CCTV as an automated sensor for firearms detection: human-derived performance as a precursor to automatic recognition

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

- Copyright 2008 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic electronic or print reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited. This paper can also be found at: http://dx.doi.org/10.1117/12.800264

Metadata Record: https://dspace.lboro.ac.uk/2134/6298

Version: Published

Publisher: © 2008 Society of Photo-Optical Instrumentation Engineers

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
CCTV as an automated sensor for firearms detection: human-derived performance as a precursor to automatic recognition

Iain T. Darker, Alastair G. Gale, Anastassia Blechko

Applied Vision Research Centre, Ergonomics & Safety Research Institute, Garendon Wing, Holywell Park, Loughborough University, Loughborough, Leicestershire, UK, LE11 3TU

ABSTRACT

CCTV operators are able to detect firearms, via CCTV, but their capacity for surveillance is limited. Thus, it is desirable to automate the monitoring of CCTV cameras for firearms using machine vision techniques. The abilities of CCTV operators to detect concealed and unconcealed firearms in CCTV footage were quantified within a signal detection framework. Additionally, the visual search strategies adopted by the CCTV operators were elicited and their efficacies indexed with respect to signal detection performance, separately for concealed and unconcealed firearms. Future work will automate effective, human visual search strategies using image processing algorithms.

Keywords: CCTV, firearm, signal detection, human performance, automation

1. INTRODUCTION

The present research is concerned with the detection of firearms carried on the person, via CCTV, in the course of proactive public space surveillance. In the UK, the possession of firearms is highly restricted and many types of firearm are prohibited altogether. Even where a firearm is possessed legitimately, it is illegal to carry it in a public place without legal authority or reasonable excuse; this applies to air weapons and imitation firearms, too. Thus, it is reasonable to bring to the attention of the authorities any firearm that is detected in a public place, in the UK.

Despite a backdrop of stringent firearms-related legislation, the carrying of firearms on the streets of the UK is a significant issue of increasing prominence [1]. Since 1981, the long-term trend has been for the level of gun crime in England and Wales to increase [2], reaching an all-time high in the year ending March 31st 2004 (although subsequent years saw reductions in the numbers firearms offences) [3]. Even so, according to official figures gun crime is a relatively small problem. In the year ending 31st March 2007 gun crime accounted for just 0.3 percent of all offences [3]. However, it is noted that firearms offences could be substantially underreported. The majority of firearms offences may occur within criminal circles [4], particularly in relation to gang membership or criminal activities like drug dealing [5]. The predilection for recourse to the law amongst these circles is likely to be low: a preference for personal retributive action, a fear of being labelled a grass, and an antipathy for the authorities prevent at least some from reporting crime [5, 6]. It is, therefore, difficult to obtain a reliable estimate of the level of gun crime in the UK, but an upper bound to the scale of the problem might be inferred by considering the number of firearms in circulation. The Small Arms Survey estimates that in the year 2000 there were 1,793,712 registered firearms and some 4,000,000 unregistered firearms in Britain, approximately one firearm for every ten people [7]. Thus, it appears that Britain hosts a considerable number of firearms, the majority of which may be possessed illegally, but all of which may find nefarious use. The figures become yet more concerning when it is considered that a single firearm can be used in multiple crimes over an extended period [5, 6].

The crux of the issue addressed by the present study is that firearms might be carried in public places with considerable frequency. Firearms offences can involve high-street business premises, street robbery, drug markets at street-level, and night clubs and other music venues [4-6]; whilst scenarios in which a firearm is carried routinely and casually, as a status symbol or as an accessory, have become evident in some British subcultures [5]. Given the vast numbers of firearms in circulation, this could place a substantial number of firearms in public spaces and the extent of this problem is likely to be both underestimated and on the increase.

The present study, which is part of the MEDUSA project [8], aims to increase the efficacy with which firearms that are carried in public places can be detected. The ultimate goal is to bring about a significant reduction in the levels of gun
crime. Public space surveillance that is mediated via CCTV networks offers a medium for the detection of firearms that are carried on the street. CCTV cameras are concentrated in areas where firearms are known to be carried such as town and city centres, and housing estates [9], whilst CCTV operators are able to predict [10] and proactively search out [11] lawless behaviour. Indeed, instances in which CCTV operators have been able to detect firearms in the course of the proactive monitoring of CCTV networks have been documented [12]. However, the capacity of CCTV operators for surveillance is limited: a single operator is required to monitor the live feeds from tens to hundreds of cameras [9]; the amount of time during which vigilance can be effectively maintained is limited to about an hour [13], whilst working hours far exceed this threshold [9]; the ability of a person to spot suspicious behaviour is fallible [10]; assignments of suspicion can be based on little more than personal prejudices [11]; and boredom or prurience may induce CCTV operators to indulge in activities outside of their job descriptions [11, 14]. With respect to gun crime, the problem may be exaggerated by the fact that incidents involving firearms are spotted with relative rarity in comparison to other events observed via CCTV. An analysis of all CCTV incidents logged between March 2006 and February 2008 at a CCTV control room which serves a large UK city (mid-year population estimate, 2006: 300,000 to one significant figure) revealed that of the 2253 incidents spotted through proactive monitoring, only 10 involved firearms [15]. Thus, CCTV operators have little opportunity to learn the visual indicators of gun crime in the course of their duties.

