Acceptable bit-rates for human face identification from CCTV Imagery

Anastasia Tsifouti[a,b], Sophie Triantaphillidou[b], Efthimia Bilissi[b], Mohamed-Chaker Larabi[c]

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

ABSTRACT

The objective of this investigation is to produce recommendations for acceptable bit-rates of CCTV footage of people onboard London buses. The majority of CCTV recorders on buses use a proprietary format based on the H.264/AVC video coding standard, exploiting both spatial and temporal redundancy. Low bit-rates are favored in the CCTV industry but they compromise the *image usefulness* of the recorded imagery. In this context *usefulness* is defined 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) Collection of representative video footage. 2) The grouping of video scenes based on content attributes. 3) Psychophysical investigations to identify key scenes, which are most affected by compression. 4) Testing of recording systems using the key scenes and further psychophysical investigations. The results are highly dependent upon scene content. For example, very dark and very bright scenes were the most challenging to compress, requiring higher bit-rates to maintain useful information. The acceptable bit-rates are also found to be dependent upon the specific CCTV system used to compress the footage, presenting challenges in drawing conclusions about universal 'average' bit-rates.

Keywords: CCTV recording systems, Image usefulness, Psychophysics, Face identification, H.264/AVC

1. INTRODUCTION

Closed circuit television (CCTV) is used on public buses to prevent crime, identify offenders/actions and for insurance purposes [1, 2]. The objective of this investigation is to produce recommendations for acceptable bit-rates for CCTV footage of faces of people onboard London buses. The findings are incorporated within the Transport for London standards.

The acceptable bit-rates were derived from typical bus footage (scenes) of human faces. Short scenes were grouped based on their inherent properties; for example scene brightness, camera to subject distance, angle of the face to the camera and level of busyness (spatial - temporal information). The majority of CCTV recorders on London buses use proprietary formats based on the H.264/AVC video coding standard [1, 3]. Two psychophysical investigations were conducted. The first was used to identify the *key scenes*, which are most affected by compression using an implementation of H.264/AVC. The second was used to identify acceptable bit-rates of the key scenes using five of the most commonly used CCTV recording systems on buses.

Observers (experienced civilian analysts and police staff) were polled to give their opinion on what they considered to be acceptable reduction of information from an 'uncompressed' reference source by answering with a *yes* or *no* to the question "Is the compressed version (s) as useful as the reference in terms of facial information?" We present a background summary of the relevant aspects of *image usefulness* with respect to face identification and video

Image Quality and System Performance X, edited by Peter D. Burns, Sophie Triantaphillidou,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8653, 865305 • © 2013 SPIE-IS&T
CCC code: 0277-786X/13/$18 • doi: 10.1117/12.2004140

compression. The subsequent sections describe the experimental methodology, the analysis of the results and our conclusions.

1.1 Image usefulness with respect to face identification

Image usefulness is a visio-cognitive attribute of image quality that relates to "the degree of apparent suitability of the reproduced image to satisfy the correspondent task" [4, 5]. In this context, the specific task requires enough useful facial information to remain in the compressed image in order to allow a specialist to identify a person in the video footage. Image usefulness should not be confused with image fidelity (no visible distortion or loss of information [6]). For example, an image with visible compression artefacts is considered distorted, but if the artefacts do not hide any facial information it maintains its usefulness. There are two main factors affecting the performance of compression:

1. *Content of the scene.* Compression performance is highly dependent on scene content [7-10]. Scenes with different motion properties (temporal differences), regions (spatial differences) and combinations of different spatial - temporal properties will require different bit-budgets leading to different levels of compression. Figure 1 provides an example of the H.264/AVC encoder performance under different illumination conditions.

Figure 1 As the bit-rates decrease 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 sensitive to compression and perhaps requiring lighter compression to achieve positive responses from observers in comparison to the well-illuminated scene in the middle.

2. *Compression algorithm and its properties and settings*. For example, whether the algorithm is based on Discrete Cosine Transform (DCT), or Discrete Wavelet Transform (DWT), settings of intra–coding or inter–coding, quantization parameters, reference frame selection, use of post-processing tools and more.

Research on identification of humans from facial imagery has shown that we have an excellent ability to identify known faces (e.g. derived from memory of faces that were learned minutes, hours or years ago) and our performance drops dramatically when faces are unknown [11-15]. For instance, a study on face identification from poor quality CCTV has shown that participants performed poorly when identifying unknown individuals and very well with known individuals [11]. The same authors explored the factors (gait, body, face) that enabled positive identification of the known individuals. The authors obscured the body, face, or gait and results suggested that, even from poor quality imagery, the face was the most crucial element in identification.

1.2 Video compression on London Buses

The majority of the CCTV recorders on London buses use a proprietary format based on the H.264/AVC encoder. H.264/AVC is a hybrid video encoder, exploiting both spatial and temporal redundancy and uses a 4×4 integer transfer (an approximation of the 4x4 DCT). H.264/AVC produces blocking artefacts that become more visible at low bit-rates [16, 17]. Low bit-rates are favoured in the CCTV industry, since they allow more hours of recording and lower the costs of the system. However, they compromise the image usefulness of the recorded imagery.

The ITU-T and ISO/IEC JTC 1 video standards (such as H.264/AVC) specify only the decoding part in order to ensure interoperability and syntax capability between different technologies implementing the standard. This allows developers

of compression algorithms to optimise compression implementations. Image quality is not specified in the standards and different implementation of encoders will produce different 'compressed qualities'. For example, the performance of detection systems was affected by the specific implementation of the software used for the H.264/AVC encoder [18].

2. METHODOLOGY

2.1 Overview

The investigation comprised four steps: 1) Collection of representative video footage, 2) Characterization and grouping of video scenes based on content attributes. 3) Identification of key scenes, which are most affected by compression using an industry standard H.264/AVC compression. 4) Testing of five CCTV recording systems, which are commonly used on London buses using the key scenes.

2.2 Collection of representative video footage

Variable lighting conditions encountered by buses create a challenging environment for CCTV systems. When the sun illuminates one side of the bus, some areas are over- and some others under- illuminated. As the bus moves, the windows allow illumination from different directions, causing the areas of over- and under- illumination to vary rapidly (uncontrollable lighting conditions). On the contrary, an overcast day will produce diffuse light and uniform illumination (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 (controlled lighting conditions). The following conditions were used for data collection.

