIN) and maximum (MAX) values. The range, mean, median and standard deviation (STD) are measures of variability of the obtained values of the scenes in the group.

Table VI shows the results from the Wilcoxon Rank Sum Test [61]. This is a non-parametric test that ranks the values of two independent samples and compares the differences between the two rank totals. This method focuses on the median rather than the mean. It derives the *p* statistical value at 0.05 significance level, below which two groups will be considered as statistically different. This method allows gathering the similar groups into a single one.

TABLE VI
WILCOXON RANK SUM TEST

(I) GROUP	(J) GROUP	MEAN DIFFERENCE	p	h
		SCENE BRIGHTNESS		
MED (BUS)	MED (DAY)	49	0.841	0
	LOW (DAY)	651	0.016	1 *
	HIGH (DAY)	191	0.548	0
	MIXED DAY	19	0.662	0
MED (DAY)	LOW (DAY)	700	0.016	1 *
	HIGH (DAY)	240	0.548	0
	MIXED (DAY)	68	0.931	0
LOW (DAY)	HIGH (DAY)	460	0.063	0 *
	MIXED (DAY)	631	0.009	1 *
HIGH (DAY)	MIXED (DAY)	171	0.931	0
		CAMERA TO SUBJECT DISTANCE		
CLOSE	FAR	134	0.097	0 *
		ANGLE OF FACE TO THE CAMERA		
FRONTAL	TILTED	122	0.765	0
		SCENE BUSYNESS		
HIGH SPA. HIGH TEMP.	LOW SPA. LOW TEMP.	510	0.007	1 *
	HIGH SPA. LOW TEMP.	442	0.016	1 *
	LOW SPA. HIGH TEMP.	405	0.050	0 *
LOW SPA. LOW TEMP.	HIGH SPA. LOW TEMP.	68	0.343	0
	LOW SPA. HIGH TEMP.	106	0.286	0
HIGH SPA. LOW TEMP.	LOW SPA. HIGH TEMP.	38	0.788	0

The (I) group is compared against the (J) group. When the p – value is less then 0.05 then the groups are significantly different. Significantly different groups have scored 1 in the *h* column and marked with an asterisk. The 0 values in the *h* column that have been marked with an asterisk are results that are marginally significant.

The variability measures (in particular range and standard deviation) in table V are not small enough to allow a single model of a psychometric curve to be representative of all the scenes in each group. Instead, table VI reveals the similarity/difference between each grouped category. For example, when two groups are similar then they could be further classified to the same group (e.g. no significant difference between 'medium brightness – bus illumination' and 'medium brightness – daylight' scene groups).

The results have shown that the 'low – daylight' brightness group is significantly different from all the other brightness groups except for group 'high – daylight', which it is marginally significant. The groups 'medium – bus illumination', 'medium – daylight', 'high – daylight' and 'mixed – daylight' can be further classified to the same group as there is not a significance difference among them. The 'low – daylight' scenes were affected more by compression than the rest of the brightness groups as the mean value of the scenes for the 75% of *yes* responses is at 1233kbps where for the rest of the brightness scenes is less than 773kbps (see table V).

There is marginally significant difference between the two camera to subject distance groups (table VI) were scenes in the far distance group (mean value of 75% *yes* responses at 784kbps) were affected more by compression than the close distance group (mean value at 75% *yes* responses at 652kbps - see table V).

There was no a significance difference between the two angle of face to camera plane groups so they could be further classified to the same group. This requires a further investigation with perhaps higher degrees of tilted angles.

The busyness of the scenes affected compression performance. Scenes with ‘high spatial – high temporal’ busyness were significantly different from all the other busyness groups, except for group ‘low spatial – high temporal’ which it is marginally significant (see table VI). All the busyness groups excluding the group of ‘high spatial – high temporal’ can be classified into one group. The scenes in the ‘high spatial – high temporal’ group have given a mean value of 1010kbps for the 75% *yes* responses whereas for the other groups it is around 550kbps (see table V).