Intelligent, automatic CCTV systems can, potentially, circumvent some of the problems associated with human-mediated surveillance [16]. Consequently, it is desirable to automate the monitoring of CCTV cameras for firearms using machine vision techniques [8]. The experiment reported here addresses three questions which arise from this assertion. 1) Is it possible to use CCTV to detect firearms in a reliable manner? 2) What visual cues might support reliability in the CCTV-mediated detection of firearms? 3) How does concealment of the firearm influence the reliability of, and the visual cues used in, firearm detection? The present approach to answering these questions was to adopt techniques from the field of human, experimental psychology in order to determine how trained CCTV operators perform a CCTV-based firearm detection task. Any strategies found to be effective could then be used to inform the design of automatic, image processing algorithms for the detection of firearms.

CCTV operator performance in the detection of firearms, via CCTV, was assessed within a signal detection framework by means of a psychophysical experiment. The firearm was treated as a signal and two types of CCTV video clip were generated: signal absent (a noise only condition) and signal present (a noise plus signal condition). In order that the findings should reflect the detection of a firearm per se, and not simply the detection of an object, the signal absent condition involved an innocuous object matched to the firearm for approximate size, weight, and colour, whilst the signal present condition involved the replacement of this innocuous object with a firearm. The signal absent and signal present conditions were each subdivided into two further conditions: concealed and unconcealed; the object (a firearm or an innocuous object) was either concealed on a person, in or under their clothing, or the object was unconcealed and held in plain view. The influence of concealment was tested in a between-participants design. CCTV operators were required to determine whether the signal was present or absent in a given clip of CCTV footage, using a six-point scale of confidence which ranged from “definitely no” through to “definitely yes”. On the basis of the response frequency distributions across the confidence categories, for the signal present and signal absent conditions, it was possible to determine the receiver operator characteristic (ROC) of each CCTV operator [17]. Thus, it was possible to infer whether or not firearms can be detected reliably via CCTV, in the concealed and unconcealed conditions. After performing the firearm detection task, the frequencies with which various strategies were used by CCTV operators in the detection task were elicited through self-report. Frequency of use for each strategy was correlated with gun detection performance, across participants, to determine which strategies were associated with effective gun detection. On this basis of this information it was also possible to infer visual cues that are associated with reliable gun detection. Finally, these two analyses were performed independently in terms of concealed objects and unconcealed objects. It was, therefore, possible to determine how concealment influenced reliability and strategy in firearm detection. Ultimately, these factors will be considered in the design of image processing algorithms to automate the detection of firearms via CCTV.

Mock CCTV footage was used in the experiment. “Real-life” CCTV footage of gun crime in the UK is hard to come by, at least by means of retrospective collection. It has already been noted that firearms are rarely spotted by CCTV operators in the course of proactive surveillance [15]. Additionally, when an incident is spotted, it is not necessarily archived and any footage that is not archived is destroyed. For town centre CCTV systems the guideline is to retain CCTV footage for a month in order to allow incidents to come to light [18]. Between March 2006 and February 2008, the CCTV control room mentioned above archived CCTV footage of just one incident involving a firearm [15]; similarly, between 2003 and 2006, another CCTV control room which serves another large UK city (mid-year population estimate, 2006: 400,000 to one significant figure) archived CCTV footage of only five incidents involving a firearm [19].
The use of mock CCTV footage allowed a high degree of control over the experiment design with respect to the scenario portrayed and the presentation of the firearm. The scenario that was addressed involved an individual walking whilst carrying a firearm, either in plain view in the hand or concealed on their person. This scenario had been observed in a small sample of real-life, UK CCTV footage of gun crime [12] and was selected in consultation with CCTV operators and firearms experts from the UK National Firearms Centre. The intention was to reproduce a situation that might occur either in the course of casual, illicit gun carrying or prior to the commissioning of an offence.

2. METHOD

2.1 Participants

Nineteen CCTV operators took part in the experiment. Data from three participants were discarded as, for each of these participants, the detection model did not account for all the variance in the data; the recommendation is to reject such data [20]. The data of the remaining 16 participants were retained (all were engaged in public space surveillance, 14 in council-managed city or town centre CCTV rooms and the remainder at a university campus CCTV control room; 14 men; age range: 18 – 57 years, $M = 39, SD = 11$; range of years of experience as a CCTV operator: 0.33 – 19, $M = 5, SD = 5$; range of hours per shift spent monitoring cameras: 1 – 12, $M = 8, SD = 4$). Each participant was offered £10 for their time and after completing the experiment reported here, they also completed a questionnaire on CCTV camera handling and the use of CCTV in the proactive detection of crime.

2.2 Apparatus

When recording the mock CCTV footage, the analogue output of a CCTV camera was converted to a digital signal by a Canopus analogue to digital video converter (Grass Valley, version ADVC 110, Canopus Co. Ltd, 2004-2006). The digital signal was then transferred to a computer (Viglen Genie, Viglen Ltd.; Microsoft Windows XP version 2002, service pack 2; Intel® Xeon™ processor 3.20 GHz; 2.00 GB of RAM) via firewire where it was sampled at 25 frames per second and captured straight to hard drive in a compressed digital video format (compression 10:1, avi format) using DV Rack software by Serious Magic™ (2005, version 1.2.1490.0). The video clip files were named during live recording, from a filming schedule, in a systemic fashion which allowed the exact experiment conditions contained in the video clip to be identified in the filename.