- Camera system: A consumer quality mini digital video (DV) camcorder was used for all the filming on an automated exposure setting, which was chosen to replicate what happens with CCTV cameras. The cameras' set up was according to TfL's recommendations (i.e. camera views – 10 camcorders were installed).
- Illumination conditions: Sunny day and bus illumination (lights on bus during night time), together with some exterior illumination e.g. from shops.
- Participants: 26 actors from various ethnicities, ages and gender acting as the bus passengers.

2.3 Comparison between CCTV camera and DV camcorder

There are several companies that provide CCTV systems to London buses; these vary in terms of quality. Figure 2 illustrates the Spatial Frequency Response (SFR) [19] of one sample CCTV camera used on buses compared against the SFR of the DV camcorder used for the collection of the representative video footage. The DV camcorder response indicates sharpening in the vertical orientation in the low and mid frequencies. Further, it has a much greater optical resolution (the SFR falls to 0.1 at nearly 4 c/pixels) and produces sharper images (0.5 SFR corresponds to approximately 2.7 c/pixels) than the CCTV (optical resolution limit at less than 3 c/pixels and 0.5 SFR at less than 2 c/pixels).

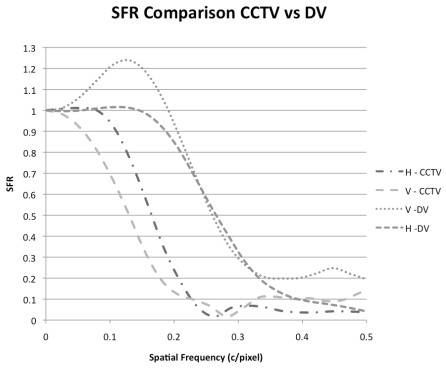

Figure 2 Horizontal (H) and Vertical (V) SFR of CCTV camera and DV camcorder.

The DV camcorder was chosen over a CCTV camera mainly because of convenience and cost. To record in an uncompressed format using the output of a CCTV camera requires expensive specialized equipment (e.g. signal

converters, hard drives, laptops, portable chargers), which in return require a considerable amount of time to set up (the bus was lent to us for a day). Also, the quality of the CCTV systems currently installed on buses varies considerably and these variations are unknown.

Consumer DV camcorders are generally considered to produce higher quality than CCTV systems, but as video system technology develops, CCTV systems are expected to approach the image quality of the consumer industry.

2.4 Preparation of the 'Uncompressed' reference

The 'uncompressed' reference was recorded in a DV format, at 25 megabits per second (Mbits/s), 4:2:0 chroma subsampling, and at full D1 PAL resolution (720 × 576). The reference was further compressed using MPEG-2 compression at approximately 25Mbits/s (4:2:0 chroma subsampling) and was supplied on a DVD to five suppliers of bus CCTV systems. The suppliers were asked to create a set number of compressed versions of the reference footage and return them for use in the experimental testing.

Both DV and MPEG-2 encoders are based on the DCT algorithm [20]. The main difference between DV and MPEG-2 is in the temporal domain. DV exploits only spatial redundancy whereas MPEG-2 exploits both spatial and temporal redundancy. An initial empirical experiment was conducted to understand the visible differences between the two encoders. The experiment involved careful observation of a number of compressed scenes, with various scene properties. No visibly noticeable difference was found in the compressed scenes. Figure 3 illustrates an example comparison. The compression bit-rates used in the CCTV industry are much lower than 25 Mbits/s. Thus, the additional compression of the reference using the MPEG-2 encoder should not affect the results.

Figure 3 Comparison of two images compressed at 700kbps, with MPEG-2 (on the right) and DV (on the left) as the references.

The scenes were captured using interlaced scanning at 25 frames per second (fps). In interlace scanning each video frame consists of two fields captured at different times in a successive order. It has been observed that CCTV recording systems sometimes record the two fields as one frame (because of miscommunication between transmission and recording) and this creates the interlace effect (see Figure 4). In order to avoid the interlace effect in the compressed scenes one of the fields (i.e. the odd line numbers) was removed from the reference. Thus, the reference consists of 25 fields per second and not 25 frames per second.

Figure 4 Illustration of the interlace effect

2.5 The characterization and grouping of video scenes

Since compression performance is dependent on scene content, the various captured scenes were characterized and grouped based on local scene attributes (see below camera to subject distance, scene brightness and angle of face to the camera) and global scene attributes (see below 'busyness').

Scenes of 20 seconds duration were chosen because video compression algorithms apply temporal information reduction and require some time to adjust to the scene content. The local characterization techniques discussed below focused only

on 8 fields of 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 face in these 8 fields was classified into a selected scene group. In total, 28 scenes were grouped, of which 2 were used for training the observers. The training scenes were not included in the results. The characterization techniques were the following:

I. **The camera to subject distance** was deduced objectively, by measuring the interpupillary distance in pixels. The scenes were classified into two groups: close (40 pixels distance, +/- 3 pixels) and far (20 pixels distance, +/- 3 pixels)

II. **The scene brightness** was deduced objectively from measuring skin lightness CIELAB L^*. The skin lightness was affected by both the scene illumination and the colour of the person's skin. Lightness levels ranged from 0 (no lightness – black) to 100 (maximum lightness– white). An average of four areas on the faces 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 scenes fell into 5 groups of L^* measures using two types of illumination (daylight and bus illumination):

1) Medium lightness (bus illumination): $L^* \approx 50$,
2) Medium lightness (daylight): $L^* \approx 50$,
3) Low lightness (daylight): $L^* \approx 10$,
4) High lightness (daylight): $L^* \approx 90$,
5) Mixed lightness (daylight): $L^* \approx 90$ and $L^* \approx 50$, (i.e. approximately half of the face had $L^* \approx 90$ and the other half $L^* \approx 50$)

The two medium skin lightness groups (bus illumination and daylight) differ in terms of type of illumination. It was observed that the camcorder produces more noisy imagery under bus illumination at night than under daylight illumination during the day. Daylight conditions result to higher levels of illumination than the bus illumination. To compensate the exposure for decreased levels of illumination when it is dark outdoors and the bus lights are on, the camcorder increases the ISO settings, resulting to increased noise levels. It was considered important to include bus illumination in the investigation.