4 Results from Testing of the CCTV Systems with the Selected Key Scenes

One scene from each of the following four scene brightness groups was selected: ‘high – daylight’, ‘medium –daylight’, ‘medium – bus illumination’ and ‘mixed – daylight’. A further two scenes from the ‘low – daylight’ group were selected. All six scenes, were these most affected by the compression. These key scenes (illustrated in Figure 10) were given to the CCTV suppliers for further investigation of the acceptable compression bitrates on London buses.

Fig. 10. The selected key scenes.

Figure 11 illustrates an example of the output of the CCTV systems (labelled A, B, C, D and E) for key scene S12. As mentioned above, in most cases the CCTV systems exported, 1 field from the 8 reference fields of the face. Even a small changeability in terms of subject to camera distance within each individual scene has affected the obtained results. For example, system C in Figure 11 has obtained more *yes* responses at 736kbps than at 1120kbps because at 736kbps the face is closer to the camera. This could have been completely controlled by using still images, but it would not have replicated reality.

Additionally, Figure 11 illustrates an example of the visual differences between the outputted images from each CCTV system. System C has brightener the scene (by enhancement) whereas compression artefacts are more visible for systems D and E. The systems are behaving differently among them even though all of them are based on the H.264/MPEG-4 AVC compression standard. This presents challenges in drawing conclusions about universal ‘average’ bitrates.

Fig. 11. An example of the outputs of the CCTV systems (labelled A, B, C, D, and E) for key scene 12. The images on the top row are the 8 images of the reference. The second row shows the exported images from system A at different kbps (e.g. between 10 – 1504 kbps). The third row shows the exported images from System B and so on.

The results from the second psychophysical investigation illustrate the probabilistic nature of CCTV systems. For example, by reducing the frame-rate from 25 to 4 has outputted one image from the eight images of the face. This outputted one image might be the worst, or the best-case scenario from the eight available images of the face. Even a slight difference in terms of camera to subject distance within each individual scene has been shown to affect the results for the CCTV systems. For this

reason, the analysis of the results is based on the performance of all five CCTV recording systems for each key scene. The same curve fitting method from the first psychophysical experiment was applied. Three curves were fitted for each key scene: a) the worst performance to the minimum points (lowest fit), b) the middle performance to the average points (average fit), and c) the best performance to the maximum points (highest fit). The lapse rate (λ) was fixed for most fittings at 0.01 except for $S12_{HIGHEST}$ were the value was fixed at 0.02 (i.e. produced a more acceptable fit to the data points).

Figures 12 and Table VII present the results obtained from the testing of the CCTV systems. TfL was after the absolute minimum bitrate to maximize data storage, so the 60% of observers *yes* responses was used instead of the standard 75%.

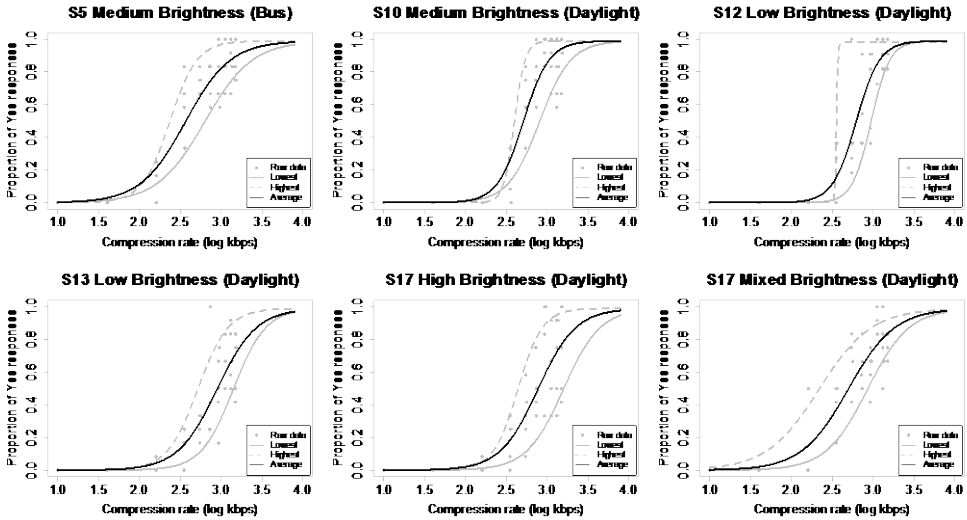


Fig. 12. Psychometric curves for each key scene. Three curves were fitted to all the data points derived from all the systems for each scene: a) worst performance curve using the lowest points (lowest fit) b) average performance curve using the average points (average fit), and c) best performance curve using the highest points (highest fit).