For the gun detection experiment, stimulus presentation and data collection were automated using a laptop computer (Toshiba Tecra, Toshiba Ltd.; Microsoft Windows XP version 2002, service pack 2; Genuine Intel® Centrino Duo™ T2600 processor 2.16 GHz; 1.00 GB of RAM) controlled by a program developed in-house (using Microsoft Visual Studio .net 2005, Visual Basic).

2.3 Stimuli

The stimuli were 2-second video clips of individual actors (hereafter referred to as surveillance targets) walking along a walkway against a neutral backdrop (a breezeblock wall). In a number of video clips the surveillance target was required to carry an object: either a firearm or an innocuous object matched to the firearm for approximate size, weight, and colour (a plastic drinks bottle filled with a dark-coloured liquid). The firearm or bottle was either held in plain view in the hand, or it was concealed on their person, in or under their clothing. In either case the object was positioned between the surveillance target’s body and the camera used to record the footage. The camera was positioned so to produce a view that might be obtained from a CCTV camera: an elevated, three-quarter angle view of the walkway. The camera was positioned at a height of 2.8 metres, approximately 9.8 metres from the centre of the walkway along an axis subtended at an angle of 50° to the backdrop of the walkway. The camera view was panned out to show 13.5 metres of walkway, although during the 2-second clip surveillance targets appeared only within the central third of the walkway. The clip was cut to show 1 second of walking either side of the centre point of the walkway in order to capture the surveillance targets in full stride. Seven hundred and ninety two video clips were used in the experiment: 72 clips of people carrying unconcealed firearms; 72 clips of people carrying concealed firearms; 216 clips of people carrying unconcealed bottles; 216 clips of people carrying concealed bottles; 216 clips of people carrying neither a firearm nor a bottle. Twelve different surveillance targets were filmed to produce these video clips and the numbers of each type of video clip were distributed equally across these 12 surveillance targets. The clips involved three sizes of firearm (a .32 calibre revolver, a Glock automatic pistol, and a sawn-off shotgun representing small, medium, and large firearms respectively) matched to three sizes and weights of bottle. Surveillance targets were filmed walking towards and away
from the camera to produce ventral and dorsal views of the surveillance target, respectively. These variations were distributed equally across surveillance targets (1 to 12), concealment conditions (concealed, unconcealed), and views of the surveillance target (ventral, dorsal), and in the appropriate proportions across object types (ratio: 1 firearm: 3 bottle: 3 no additional object) at each object size (small, medium, large). At a viewing distance of 450mm each surveillance target subtended approximately 4° visual angle and the firearms and bottles subtended between 0.1° and 0.5° visual angle, depending on type. An example video clip involving a ventral view of a surveillance target walking whilst carrying an unconcealed firearm (the Glock automatic pistol) is featured in Video 1.

![Video 1. Stimulus video clip of a surveillance target walking towards the camera whilst carry an unconcealed firearm.](http://dx.doi.org/10.1117/12.800264.1)

2.4 Design

The experiment was designed to facilitate analyses of both gun detection performance and detection strategy, as a function of whether the firearm was concealed or not. The influence of concealment was tested in a between-participants design. Participants were pseudo-randomly assigned to view either concealed firearms and bottles, or unconcealed firearms and bottles with the stipulation that, of those participants whose data fitted the signal detection model, half would participate in the concealed condition and half would participate in the unconcealed condition. Thus, of the 16 participants whose data fitted the signal detection model, 8 viewed the concealed condition and 8 viewed the unconcealed condition.

Each participant saw 504 experimental video clip stimuli which were determined by the concealment condition to which the participant had been assigned: Concealed: 72 clips of people carrying concealed firearms, 216 clips of people carrying concealed bottles, 216 clips of people carrying no additional object (neither a firearm nor a bottle); Unconcealed: 72 clips of people carrying unconcealed firearms, 216 clips of people carrying unconcealed bottles, 216 clips of people carrying no additional object (neither a firearm nor a bottle).
2.5 Procedure

CCTV operators were recruited by contacting CCTV control room managers and through speaking about the study at CCTV practitioner conferences, seminars, and meetings. Participants gave informed consent to participate in the study. They then completed a questionnaire to record demographic information and to elicit information about their current role, length of service, work hours, degree of proactive monitoring of CCTV cameras, work history, and CCTV-related training.

After completing the questionnaire, each participant sat a ten-minute video-based test to assess their ability to read body language (findings relating to this test are not reported here). Each participant then undertook the firearm detection experiment which took about three-quarters of an hour to complete. This involved the presentation of 8 practice video clips followed by the presentation of 504 experimental video clips. Participants were required to view each video clip and then, after each clip, they were prompted for a response: Was a firearm present in the video clip that you have just viewed? The response was made on a six point scale which indicated the participant’s confidence regarding whether or not a firearm was present in the clip: Definitely no; Probably no; At a guess, no; At a guess, yes; Probably yes; Definitely yes. The response was recorded automatically and the next video clip was initiated when the response had been submitted.