III. **The angle of face to the camera** was deduced subjectively by visual inspection. Two groups were derived: tilted angle and frontal angle. Figure 5 illustrates an example. 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) the image is classified as tilted.

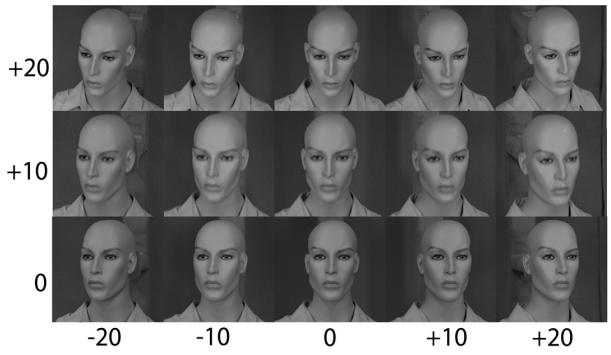

Figure 5 Partial groups of facial angles in degrees

IV. **The scene busyness** was deduced objectively, by measuring the global spatial and temporal properties of the scenes. An objective measure using ITU specifications, was implemented [21]. The spatial information was extracted by using the standard deviation of Sobel filtered fields and the maximum value represented the spatial information for the scene. The temporal information was obtained by using the standard deviation of the field differences and the maximum value represented the temporal information for the scene.

Figure 6 The 26 scenes grouped based on skin lightness.

Figure 6 includes all 26 scenes used in the psychophysical investigations. Table 1 lists the groups that each scene belongs to. For example, scene 1 (s1) belongs to the following groups: medium lightness (bus illumination), close camera to subject distance, frontal angle to the camera and low spatial - low temporal busyness.

Table 1 Scene grouping. Each scene from figure 6 belongs to different groups. The total indicates the total number of scenes in the specific group.

	Scene Brightness (Skin lightness)					**Distance**	
	Medium (Bus illumination)	Medium (Daylight)	Low (Daylight)	High (Daylight)	Mixed (Daylight)	Close	Far
	s1,s2,s3 s4,s5	s6,s7,s8, s9,s10	s11,s12,s13 s14,s15	s16,s17,s18 s19,s20	s21,s22,s23 s24,s25,s26	s1,s2,s3 s6,s7,s11,s12 s13,s16,s17, s21,s22,s23	s4,s5,s8,s9,s10 s14,s15,s18,s19, s20, s24,s25,s26
Total	**5 Scenes**	**5 scenes**	**5 Scenes**	**5 Scenes**	**6 Scenes**	**13 Scenes**	**13 Scenes**
	Busyness					**Angle**	
	High Spatial - High Temporal	Low Spatial - Low Temporal	High Spatial - Low Temporal	Low Spatial– High Temporal		Frontal	Tilted
	s10,s11,s12,s14,s15 s17,s18,s20,s23	s1,s2,s3,s4,s7s13	s8,S9,s16,s19,s24, s25,s26	s5,s6,s21,s22		s1,s2,s3,s6 S7,s9,s11,s12 S14,s17,s18,s24 S26	s4,s5,s8,s10 S13,s15,s16 S19,s20,s21 S22,s23,s26
Total	**9 scenes**	**6 scenes**	**7 scenes**	**4 scenes**		**13 Scenes**	**13 Scenes**

2.6 Identification of key scenes

A first psychophysical investigation was carried out to identify the key scenes (i.e. those affected most by compression). The experiment was conducted using the video coding standard H.264/AVC. The MPEG Streamclip implementation encoder was employed to compress the scenes at selected target bit-rates. Implementation encoders such as Joint Model (JM) and FFpmeg (i.e. these are verification models used for compliance testing of 'industrial' implementations) 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 bit-rate. These verification models tend to apply 'high quality' compressions whilst encoders in the consumer and CCTV industry apply 'lower quality' compressions [18]. 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 bit-rate control (i.e. no GOP size or B frames were selected). Furthermore, the security recording systems on buses employ only bit-rate control.

Most of the scenes were applied 9 different bit-rates, whilst some 'difficult' ones 12 different bit-rates, all at 25 fields per second. The levels and ranges of compression were selected empirically, after careful examinations, to enable the derivation of an accurate psychometric curve [22]. The compression bit-rates used were the following in kilobits per second (kbps):

a) 9 bit-rates: 400kbps, 600kbps, 800kbps, 1000kbps, 1200kbps, 1400kbps, 1600kbps, 1800kbps, 2000kbps
b) 12 bit-rates: 600kbps, 800kbps, 1000kbps, 1200kbps, 1400kbps, 1600kbps, 1800kbps, 2000kbps, 2200kbps, 2400kbps, 2600kbps, 2800kbps

The International Telecommunication Union (ITU) provides guidance on the assessment of video image quality. For example, when image fidelity (i.e. no visible distortion from reference imagery) is assessed, the ITU recommends that the reference video runs simultaneously on a single monitor, along with the reduced quality version [23]. This technique helps the observers to make a direct judgment based on what they see and does not rely on memory. The ITU does not include explicit recommendations for image usefulness assessment. A similar methodology to the fidelity assessment was used for the assessment of image usefulness.

In the psychophysical experiment the observers were presented each time with four versions of the same scene, running simultaneously on a single computer monitor. These were presented on a mid grey background, at 25 fields per second, as illustrated in figure 7. The top left is the reference scene and the other three are compressed versions of the reference scene. 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 (see figure 7). The observers could see the displayed compressed versions field by field and as many times as they wished before making their judgment.

Figure 7 Example of the test display used in the identification of the key scenes.

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 in order to eliminate reflections and minimize monitor flare. The specialist observers were asked to wear glasses if they would normally do so in front of a monitor.

The observers were 7 Metropolitan Police Service (MPS) police officers, 10 MPS surveillance officers, and 10 Bus analysts. Table 2 provides a summary of the observers' average years of experience and purpose of use of security imagery.

Table 2 Observers' background

	Bus Analysts	**MPS Police Officers**	**MPS Surveillance Officers**
Average years of experience in assessing security imagery	5 years	9 years	18 years
Use of security imagery	To identify for insurance purposes and bus issues, gathering evidence for the police	To identify and provide evidence to court mainly for volume crime (e.g. antisocial behaviour, assaults)	To monitor activities and behaviours, 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(s) 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) for setting up recording systems for the London buses. If you are paired while doing the experiment, you are allowed to discuss your thoughts 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 results in section 3.1 present a comparison among the different groups of observers, camera to subject distance, scene brightness, angle of face to the camera and busyness. Section 3.1 also includes the identified key scenes (see Figure 12).