TABLE VII
DATA FROM CURVE FITTING FOR EACH KEY SCENE FROM THE CCTV SYSTEMS

SCENE NAME	α	SE	β	SE	pDEV	60% (KBPS)
$S5_{LOWEST}$	2.799	0.063	3.434	0.709	0.495	840
$S5_{AVERAGE}$	2.563	0.075	3.639	0.822	0.898	480
$S5_{HIGHEST}$	2.37	0.076	5.958	4.513	0.510	277
$S10_{LOWEST}$	2.906	0.047	5.216	1.033	0.083	975
$S10_{AVERAGE}$	2.707	0.042	6.327	1.829	0.653	596
$S10_{HIGHEST}$	2.599	0.03	15.27	8.094	0.640	423
$S12_{LOWEST}$	2.974	0.031	8.717	1.982	0.805	1055
$S12_{AVERAGE}$	2.792	0.042	6.959	1.504	0.895	714
$S12_{HIGHEST}$	2.552	0.001	127.817	6.156	0.174	357
$S13_{LOWEST}$	3.149	0.066	1.588	1.588	0.650	1716
$S13_{AVERAGE}$	2.949	0.062	4.14	1.038	1.00	1131
$S13_{HIGHEST}$	2.713	0.052	5.356	1.443	0.090	621
$S17_{LOWEST}$	3.199	0.092	4.444	1.175	0.725	1977
$S17_{AVERAGE}$	2.887	0.055	4.324	1.04	0.950	969
$S17_{HIGHEST}$	2.642	0.056	6.65	1.77	0.685	509
$S25_{LOWEST}$	2.932	0.059	3.923	0.862	0.348	1100
$S25_{AVERAGE}$	2.695	0.068	3.502	0.827	0.808	657
$S25_{HIGHEST}$	2.338	0.107	3.114	0.656	0.273	299

Illustrates the parameter estimates along with their estimated standard error (SE) for each curve. The goodness of fit is given by the pDev value. The 60% of positive responses, in kbps, for each curve is also provided.

5 Comparison between CCTV and Industry

Table VIII, shows a comparison between the results from the consumer industry compressor (MPEG Streamclip – SC) in the first investigation and from the CCTV systems in the second investigation at

60% of *yes* responses for each key scene. This comparison helps to understand the performance of CCTV recording systems and thus employ appropriate testing methods for such systems.

Table VIII
A comparison between CCTV and Industry compressors at 60% of *yes* responses for each key scene.

SCENE NAME	S5	s10	S12	s13	S15	s25
INDUSTRY (SC)	670	752	1255	1231	1285	711
CCTV LOWEST FIT	840	975	1055	1716	1977	1100
CCTV AVERAGE FIT	480	596	714	1131	969	657
CCTV HIGHEST FIT	277	423	357	621	509	299

The performance of the consumer industry compression at 60% of *yes* responses is in most cases in the middle between the worst (lowest fit) and average (average fit) values of the CCTV systems. Also, the CCTV systems for all the fits have performed better than the consumer industry compression for scene 12 (required less bitrate to maintain facial information). This is because the CCTV systems have enhanced the dark areas by making them look brighter and thus revealed more information within the image. Additionally, it is observed that the CCTV systems might have performed some sharpening to the images, as a result making the information more visible. This does not mean that the image itself will have more information than the consumer industry compressed version. For example, the highest and average curve fits of the CCTV systems outperformed the consumer industry compressor by requiring less bitrate.

6 Conclusion

Acceptable bitrates for video compression depend largely upon scene content properties. Very dark scenes, far distance scenes and scenes with high levels of spatial-temporal busyness were found the most challenging to compress, requiring higher bitrate to maintain useful facial information, necessary for face identification.

Less facial information in the reference scene requires higher bitrates (lighter compression) to sustain the useful information. This can be seen in the results derived for scene brightness and camera to subject distance groups. The low brightness and far distance groups could be considered as having less useful facial information in the reference (in comparison to the medium brightness and close distance groups) and they were affected more by compression than the rest of the groups.