The video clips were presented in a pseudo-randomised order such that they appeared in cycles of 42 clips. The cycles were designed to balance the presentation of video clips across the experiment with respect to type of object (firearm, bottle, none), size of object (small, medium, large), and view of the surveillance target (ventral, dorsal). Each cycle of 42 clips involved a ventral clip of each of the three sizes of firearm, a dorsal clip of each of the three sizes of firearm, three ventral clips of each of the three sizes of bottle, three dorsal clips of each of the three sizes of bottle, nine ventral clips involving neither a gun nor a bottle, and nine dorsal clips involving neither a gun nor a bottle. The clips that were presented in each cycle were selected at random with the stipulation that they had not been presented previously to that participant. The order of the clips within a cycle of 42 clips was also pseudo-randomised such that every 14 clips a participant saw one ventral clip involving a firearm, one dorsal clip involving a firearm, three ventral clips involving a bottle, three dorsal clips involving a bottle, three ventral clips involving neither a firearm nor a bottle, and three dorsal clips involving neither a firearm nor a bottle. The order of the video clips was randomised within this sub-cycle of 14 clips. A different pseudo-randomised order was generated for each participant using the method described.

It was necessary to employ these pseudo-randomising algorithms because different numbers of clips contributed to different signal detection conditions and the stimulus presented was generated on the basis that that stimulus should not have been presented previously to that participant. If simple, random selection had been used, the laws of probability would have dictated that more common types of clip (the bottle and no additional object clips) would have been more likely to be selected for presentation in the early stages of the experiment, with the likelihood of selecting a firearm clip for presentation not achieving equality with the other types of clip until the later stages of the experiment. The pseudo-randomisation procedure prevented such bias and also balanced the potential influences of fatigue and other zeitgeist-related factors evenly across the experimental conditions.

After the experiment, participants completed a questionnaire regarding the strategies that they had used to decide whether or not a firearm had been present in the video clips. They then completed a questionnaire regarding CCTV camera handling and the use of CCTV to proactively detect crime.

2.6 The measurement of gun detection performance

Gun detection performance was analysed within a signal detection framework [17] and indexed using ROC parameters. The signal detection analysis measures the ability of a sensor to detect a signal against a background of noise. In this instance the sensor is the human visual system and the signal is a firearm (and the firearm’s influence upon the person carrying it). The noise condition tested here involved an innocuous object (and the innocuous object’s influence upon the person carrying it). The innocuous object was a plastic drinks bottle which was matched to the firearm for approximate size, weight, and colour.

Estimates of ROC parameters were obtained by means of a maximum likelihood algorithm (RSCORE 5.5.7) based on a Gaussian probability distribution [20]. The measure of detection sensitivity adopted for the present analysis was $d_a$, which is a standardised ($z$-score based) measure of the distance between the means of the noise and noise plus signal response frequency distributions over a range of confidence levels relating to confidence that signal is present or absent.
The ROC parameter $d_{a}$ offers a direct index of sensitivity to the signal, it is appropriate when the two distributions are of unequal-variance, and it is numerically equal to $d'$ when the two distributions are of equal variance.

### 2.7 The assessment of gun detection strategy

Detection strategy was assessed after participants had viewed all 504 experimental video clips in the concealment condition to which they had been assigned. Participants were asked: In the video clips that you have just seen, how did you decide whether or not a person was carrying a gun? In response to this question they were asked to rate, on a 5-point scale, a series of statements about potential task strategies with respect to the frequency with which they had employed that strategy: 1) not at all; 2) very little; 3) sometimes; 4) a lot; 5) all the time. The strategy statements were derived from feedback to a pilot study ($n = 8$) which employed the same video clips as used here, as well as on the basis of literature reviews regarding the use of CCTV, surveillance and inspection strategies, and the influences of firearms on behaviour [8, 12].

In the concealed condition, the strategies used by participants were elicited by asking them to rate, on the 5-point scale, a series of statements which were tailored to the concealed condition: I guessed; I looked at how they walked; I looked at how fast they were walking; I looked at the length of their stride; I looked at how they swung their arms; I looked at the expression on their face; I looked at their posture; I looked at their demeanour or overall behaviour; I looked at the shadow on the wall; I looked at how their clothing hung; I looked at how their clothing moved; I looked for a bulge in their clothing; I looked for a gun sized bulge in their clothing; I looked for a gun shaped bulge in their clothing; I looked at where they put their hands; I looked at whether their hands were hidden or not; I looked for possible hiding places in their clothing.

In the unconcealed condition, the strategies used by participants were elicited by asking them to rate, on the 5-point scale, a series of statements which were tailored to the unconcealed condition: I guessed; I looked at how they walked; I looked at how fast they were walking; I looked at the length of their stride; I looked at how they swung their arms; I looked at the expression on their face; I looked at their posture; I looked at their demeanour or overall behaviour; I looked at the shadow on the wall; I looked at how their clothing hung; I looked at how their clothing moved; I looked for a gun sized object in the hand; I looked for a gun shaped object in the hand; I looked for a gun coloured object in the hand; I looked at how they held the object in their hand; I looked at how the object moved in their hand; I looked at how light reflected off of the object in their hand.