2.7 Testing of CCTV systems

Five suppliers of CCTV recording systems were given the previously identified key scenes and instruction on the amount of compression they should apply to them. The five systems used were based on H.264/AVC compression. The scenes were compressed at 4 fields per second, which was the requirement by TfL. The compressed bit-rates, in kbps were: 10kbps, 160kbps, 352kbps, 544kbps, 736kbps, 928kbps, 1120kbps, 1312kbps, 1504kbps. As mentioned above only the 8 fields from scenes of 20 second duration were judged by the observers. 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 of 1 field from the 8 fields with the face.

In this second psychophysical experiment, the methodology detailed in section 2.6 was followed with two modifications: i) the mode of presentation of the experiment (see figure 8) and ii) the number of observers involved. It was noticed from the previous experiment that the observers would occasionally pay attention to the surrounding areas in the image of the faces. This possibility was eliminated, by cropping the surrounding areas.

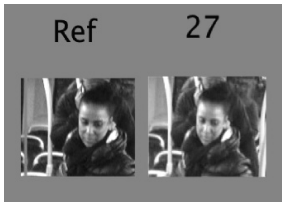

Figure 8 Example of test display in the testing of the CCTV systems

Observers consisted of 2 MPS police officers and 9 Bus analysts. All observers had participated in the first psychophysical investigation. The output of each CCTV recorder (1 field) was judged against the reference (8 fields).

The results from the performance of the five CCTV recording systems using the key scenes are presented in section 3.2.

3.RESULTS

3.1 Results from the identification of key scenes

Figures 9 to 11 and tables 3 to 5 present results from the first psychophysical investigation (identification of key scenes). Values in Tables 3 to 5 were calculated in kbps and are derived from 50%, 60%, 70% and 75% proportion of observers *yes* responses. These were obtained by fitting sigmoid functions to the data. Additionally, the discrepancy between the actual data and the fitted curve is given by the residual values in the tables.

The average bit-rates of each group of scenes are shown. As noted previously, scenes had a variety of bit-rates applied to them depending on their sensitivity to compression (see section 2.6). The figures include error-bars, indicating the

standard uncertainty (standard deviation divided by the square root of the number of samples[24]) of the observers' responses. The standard uncertainty illustrates the uncertainty of the spread of values and of an average.

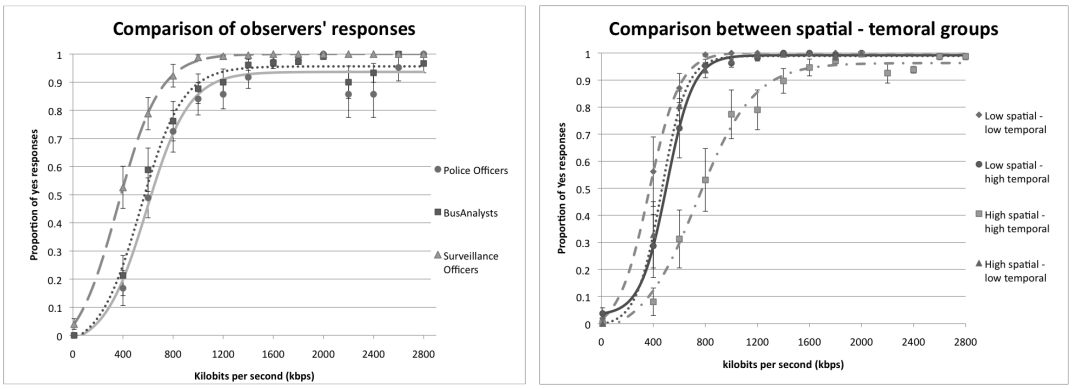

Figure 9 Psychometric curves and data points a) for the different groups of observers (left) and b) for different spatial - temporal groups (right).

Table 3 Results, in kbps, for the different groups of observers and for the different spatial - temporal groups.

Proportion of Yes responses	Police Officer	Bus Analysts	Surveillance officers	Low spatial – Low temporal	Low spatial – High temporal	High spatial – High temporal	High spatial – Low temporal
50%	627	574	385	373	499	777	468
60%	699	642	449	419	514	866	507
70%	784	721	521	470	588	967	550
75%	835	768	562	500	614	1026	575
Residual	0.22	0.12	0.01	0.02	0.03	0.12	0.03

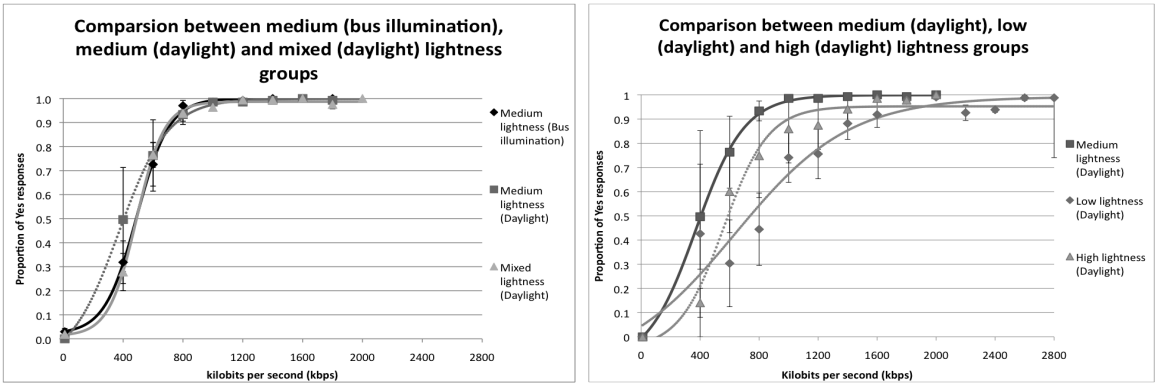

Figure 10 Psychometric curves and data points for the groups with different skin lightness.

Table 4 Results, in kbps for the groups with different skin lightness.