The application (linked to TfL) was seeking the absolute minimum bitrate to maximize data storage, so a 60% of observers *yes* responses was recommended to be used on London buses, which is higher then the absolute threshold of 50%. It was recommended that, during daytime, when there is variable illumination, to set the bitrate to approximately 1977kbps (derived from the worst performance curves, scene 17) and during night-time, when the bus illumination is on, to reset the bitrate to around 840kbps (derived from the worst performance curve for constant bus illumination, scene 5). The findings of this study can be easily extended to others applications.

Future work will involve further investigation into: i) sharpness assessments of CCTV cameras systems using the SFR measure, ii) assessment of additional scenes with more groups (e.g. more subject to camera distances, angle to the camera plane and brightness variations), iii) how face distinctiveness affects acceptable compression levels, iii) the relationship between image usefulness and frame-rate, iv) the use of the same scenes to assess performance of automated face recognition systems, and v) similarly to the use of facial information, garments can be characterized and investigated.

Additionally, findings of this and future investigations could be employed in the creation of quality metrics. For example, a study by Maalouf et al [62] has focused on monitoring quality of legal evidence images in video-surveillance applications by using a combination of a tracking algorithm, a quality metric and a super-resolution algorithm. Furthermore, a more challenging task will be to define quantitatively the relationship between video parameters (e.g. frame rate, bitrate) and image attributes (e.g. busyness, lightness) with the acceptability of usefulness of the face.

Acknowledgment

The authors thank the Home Office Centre for Applied Science and Technology (CAST) for funding this work. Fell David from TfL for assisting in the execution of the methodology. Also, Steve Bleay and Tony Clark from the Home Office CAST for advice on the methodology and analysis of results.