### 3. RESULTS

An alpha-level of .05 was adopted as the threshold for significance (* = significant). The data were first subjected to tests of normality and homogeneity of variance. Then tests were performed to determine whether the CCTV operators in the concealed and unconcealed viewing groups were matched in terms of age, years of experience as a CCTV operator, and hours per shift spent proactively monitoring CCTV cameras. Subsequently, the sensitivities to concealed and unconcealed firearms detection were assessed and compared on the basis of the ROC parameter, $d_{a}$. Then, the frequency of use ratings for each gun detection strategy were compared within each concealment condition and were correlated with gun detection sensitivity ($d_{a}$) for each concealment condition.

#### 3.1 Tests of normality and homogeneity of variance

Shapiro-Wilk tests were conducted to assess the normality of the data distributions within each level of the concealment factor. Levine tests were conducted to assess homogeneity of variance in the data, across the two levels of the concealment factor. The data relating to the ages of the participants and sensitivity to the firearm ($d_{a}$) exhibited normality and homogeneity of variance across both levels of the concealment factor and so were subsequently subjected to parametric analyses (Shapiro-Wilk tests: age, concealed, $W(8) = 0.932, p = .530$; age, unconcealed $W(8) = 0.955, p = .766$; $d_{a}$, concealed, $W(8) = 0.916, p = .398$; $d_{a}$, unconcealed, $W(8) = 0.940, p = .615$; Levine tests: age, $F(1, 14) = 1.14, p = .304$; $d_{a}$, $F(1, 14) = 4.32, p = .057$). The data relating to years of experience as a CCTV operator did not exhibit normality in the concealed level of the concealment factor, whilst hours spent proactively monitoring CCTV cameras per shift did not exhibit normality in the unconcealed level of the concealment factor (Shapiro-Wilk tests: years of experience, concealed, $W(8) = 0.903, p = .307$; years of experience, unconcealed $W(8) = 0.805, p = .032$*; hours monitoring, concealed, $W(8) = 0.765, p = .012$*; hours monitoring, unconcealed, $W(8) = 0.897, p = .273$; Levine tests: years of experience, $F(1, 14) = 3.84, p = .070$; hours monitoring, $F(1, 14) = 0.41, p = .534$). Consequently, years of
experience as a CCTV operator and hours spent proactively monitoring CCTV cameras per shift were subsequently subjected to nonparametric analyses.

3.2 Tests of differences between the concealment groups in terms of age, years of experience as a CCTV operator, and hours per shift spent proactively monitoring CCTV cameras

The CCTV operators who saw concealed (c) objects and the CCTV operators who saw unconcealed (u) objects did not differ significantly in terms of age ($M_c = 37.13$, $SD_c = 9.60$ vs. $M_u = 40.25$, $SD_u = 13.40$; ANOVA: $F(1, 14) = 0.29$, $MSE = 39.06$, $p = .600$, $\eta^2 = 0.02$), years of experience as a CCTV operator ($Md_{nc} = 5.00$ vs. $Md_{nu} = 2.25$; Mann-Whitney test: $U = 27.00$, $p = .627$, $r = -0.13$), or hours per shift spent proactively monitoring CCTV cameras ($Md_{nc} = 9.25$ vs. $Md_{nu} = 7.75$; $U = 32.00$, $p = 1.000$, $r = 0.00$).

3.3 The influence of concealment on sensitivity to the firearm

Sensitivity ($d_a$) to the firearm when concealed was significantly lower than sensitivity to the firearm when unconcealed ($M_c = -0.31$, $SD_c = 0.16$ vs. $M_u = 2.59$, $SD_u = 0.36$; ANOVA: $F(1, 14) = 430.83$, $MSE = 33.65$, $p = .000^*$, $\eta^2 = 0.97$). (See Figure 1.)

![Fig. 1 Mean sensitivity ($d_a$) to firearms in the concealed and unconcealed conditions.](image)

3.4 Concealed: frequency of use ratings by strategy, and correlations between frequency of use ratings and sensitivity to the firearm

The median frequency of use ratings assigned to strategies in the concealed condition were subjected to a Friedman’s ANOVA and it was found that frequency of use rating varied significantly by strategy ($\chi^2(16) = 49.45$, $p = .000$). The critical difference between the mean ranks of the Friedman’s ANOVA was calculated [21]. This critical difference was used as a basis for defining the most and least frequently used strategies. The most frequently used strategies were defined as those whose mean rank differed significantly from the lowest mean rank. The least frequently used strategies were defined as those whose mean rank differed significantly from the highest mean rank. The most frequently used strategies were to look for a bulge in clothing, to look at the location of the hands, to look at whether the hands were hidden or not, and to look at how the clothing hung. The least frequently used strategies were to look at the length of stride and to look at facial expression. (See Table 1, the most frequently used strategies appear in italics above the upper dashed line, the least frequently used strategies appear in italics below the lower dashed line.)
Additionally, for each strategy the frequency of use ratings were subjected to a Spearman’s correlation against sensitivity to the firearm, across participants. There was a significant, positive correlation between frequency of looking at how clothing hung and sensitivity to the concealed firearm. Also, a positive correlation between frequency of looking for a bulge in the clothing and sensitivity to the concealed firearm tended towards significance and given the degree of correlation \( r_s = .63 \) this relationship may have achieved statistical significance in a larger sample. (These correlations are summarized by strategy, in Table 1.)