Proportion of Yes responses	Medium (Bus illumination)	Medium (Daylight)	Low (Daylight)	High (Daylight)	Mixed (Daylight)
50%	488	406	739	590	490
60%	532	468	884	653	529
70%	581	537	1048	725	572
75%	608	577	1143	769	596
Residual	0.03	0.02	0.25	0.15	0.03

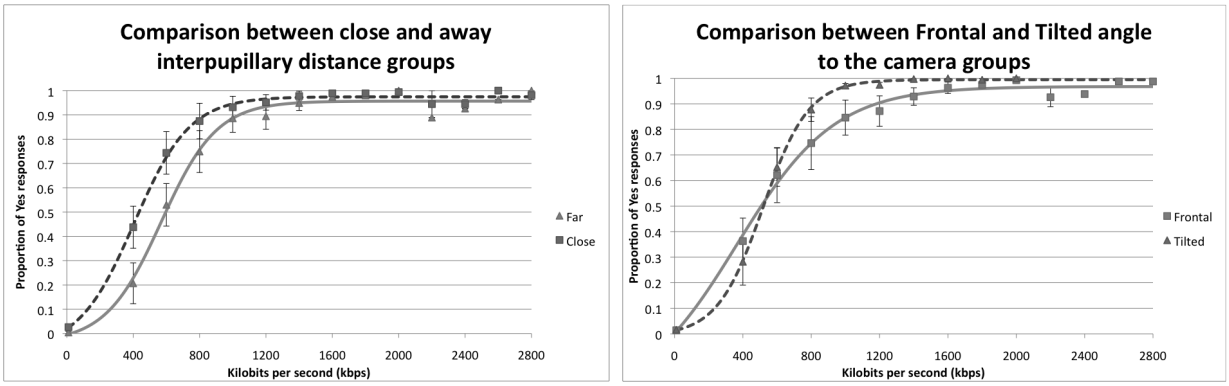

Figure 11 Psychometric curves and data points a) for the interpupillary distance groups (left figure) and b) for angle to the camera groups (right figure).

Table 5 Results, in kbps, from curve fitting for the interpupillary distance and angle to the camera groups.

Proportion of Yes responses	Far	Close	Frontal	Tilted
50%	597	435	508	521
60%	667	502	612	575
70%	746	578	734	635
75%	795	622	808	669
Residual	0.11	0.06	0.08	0.02

One scene from each of the following four groups was selected: 'high lightness - daylight', 'medium lightness -daylight', 'medium lightness – bus illumination' and 'mixed lightness - daylight'. A further two scenes from the 'low lightness' group were selected. All six were the scenes, which were affected the most by the compression. These key scenes (shown in Figure 12) were given to the CCTV suppliers for further investigation of the acceptable compression bit-rates.

Figure 12 The identified key scenes

3.2 Results from testing of the CCTV systems with the selected key scenes

Figures 13 to 18 and tables 6 to 11 present the results obtained from the testing of the CCTV systems. The analysis is based on the performance of each CCTV recording system with each key scene. Also, the averages of all systems for each key scene are included in order to understand if they could be used to advice on acceptable bit-rates. The figures include error-bars indicating the standard error of the observers' responses. The systems are labeled A, B, C, D and E.

The tables include values in kbps, obtained from 50%, 60%, 70% and 75% proportion of observers *yes* responses. These were obtained again by fitting sigmoid functions to the data. Results with residuals higher than 0.4 are highlighted in gray. NA stands for not applicable and it is when the scoring of the recording system for a scene has not obtained the required proportion of observers' *yes* responses.

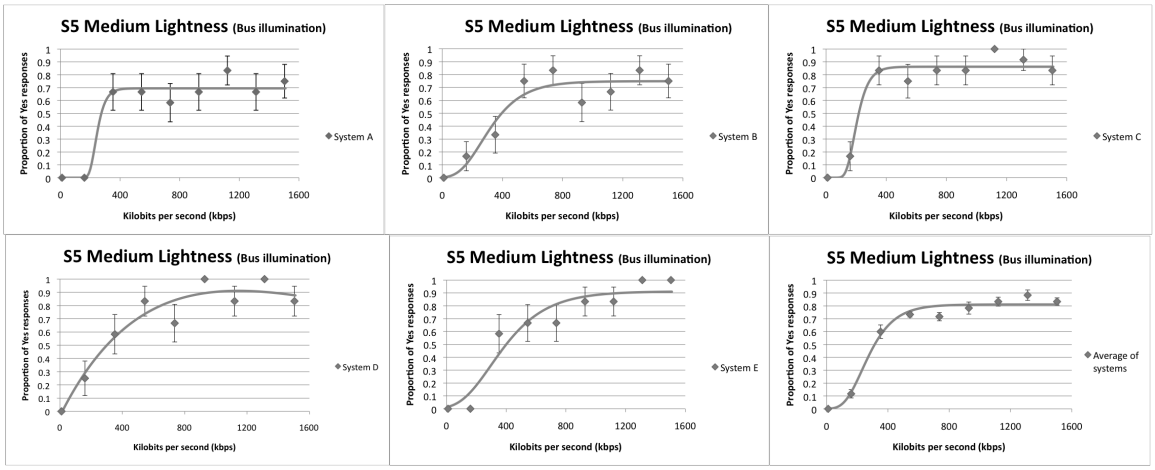

Figure 13 Psychometric curves and data points for scene 5 (medium lightness - bus illumination).

Table 6 Results, in kbps, for scene 5 (medium lightness, bus illumination).