[1] G. Hutton, and D. Johnston, Police manual evidence and procedure, Blackstone press Ltd, 1998/1999.
[2] M. Gill, and A. Spriggs, Assessing the impact of CCTV, Home Office Research Study 292, Home Office Research, Development and Statistics Directorate, (2005).
[3] K. Bashir, X. Tao, and G. Shaogang, Feature selection on gait energy image for human identification, in: *IEEE Int. Conf. Acoustics, Speech and Signal Processing, ICASSP*, Las Vegas, NV, 2008, pp. 985-988.
[4] D. Fell, private communications, June 2011.
[5] Transport for London (TfL). (2012). CCTV [Online]. Available: http://www.tfl.gov.uk/termsandconditions/22246.aspx.
[6] Y. Moses, Y. Adini, and S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, Computer Vision – ECCV '94 800 (1994) 286-296.
[7] H. Hill, P. G. Schyns, and S. Akamatsu, Information and viewpoint dependence in face recognition, Cognition 62(2) (1997) 201-222.
[8] G. Davies, and S. Thasen, Closed-circuit television: how effective an identification aid?, Br. J. of Psychology 91 (2000) 411-426.
[9] R. Kemp, N. Towell, and G. Pike, When seeing should not be believing: photographs, credit cards and fraud, Applied Cognitive Psychology 11(3) (1997) 211-222.
[10] H. Hill, and V. Bruce, The effects of lighting on the perception of facial surfaces. J. of Experimental Psychology: Human perception and performance. 22(4) (1996) 986-1004.
[11] A. J. O'Toole, S. Edelman, and H. H. Bülthoff, Stimulus-specific effects in face recognition over changes in viewpoint, Vision research 38(15-16) (1998) 2351-2363.
[12] V. S. Ramachandran, Perceiving shapes from shading, Scientific American 259(2) (1988) 76-83.
[13] S. N. Yendrikhovskij. Image quality: Between science and fiction. In: *Proc. IS & T PICS* 1999: Imager Processing, Image Quality, Image Capture, Systems Conference; 1999 April; Savannah (Georgia), pp. 173-178.
[14] Rec. ITU-T P.912, Subjective video quality assessment methods for recognition tasks, in series P: Terminals and subjective and objective assessment methods, (2008).
[15] S. N. Yendrikhovskij, Image quality and colour characterization, in: Colour image science: Exploiting digital media, L. McDonald, and M. Ronnier Luo, Ed. Chichester, UK: John Wiley and Sons, 2002, pp. 393-420.
[16] L. Janowski, M. Leszczuk, M. C. Labari and A. Ukhanova. Recognition tasks. In: Quality of experience: Advanced concepts, applications and methods, S. Moller, and A. Raake, Ed. Springer: T – Labs series in telecommunication services, Berlin, Germany, 2014, pp. 383 – 394.
[17] D. A. Silverstein, and J. E. Farrell. The relationship between image fidelity and image quality. In: Proc. Image Processing, 1996, pp. 881- 884.
[18] J. Zhao, Effect of JPEG2000 compression on fingerprint image quality, MSc thesis. Dept Imaging Science, University of Westminster, 2006.
[19] M. Bosch, Z. Fengqing, and E. J. Delp, Segmentation-based video compression using texture and motion models, IEEE J. of Selected topics in signal processing. 5(7) (2011) 1366-1377.
[20] G. J. Sullivan, and T. Wiegand, Video compression from concepts to the H.264/AVC standard, in: *IEEE Proc. 93(1)* (2005) 18-31.
[21] T. Sikora, Trends and perspectives in image and video coding, in: *IEEE Proc. 93(1)* (2005) 6-17.
[22] T.87, Information technology – Lossless and near lossless compression of continuous tone still images baseline, ITU-T Rec., 1998.
[23] G. K. Wallace. The JPEG still picture compression standard. IEEE Trans. Consumer electronics. 38(1) (1992) xviii-xxxiv.
[24] N. R. Axford, Digital image processing and manipulation, in: The manual of photography: Photographic and digital imaging, 9th ed. R. E. Jacobson, S. F. Ray, G. G. Attridge, and N. R. Axford, Ed. Focal Press, Elsevier science Ltd, 2000, pp. 428-446.
[25] P. D. Symes, Video compression demystified, McGraw-Hill. 2001.
[26] A. Tsifouti, S. Triantaphillidou, E. Bilissi and M.-C. Larabi, Acceptable bit rates for human face identification from CCTV imagery. In: Proc. SPIE 8653, Image quality and system performance X. San Francisco, USA. 865305, 2013.
[27] M. Klima, and K. Fliegel. Image compression techniques in the field of security technology: examples and discussion. In: Proc. Security Technology: 38th Annual Int. Carnahan Conf., 2004, pp. 278-284.
[28] M. Klima, and V. Kloucek. Some remarks on very high-rate image compression and its impact on security image data subjective evaluation. In: Proc. Security Technology: 36th Annual Int. Carnahan Conf., 2002, pp. 198-201.
[29] E. Allen, S. Triantaphillidou, and R. E. Jacobson. Image quality comparison between JPEG and JPEG2000. I. Psychophysical investigation. J. of imaging science and technology. 41(3) (2007) 248-258.
[30] S. Triantaphillidou, E. Allen, and R. E. Jacobson, Image quality comparison between JPEG and JPEG2000. II. Scene dependency, scene analysis, and classification. *J. of imaging science and technology. 51(3)* (2007) 259-270.