Table. 1. Mean Friedman ranks, median and range of frequency of use ratings, and Spearman’s correlations between frequency of use ratings and sensitivity to the firearm for each strategy, in the concealed condition.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Friedman's ANOVA</th>
<th>Frequency of use rating</th>
<th>Spearman's correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean rank</td>
<td>Median</td>
<td>Range</td>
</tr>
<tr>
<td>Bulge in clothing</td>
<td>14.19</td>
<td>5.0</td>
<td>4 to 5</td>
</tr>
<tr>
<td>Hands, location</td>
<td>12.25</td>
<td>4.5</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Hands, hidden</td>
<td>12.25</td>
<td>4.0</td>
<td>3 to 5</td>
</tr>
<tr>
<td>Clothing, hang</td>
<td>11.56</td>
<td>4.5</td>
<td>3 to 5</td>
</tr>
<tr>
<td>Walk</td>
<td>10.75</td>
<td>4.0</td>
<td>3 to 5</td>
</tr>
<tr>
<td>Bulge, gun shaped</td>
<td>10.25</td>
<td>4.0</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Demeanour</td>
<td>10.13</td>
<td>4.0</td>
<td>1 to 5</td>
</tr>
<tr>
<td>Clothing, movement</td>
<td>9.88</td>
<td>4.0</td>
<td>1 to 5</td>
</tr>
<tr>
<td>Bulge, gun sized</td>
<td>9.88</td>
<td>4.0</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Speed of walk</td>
<td>8.81</td>
<td>3.5</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Posture</td>
<td>8.31</td>
<td>3.0</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Arm swing</td>
<td>7.19</td>
<td>3.0</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Clothing, hiding places</td>
<td>7.19</td>
<td>3.5</td>
<td>1 to 4</td>
</tr>
<tr>
<td>Guesswork</td>
<td>6.88</td>
<td>3.0</td>
<td>1 to 4</td>
</tr>
<tr>
<td>Shadow on the wall</td>
<td>6.38</td>
<td>2.5</td>
<td>1 to 5</td>
</tr>
<tr>
<td>Length of stride</td>
<td>4.38</td>
<td>2.0</td>
<td>1 to 4</td>
</tr>
<tr>
<td>Facial expression</td>
<td>2.75</td>
<td>1.5</td>
<td>1 to 4</td>
</tr>
</tbody>
</table>

3.5 Unconcealed: frequency of use ratings by strategy, and correlations between frequency of use ratings and sensitivity to the firearm

The median frequency of use ratings assigned to strategies in the unconcealed condition were also subjected to a Friedman’s ANOVA and it was found that frequency of use rating varied significantly by strategy \( \chi^2(16) = 60.02, p = .000 \). Again, the critical difference between the mean ranks in the Friedman’s ANOVA was calculated [21]. The critical difference was used to define the most and least frequently used strategies in the same manner as described for the concealed condition (see section 3.4). The most frequently used strategies were to look for a gun shaped object in the hand and to look for a gun sized object in the hand. The least frequently used strategies were to use guesswork, to look at the length of stride, and to look at facial expression. (See Table 2, the most frequently used strategies appear in italics above the upper dashed line, the least frequently used strategies appear in italics below the lower dashed line.)

Again, for each strategy the frequency of use ratings were subjected to a Spearman’s correlation against sensitivity to the firearm, across participants. There was a significant, positive correlation between frequency of looking at demeanour or overall behaviour and sensitivity to the unconcealed firearm. There was also a significant, positive correlation between frequency of looking at demeanour or overall behaviour and sensitivity to the unconcealed firearm. (These correlations are summarized by strategy, in Table 2.)
Table 2. Mean ranks, median and range of frequency of use ratings, and Spearman’s correlations between frequency of use ratings and sensitivity to the firearm for each strategy, in the unconcealed condition.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Friedman’s ANOVA Mean rank</th>
<th>Frequency of use rating</th>
<th>Spearman’s correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median</td>
<td>Range</td>
</tr>
<tr>
<td>Object, gun shaped</td>
<td>14.56</td>
<td>5.0</td>
<td>4 to 5</td>
</tr>
<tr>
<td>Object, gun sized</td>
<td>13.13</td>
<td>5.0</td>
<td>4 to 5</td>
</tr>
<tr>
<td>Object, light reflection</td>
<td>12.06</td>
<td>5.0</td>
<td>3 to 5</td>
</tr>
<tr>
<td>Object, way held</td>
<td>11.63</td>
<td>4.0</td>
<td>4 to 5</td>
</tr>
<tr>
<td>Object, way moved</td>
<td>11.31</td>
<td>4.5</td>
<td>1 to 5</td>
</tr>
<tr>
<td>Walk</td>
<td>10.56</td>
<td>4.0</td>
<td>3 to 5</td>
</tr>
<tr>
<td>Arm swing</td>
<td>10.56</td>
<td>4.0</td>
<td>3 to 5</td>
</tr>
<tr>
<td>Clothing, hung</td>
<td>9.81</td>
<td>3.5</td>
<td>1 to 5</td>
</tr>
<tr>
<td>Posture</td>
<td>9.06</td>
<td>4.0</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Clothing, movement</td>
<td>8.50</td>
<td>3.0</td>
<td>1 to 5</td>
</tr>
<tr>
<td>Object, gun coloured</td>
<td>8.44</td>
<td>4.0</td>
<td>1 to 5</td>
</tr>
<tr>
<td>Speed of walk</td>
<td>8.38</td>
<td>4.0</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Demeanour</td>
<td>7.13</td>
<td>3.5</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Shadow on the wall</td>
<td>6.31</td>
<td>2.5</td>
<td>1 to 5</td>
</tr>
<tr>
<td>Guesswork</td>
<td>4.13</td>
<td>2.0</td>
<td>1 to 3</td>
</tr>
<tr>
<td>Length of stride</td>
<td>3.88</td>
<td>2.5</td>
<td>1 to 4</td>
</tr>
<tr>
<td>Facial expression</td>
<td>3.56</td>
<td>1.5</td>
<td>1 to 4</td>
</tr>
</tbody>
</table>