Proportion of Yes responses	System A	System B	System C	System D	System E	Average of systems
50%	274	395	218	300	401	314
60%	305	485	239	364	474	369
70%	NA	552	268	447	567	404
75%	NA	NA	289	504	629	524
Residual	0.19	0.32	0.19	0.27	0.29	0.12

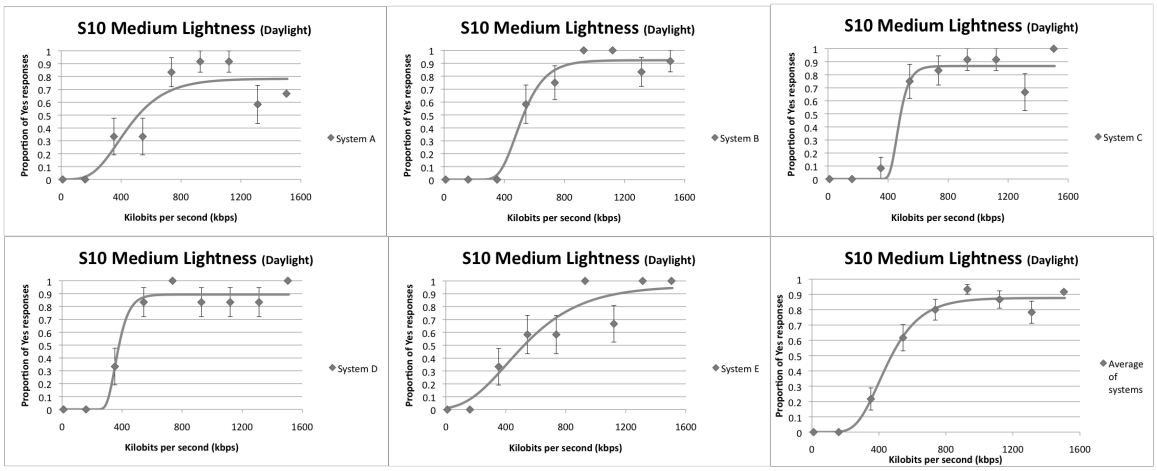

Figure 14 Psychometric curves and data points for scene 10 (medium lightness - daylight).

Table 7 Results, in kbps, for scene 10 (medium lightness - daylight).

Proportion of Yes responses	System A	System B	System C	System D	System E	Average of systems
50%	515	528	485	380	522	476
60%	603	565	503	400	611	529
70%	750	612	527	425	718	598
75%	913	642	545	443	785	647
Residual	0.41	0.18	0.26	0.19	0.34	0.12

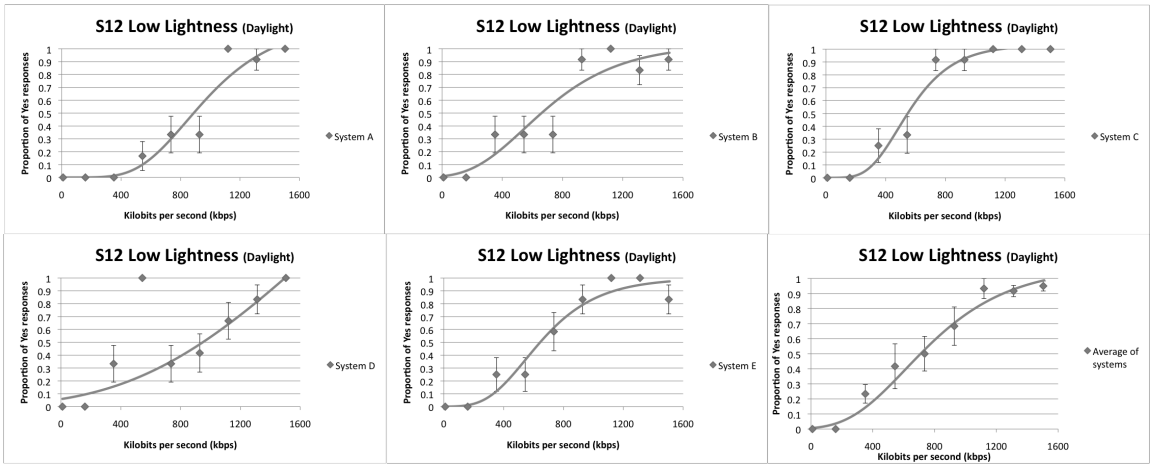

Figure 15 Psychometric curves and data points for scene 12 (low lightness - daylight).

Table 8 Results, in kbps, for scene 12 (low lightness - daylight).

Proportion of Yes responses	System A	System B	System C	System D	System E	Average of systems
50%	889	661	548	949	645	710
60%	966	763	607	1073	724	811
70%	1049	880	674	1189	819	923
75%	1095	949	713	1242	875	988
Residual	0.33	0.39	0.24	0.22	0.26	0.16

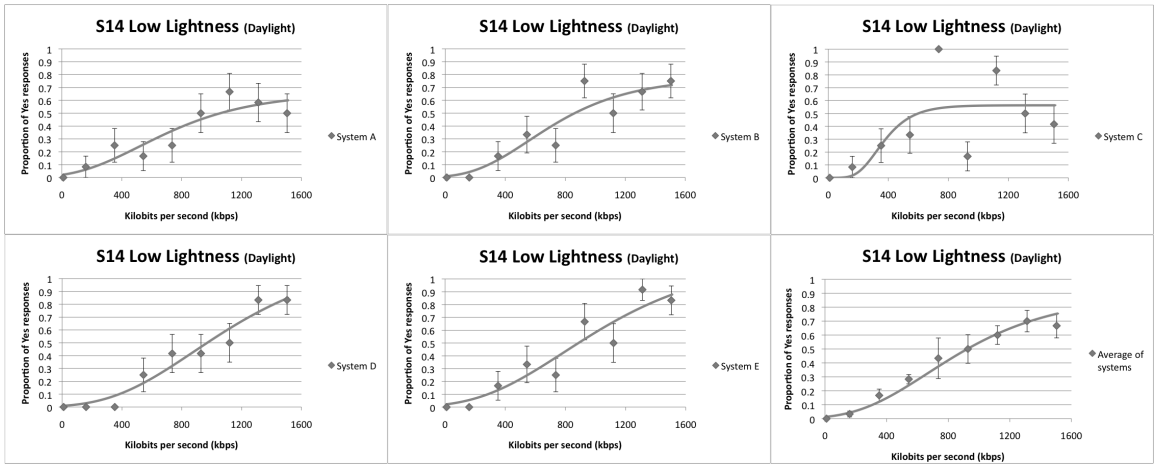

Figure 16 Psychometric curves and data points for scene 14 (low lightness - daylight).

Table 9 Results, in kbps, for scene 14 (low lightness - daylight).