[31] E. Allen, S. Triantaphillidou, R. E. Jacobson, Perceptibility and acceptability of JPEG 2000 compressed images of various scene types, in: Proc SPIE 9016. Image quality and system performance XI. San Francisco, USA, 2014.
[32] S. Triantaphillidou, Introduction to image quality and system performance, in: The manual of photography, 10th ed., E. Allen and S. Triantaphillidou, Ed. Focal Press, 2011, pp. 345-364.
[33] H. Kalva, The H.264 video coding standard. IEEE Multimedia. 13(4) (2006) 86-90.
[34] M. Ghanbari. Standard codecs: Image compression to advanced video coding. IET: Telecommunications series 49. 2003.
[35] G. Gescheider, Psychophysics: the fundamentals, 3rd ed., Lawrence Erlbaum Associates, 1997.
[36] P. G. Engeldrum, The process of scaling and some practical hints, in: Psychometric scaling: A toolkit for imaging systems development, Imcotek Press: USA, 2000, pp. 19-42.
[37] BT. 500-11, Methodology for the subjective assessment of the quality of television pictures, ITU Rec., 2002.
[38] M. Klima, J. Pazderak, M. Bernas, P. Pata, and J. Hozman. Objective and subjective image quality evaluation for security technology. In: Proc. Security Technology: 35th Annual Int. Carnahan Conf., 2001, pp. 108-114.
[39] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, Overview of the H.264/AVC video coding standards, IEEE Trans. Circuits and systems for video technology 13(7) (2003) 560-576.
[40] CASTBUS_2012, Dataset, Home Office Centre for Applied Science and Technology, Sandridge, UK, 2012.
[41] BS ISO 12233:2000, Photography. Electronic still-picture cameras. Resolution measurements, 2000.
[42] E.W. Jin, B.W. Keelan, J. Chen, J.B. Phillips, and Y. Chen, Softcopy quality ruler method: implementation and validation. In: Proc. SPIE 7242, Image Quality and System Performance VI. San Francisco, USA. 724206, 2009.
[43] C. Poynton, *Digital video and HDTV: algorithms and interfaces*. Elsevier science: USA, 2003.
[44] P.910, Subjective video quality assessment methods for multimedia applications. ITU-R Rec., 1999.
[45] A. Tsifouti, M. M. Nasralla, M. Razzak, J. Cope, J. M. Orwell, M. G. Martini, and K. Sage, A methodology to evaluate the effect of video compression on the performance of analytics systems, in: Proc. SPIE 8546, Optics and photonics for counterterrorism, crime fighting, and defence VIII, 2012.
[46] P. G. Engeldrum, Image quality and psychometric scaling, in: Psychometric scaling: A toolkit for imaging systems development, Imcotek Press: USA, 2000, pp. 1-4.
[47] F. A. A. Kingdom, and N. Prins, Classifying psychophysical experiments, in: Psychophysics: a practical introduction. Elsevier Ltd. 2010, pp. 9-37.
[48] F. A. A. Kingdom, and N. Prins, Psychophysics: a practical introduction, Elsevier Ltd, 2010.
[49] F. Wichmann, and N. J. Hill, The psychometric function: I. Fitting, sampling, and goodness of fit. Perception and psychophysics. 68(3) (2001) 1293-1313.
[50] F. A. A. Kingdom, and N. Prins, Psychometric functions, in: Psychophysics: a practical introduction. Elsevier Ltd. 2010, pp. 59-118.
[51] D. Collet, Modelling binary data. Springer science business media, B.V.1991.
[52] A. J. Dobson, Introduction to generalised linear models. Chapman and Hall: London, 1990.
[53] F. Wichmann, and N. J. Hill, The psychometric function: II. Bootstrap-based confidence intervals and sampling. Perception and psychophysics. 68(8) (2001) 1314-1329.
[54] A. Janssen and T. Pauls, How do bootstrap and permutation tests work?. The annals of statistics. 31(3) (2003) 768-806.
[55] S. Klein, Measuring, estimating, and understanding the psychometric function: A commentary. Perception and psychophysics. 63(8) (2001) 1421-1455.
[56] B. W. Keelan, Characterization of quality, in: Handbook of image quality: Characterization and prediction. Marcel Dekker Inc: New York, USA, 2002, 1-140.
[57] P. G. Engeldrum, Thresholds and just-noticeable difference, in: Psychometric scaling: A toolkit for imaging systems development, Imcotek Press: USA, 2000, pp. 53-78.
[58] P. J.B. Hancock, V. Bruce, and A. M. Burton, Recognition of unfamiliar faces. Trends in cognitive sciences. 4(9) (2000) 330-337.
[59] V. Bruce, M. A. Burton, and N. Dench, What's distinctive about a distinctive face?. The quarterly J. of experimental psychology section A. 47(1) (1994) 119-141.
[60] J. Penry, Looking at faces and remembering them: a guide to facial identification. Elek, 1971.
[61] F. Wilcoxon, Individual comparisons by ranking methods. Biometrics Bulleting. 1(6) (1945) 80-83.
[62] A. Maalouf, M.-C. Larabi, and D. Nicholson, Offline quality monitoring for legal evidence images in video surveillance applications, Multimedia tools and applications [Online] (2012). Springer US: ISSN 1573-7721 Available: http://link.springer.com/article/10.1007%2Fs11042-012-1268-9.