Critical difference 8.84 d_s (range) 1.92 to 3.02

4. DISCUSSION

The CCTV operators who viewed video clips involving concealed objects and those who viewed video clips involving unconcealed objects did not differ in terms of a number of factors which might influence surveillance ability (age, years of experience as a CCTV operator, hours per shift spent proactively monitoring CCTV cameras). Consequently, it is inferred that differences between the two groups reflect the influence of concealment on firearm detection.

Perhaps unsurprisingly, sensitivity to a concealed firearm was significantly lower than sensitivity to an unconcealed firearm. Further, sensitivity to a concealed firearm was below zero, indicating that ability to distinguish between a concealed firearm and a concealed bottle was below chance. This might reflect a tendency to mistake a concealed bottle for a concealed firearm. Consequently, the most frequently used strategies in the concealed condition are unlikely to provide a sound basis for an algorithm to support the automatic detection of a concealed firearm. Similarly, it cannot be inferred that those strategies whose frequency of use correlated positively with sensitivity to concealed firearms can provide the basis of effective, automatic image processing algorithms for the detection of concealed firearms. This is because the range of sensitivities over which these correlations occurred was entirely below zero. Thus, although the frequency with which a CCTV operator looked at how clothing hung exhibited a significant, positive correlation with sensitivity to the concealed firearm, it cannot be inferred that increasing the frequency with which this strategy is used will support high sensitivity in the detection of concealed firearms. It appears that CCTV operators cannot be expected to detect concealed firearms, at least under the present experimental conditions. Thus, there is no evidence to suggest that the strategies explored here would provide a useful basis for the design of effective, automatic image processing algorithms for concealed firearm detection.

Sensitivity to an unconcealed firearm was exceedingly high (the mean sensitivity was 2.59 standard deviations above zero sensitivity). This suggests that CCTV operators have an excellent ability to distinguish between a bottle and a firearm, even in CCTV footage where the surveillance target is small, the object of interest is smaller still, and lighting...
conditions are less than ideal (see Video 1). Thus, it might be inferred that the strategies frequently used in detecting an unconcealed firearm could form the basis of an effective image processing algorithm for the automatic detection of such firearms. The most frequently used strategies were to look for a gun sized or gun shaped object in the hand. Every participant who viewed the video clips featuring unconcealed objects used these strategies either all the time or most of the time. It is noted that neither of the most frequently used strategies exhibited significant correlations with sensitivity to the firearm. However, it is likely that the limited range and variation amongst frequency of use ratings for these strategies reduced the opportunity for a significant correlation to arise. In contrast, frequency of looking for a gun coloured object in the hand did correlate significantly and positively with sensitivity to the firearm; this strategy had frequency of use ratings across the full range of permitted responses. Thus, it might be inferred that concentrating on the visible, physical properties of the object in the hand will be most effective in the detection of unconcealed firearms. This statement might be considered a truism in the context of the experiment’s design, but it warrants discussion as the strategy has considerable validity in “real-life” situations. This issue is addressed in more depth in the General Discussion (section 5).

It is also noted that there was a significant, positive correlation between frequency of looking at the demeanour or overall behaviour of the surveillance target and the CCTV operator’s sensitivity to the unconcealed firearm. This is somewhat unexpected given that other strategies which might rely on interpreting the surveillance target’s behaviour did not correlate significantly with sensitivity. Looking at demeanour or overall behaviour received a wide range of responses in terms of frequency of use, a factor which increases the potential for a correlation, but this was also true of other more specific, behaviour-related strategies (for instance, looking at facial expression or looking at posture). It is possible that CCTV operators were not consciously aware of the specific visual cues that informed their decision-making and so ascribed this “intuition” to use of demeanour or overall behaviour. Alternatively, a tendency to report a high frequency of use for demeanour or overall behaviour could reflect certain qualities in the CCTV operator rather than any visual elements in the video clip: the correlation indicates an association between the two factors rather than a direct, causal relationship.