Proportion of Yes responses	System A	System B	System C	System D	System E	Average of systems
50%	1059	863	600	970	887	925
60%	1510	1063	NA	1104	1025	1101
70%	NA	1409	NA	1249	1175	1334
75%	NA	NA	NA	1329	1257	1495
Residual	0.26	0.31	0.70	0.21	0.31	0.15

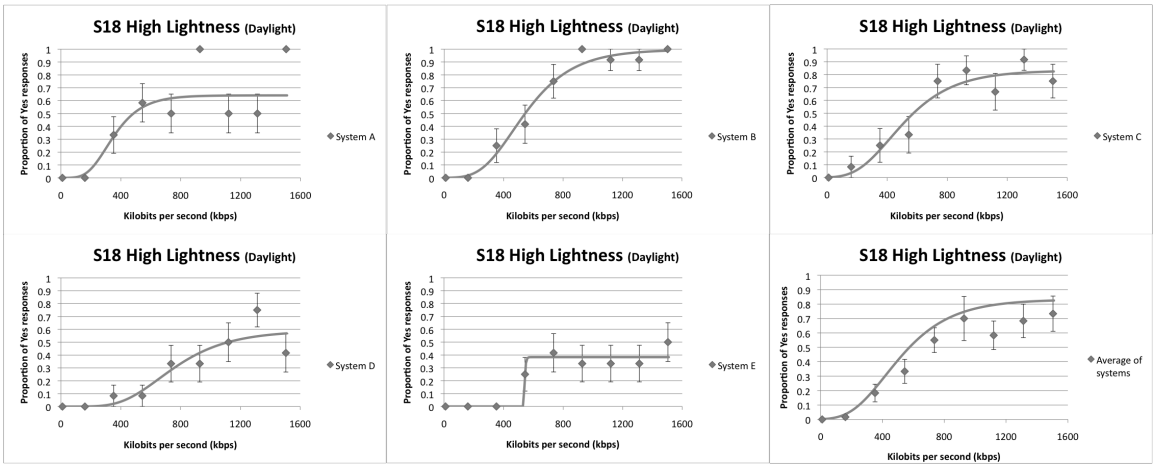

Figure 17 Psychometric curves and data points for scene 18 (high lightness - daylight).

Table 10 Results, in kbps, for scene 18 (low lightness - daylight).

Proportion of Yes responses	System A	System B	System C	System D	System E	Average of systems
50%	485	545	575	1120	NA	685
60%	659	613	677	NA	NA	859
70%	NA	692	822	NA	NA	1500
75%	NA	740	940	NA	NA	NA
Residual	0.56	0.18	0.27	0.28	0.28	0.13

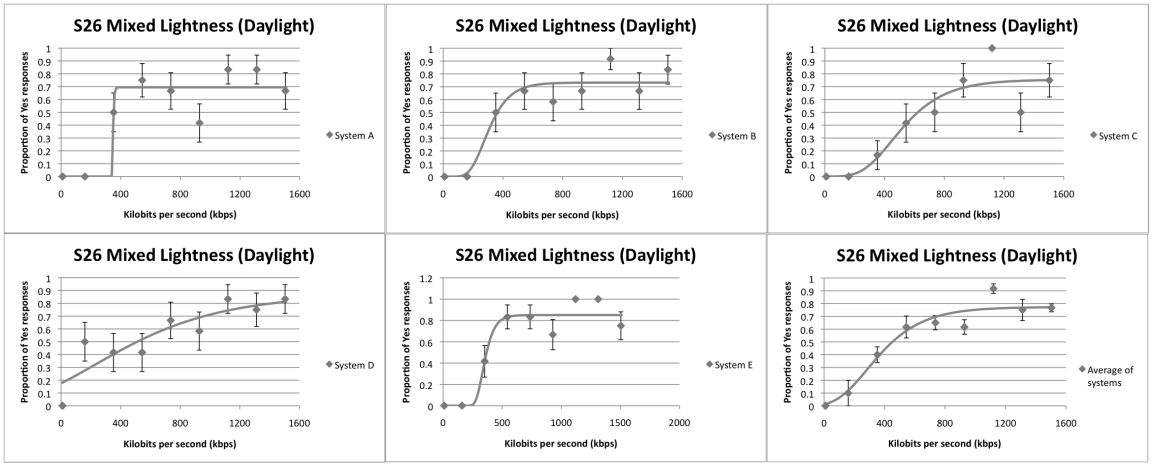

Figure 18 Psychometric curves and data points for scene 26 (mixed lightness - daylight).

Table 11 Results, in kbps, for scene 26 (mixed lightness - daylight).

Proportion of Yes responses	System A	System B	System C	System D	System E	Average of systems
50%	352	363	623	512	369	452
60%	356	424	738	704	392	562
70%	NA	565	957	962	424	750
75%	NA	NA	NA	1146	449	990
Residual	0.35	0.27	0.38	0.34	0.30	0.21

4. DISCUSION

The results have shown that police officers accepted less compression for maintaining usefulness than the bus analysts and surveillance officers (see figure 9a). The 75% of *yes* responses for the police officers were at 835kbps, for the bus analysts at 768kbps, and for the surveillance officers at 562kbps (see table 3). The bus analysts are considered as having the highest technical understanding of compression, followed up by the surveillance officers and last the police officers. The surveillance officers used to be police officers and their work, in most cases, involved monitoring (following and recording) known individuals and gathering evidence to present in court. Their experience and familiarity with the individual make them able to uncover useful information even within highly compressed scenes. The difference in the levels of compression accepted by the police officers and the bus analysts is small.

The busyness of the scenes affects compression performance as seen in Figure 9b. Scenes with high spatial - high temporal busyness required approximately 1026kbps to obtain the 75% proportion of *yes* responses, whilst all other scenes require closer to half of that number (see table3).

The hypothesis was that less information in the reference image will require higher bit-rate numbers (lighter compression) in order to sustain the original information. This can be seen in the scene brightness (Figure 10) and camera to subject distance groups (figure 11a). The low lightness and high lightness groups could be considered as having less useful information in the reference (in comparison to the medium brightness groups) and they were affected more by compression than the rest of the groups (see table 4). Additionally, scenes in the far distance group (which could be considered of having less information than the close group) were affected more than the close distance group (see table 5).

The frontal angle to the camera group required higher bit-rates for the 60%, 70% and 75% of *yes* responses than the titled angle group (see figure 11b). The titled group required higher bit rates than the frontal for the 50% of *yes* responses. This requires a further investigation, since the results are not consistent.