If the use of demeanour or overall behaviour is subserved by visual elements in the video clip, one explanation of the correlation between sensitivity to the firearm and frequency of use of demeanour or overall behaviour might be that the CCTV operator is reading the surveillance target’s body language in relation to a phenomenon known as the weapons effect [22]. Weapons are thought to prime aggressive behaviour via semantic association [23, 24]. The resultant behaviour patterns may be stimulus and context specific, stereotyped, and automatic to some degree [22, 25]. Consequently, the surveillance targets who appeared in the video clips may reveal that they are carrying a firearm through aggression in their body language, as an automatic consequence of contact with the firearm. Indeed, the moods of the surveillance targets whilst they carried the firearms and bottles were assessed within a counterbalanced, experiment design and it was found that the surveillance targets experienced higher levels of aggression-related affect whilst carrying a firearm than whilst carrying a bottle [26]. However, this raises a question: why was behaviour or overall demeanour not useful when the firearm was concealed? Perhaps, when the firearm was unconcealed, aggression-related cues in the surveillance target merely supplemented (possibly at a subconscious level) the potentially more dominant visual cues related to the shape, size, and colour of the firearm. In contrast, when the firearm was concealed and there might have been an absence of dominant visual cues, CCTV operators may have been reluctant to place a large emphasis on what might amount to no more than a gut feeling arising from an interpretation of the body language of the surveillance target.

Thus, a number of visual cues have been associated with efficacy in the detection of an unconcealed firearm via CCTV: tangible visual cues related to the size, shape, and colour of an object might be readily incorporated into image processing algorithms, whilst cues related to demeanour or overall behaviour require further clarification. However, the possibility that various visual cues interact evokes a further consideration when designing image processing algorithms on the basis of the present study. If a given cue is not implemented within the observed combination of other cues then its efficacy may be altered. Indeed, some seemingly highly effective cues might become relatively ineffective if not used alongside certain other cues. Because the present analysis is based on a series of correlations, it does not allow a determination of the degree to which a given visual cue contributes to firearms detection, alone. A regression analysis, which might have teased out the separate contributions of individual strategies, would not be appropriate in the present context as the individual strategies are not necessarily independent (and because the sample size was limited relative to the number of strategies of interest). The issue of interactions amongst strategies might be addressed through machine vision trials that implement potentially useful cues individually and in each possible combination.
5. GENERAL DISCUSSION

It has been determined that human-derived performance can inform the design of effective, automatic image processing algorithms for the detection of unconcealed firearms. However, it has not been possible to derive effective strategies for the detection of concealed firearms. Effective strategies for the detection of unconcealed firearms relate to the size, shape, and colour of an object held in the hand, and to the demeanour or overall behaviour of the person carrying the object. The nature and use of cues related to demeanour or overall behaviour require further investigation, but the cues relating to the directly visible, physical properties of the firearm may be more tangible.

It is suggested that a machine vision based system for the automatic detection of unconcealed firearms should engage in localisation of the area of interest, object detection, and object recognition. This might involve locating the hands of a human surveillance target, detecting an object held in the hand, and then attempting to recognise whether or not the object is a firearm. When defined and verified, cues relating to the target’s overall demeanour or behaviour might be used in conjunction with the directly visible, physical properties of the object to decrease the probability of a false alarm. Additionally, cues relating to demeanour or overall behaviour, and possibly other situational cues not present in the mock CCTV footage used here, might also aid in the selection of surveillance targets that are most likely to be carrying firearms.

The approach of looking for a firearm in the hand has real-world validity. It has already been noted that certain groups carry firearms on the street, routinely and casually [5]. Further, instances where a firearm has been held in plain view, in public places, whilst being carried in a casual manner (as opposed to in the course of commissioning an offence) have been observed in CCTV footage recorded at UK CCTV control rooms [12]. Additionally, in the analysis of proactively spotted CCTV incidents logged over a two-year period at a UK CCTV control room (see section 1), of the 10 incidents involving firearms, in 9 instances the firearm was spotted in plain view and in the remaining instance a known individual suspected of owning an illicit firearm was recognised [15].

However, as the present study is based on mock CCTV footage it might lack real-world validity in certain respects. Real-life CCTV footage is likely to feature a wealth of cues that are not captured in the mock CCTV footage: behavioural cues related to the intent to use the firearm in the commission of an offence; situational and demographic cues related to times, places, and persons associated with firearms; known individuals; and group and crowd behaviours. An agreement has been reached with a large, UK police force and a council-managed CCTV control for the prospective collection of real-life CCTV footage involving firearms. This real-life CCTV footage will be fed into a similar study to the one described here. A follow-up experiment based on real-life CCTV footage involving firearms will, hopefully, help assess the validity of the visual cues defined here, determine further cues, and facilitate a test of the final automatic firearm detection system. This follow-up experiment may also employ an eye-tracker to record the loci of attention throughout inspection of the video clips. Thus, inspection strategy can be determined empirically as well as through self-report and any subconscious elements of the inspection strategies can be inferred by comparing the empirical data with the self-report data. Such a comparison might help illuminate any subconscious visual cues that contribute to the use of demeanour or overall behaviour in firearm detection.

In summary, the present study is able to inform the design of image processing algorithms for the automatic detection of unconcealed firearms, via CCTV. Although the present study has relied upon mock CCTV footage, there are indications that the inspection strategies that appeared effective in the present context also have real-world validity.

ACKNOWLEDGMENTS

The MEDUSA project is funded by the EPSRC (grant number: EP/D078105/1). We would like to thank Mark Murray-Flutter of the Royal Armouries, Leeds, UK, and Ed Wallace of LGC Forensics, Leeds, UK, for consulting on and facilitating the use of real firearms. We would also like to thank the Leicestershire Constabulary for providing safe, overnight storage for the firearms used in filming the mock CCTV footage.
REFERENCES