The results from the second psychophysical investigation have shown that for each CCTV recording system different acceptable compression bit-rates were obtained, even though their proprietary formats are based on the H.264/AVC encoder. This presents challenges in drawing conclusions about universal ‘average’ bit-rates. It was also observed that CCTV recording systems perform a more ‘aggressive’ compression than the industry standard compression. For example, some systems did not even obtain 60% of observers’ *yes* responses for scenes 14 and 18 (see tables 9 and 10), at high bit-rates 1504kbps for 4 fields per second (this is equivalent to 9400kbps for 25 fields per second). This low performance of the CCTV systems could be explained as been the results of comparing 8 fields to 1 field of a face. This requires a closer look to the video data in order to understand if the outputted 1 field for each system had less information than the 8 fields in the reference.

5. CONCLUSIONS

TfL was after the absolute minimum bit-rate 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%. The imaging community normally recommends the value of 75% [22, 25]. It was recommended that during daytime, when there is variable illumination, to set the bit-rate of approximately1500kbps (derived from the worst - case performance of System A with the low lightness scene s14 - see table 9) and during nighttime, when the bus illumination is on, to reset the bit-rate to around 700kbps (constant bus illumination). Bus illumination has been shown to produce similar performance with the medium lightness and mixed lightness scenes, thus a bit rate of 700kbps will cover all these groups.

Future work will involve further investigations into: i) further sharpness assessments of the CCTV cameras used on buses using the SFR measure, ii) calculation of the errors within the data derived from the two psychophysical experiments, iii) the assessment of additional scenes with more groups (e.g. more subject to camera distances, angle to the camera and brightness variations), and iv) the relationship between image usefulness and frame rate.

Additionally, findings of this and future investigations could be employed in the creation of quality metrics. For example, a study in[26] 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.

5. REFERENCES

[1] Fell, D., Personal Communications, Transport for London (TfL), (2011).

[2] Transport for London (TfL)., *CCTV*, Available from: http://www.tfl.gov.uk/termsandconditions/22246.aspx, (2012).

[3] Wiegand, T., Sullivan, G, J., Bjontegaard, G., Lutha, A., "Overview of the H.264/AVC video coding standard", IEEE Trans. Circuits Syst. Video Techn. 13 (7), 560–576 (2003).

[4] Yendrikhovskij, S.N., "Image quality and colour characterisation", In [Colour Image Science; Exploiting Digital Media], Edited by MacDonald, L. W. and Luo, M. R., John Wiley & Sons Inc., Chichester, Chap.19, 393-420 (2002).

[5] Yendrikhovskij, S.N., "Image quality: Between science and fiction" Proc. IS & T PICS, 173 - 178 (1999).

[6] Silverstein, D.A., and Farrell, J. E., "The relationship between image fidelity and image quality", Proceedings of the 1996 International Conference of Image Processing, IEEE, 881-884 (1996).

[7] Klima, M., and Fliegel, K., "Image compression techniques in the field of security technology: examples and discussion", Security Technology, 2004. 38th Annual 2004 International Carnahan Conference. 278-284 (2004).

[8] Klima, M., and Kloucek, V., "Some remarks on very high-rate image compression and its impact on security image data subjective evaluation", Security Technology, 2002. Proceedings. 36th Annual 2002 International Carnahan Conference. 198-201 (2002).

[9] Allen, E., Triantaphillidou, S., and Jacobson, R. E., "Image Quality Comparison Between JPEG and JPEG2000. I. Psychophysical Investigation", Journal of Imaging Science and Technology, **51**(3), 248-258 (2007).

[10] Triantaphillidou, S., Allen, E., and Jacobson, R. E., "Image Quality Comparison Between JPEG and JPEG2000. II. Scene Dependency, Scene Analysis, and Classification", Journal of Imaging Science and Technology, **51**(3), 259-270 (2007).

[11] Burton, A. M., Wilson, S., Cowan, M., and Bruce, V., "Face Recognition in Poor-Quality Video: Evidence From Security Surveillance", Psychological Science, 10(3), 243-248 (1999).

[12] Bruce, V., Henderson, Z., and Newman, C., "Matching identities of familiar and unfamiliar faces caught on CCTV images", Journal of Experimental Psychology: Applied, 7(3), 207-218 (2001).

[13] Kemp, R., Towell, N., and Pike, G., "When Seeing should not be Believing: Photographs, Credit Cards and Fraud", Applied Cognitive Psychology, 11(3), 211-222 (1997).

[14] Davies, G. and Thasen, S., "Closed-circuit television: How effective an identification aid?", British Journal of Psychology, 91, 411-426 (2000).

[15] Hillstrom, A.P., Sauer, J., and Lorrain, H., "Training methods for facial image comparison: a literature review" Project Report. University of Portsmouth (2011).

[16] Kalva, H., "The H.264 Video Coding Standard", IEEE, 13(4): p. 86-90 (2006).

[17] Ghanbari, M., [Standard Codecs: Image compression to advanced video coding], IEE Telecommunications Series 49 (2003).

[18] Tsifouti, A., Nasralla, M., Razaak, M., Cope, J., Orwell, M. J., Martini, G. M., Sage, K., "A methodology to evaluate the effect of video compression on the performance of analytics system", Proc. SPIE 8546, 85460S-85460S (2012).

[19] ISO 12233:2000, Photography -- Electronic still-picture cameras -- Resolution measurements (2000).

[20] Poynton, C., [Digital video and HDTV, algorithms and interfaces] Elsevier Science (2003).

[21] ITU, Subjective video quality assessment methods for multimedia application, in ITU-R Recommendation P.910 (1999).

[22] Engeldrum, P.G., [Psychometric scaling: A toolkit for imaging systems development], Imcotek press (2000).

[23] ITU, Methodology for the subjective assessment of the quality of television pictures, in Recommendation ITU - R BT.500-11 (2002).

[24] Bell, S., "Measurement Good Practice Guide No.11 (Issue 2). A biginner's guide to uncertainty of measurement", National Physical Laboratory (1999).

[25] Keelan, B.W., [Handbook of Image Quality:Characterization and Prediction], Marcel Dekker Inc (2002).

[26] Maalouf, A., Larabi, M, C., and Nicholson, D., "Offline quality monitoring for legal evidence images in video-surveillance applications", Journal of Multimedia Tools and Applications, http://dx.doi.org/10.1007/s11042-012-1268-9, (2012).