SCIENTIFIC REPORTS

OPEN

Person identification from aerial footage by a remote-controlled drone

Received: 3 May 2017
Accepted: 5 October 2017
Published online: 19 October 2017

Markus Bindemann, Matthew C. Fysh, Sophie S. K. Sage, Kristina Douglas & Hannah M. Tummon

Remote-controlled aerial drones (or unmanned aerial vehicles; UAVs) are employed for surveillance by the military and police, which suggests that drone-captured footage might provide sufficient information for person identification. This study demonstrates that person identification from drone-captured images is poor when targets are unfamiliar (Experiment 1), when targets are familiar and the number of possible identities is restricted by context (Experiment 2), and when moving footage is employed (Experiment 3). Person information such as sex, race and age is also difficult to access from drone-captured footage (Experiment 4). These findings suggest that such footage provides a particularly poor medium for person identification. This is likely to reflect the sub-optimal quality of such footage, which is subject to factors such as the height and velocity at which drones fly, viewing distance, unfavourable vantage points, and ambient conditions.

Unmanned aerial vehicles (UAVs), commonly referred to as drones, are increasingly utilised by police and the military. In the UK, for example, a key application of drones by police organisations is to assist in searches for missing persons[1], as well as crowd control[2]. In addition, drones are routinely used by the military in operations such as reconnaissance, target acquisition, and to carry out lethal strikes[3,4]. These deployment strategies imply that drone-captured footage provides sufficient information for person identification. However, due to the variable height and velocity at which drones fly, such footage is likely to be subject to momentum, unfavourable vantage points, and unpredictable ambient conditions. For example, military drones operate from ground level up to maximum altitudes of 200 ft for micro drones[4], which are small tactical drones of up to 2 kg in weight, and up to 45,000 ft for large drones[3,4] that comprise unmanned long-endurance aircraft of over 600 kg. Moreover, the ground speed at which these drones operate varies considerably, from 0–250 kts[3,5]. In addition, drones employed in police operations record surveillance footage whilst operating at altitudes ranging from ground level to up to 400 ft, and at speeds of up to 38 kts[6]. This range in operational parameters raises the possibility that drone-captured footage can be of sub-optimal quality for person identification.

The current study reports four experiments that investigate this issue, by examining the accuracy of person identification from drone-captured footage of a football (soccer) match at a UK university. This set up presents a natural scenario that should provide relatively favourable conditions for image capture and subsequent person recognition. We recorded such footage with a commercially available remote-controlled drone, with a minimum take-off weight (MTOW) of 300 g. As classified by NATO regulation, this type of drone falls into Class I(b)[7] and is therefore comparable to micro drones in use by the UK military[3] and police force[6]. We present four experiments that utilised the footage recorded with this drone to examine the accuracy of person identification from such surveillance material.

To our knowledge, these experiments represent the first systematic investigation of person identification by human observers from aerial footage recorded by a remote-controlled drone. By contrast, a compelling body of research already exists on person identification in other applied settings, such as passport control[8,9], closed-circuit television (CCTV)[10,11], and eyewitness scenarios[12,13]. This research demonstrates that *familiar* people, who are known to an observer, can be identified with good accuracy[14,15]. This is found under challenging conditions, for example, when people are viewed in poor-quality surveillance footage[11], or heavily degraded video[16], or when they are only seen briefly[17], partially[18,19], or in unfavourable non-frontal views[20].

School of Psychology, University of Kent, Canterbury, UK. Correspondence and requests for materials should be addressed to M.B. (email: m.bindemann@kent.ac.uk)

This reliable recognition of familiar people is held to be based on sophisticated cognitive representations that build up through substantial exposure to a person's face across a range of ambient conditions[21,22]. Such experience enables the extraction of the stable visual characteristics of an individual's identity, and for the dissociation of this information from ambient factors that interact with a person's appearance, such as variation in lighting or viewing direction[23]. The exact nature of these representations remains under investigation, but might reflect a cognitive "average" of the encounters with a face[22,24], with dimensions that capture the different ways in which a person's appearance can vary around such an average[21,25]. Such approaches view familiarity as a continuum, from unknown to well-known faces. Consequently, whether a specific point exists on this continuum at which faces can be defined as "familiar" is an open question. What is clear, however, is that when the cognitive representations of familiar faces are firmly established, these allow for recognition to generalise across a broad range of conditions, and to succeed even with very limited visual information[11,18,19].

By contrast, the identification of unknown or unfamiliar people, of whom an observer has no prior experience, is error-prone, even under seemingly good conditions. For example, when observers try to identify a target from a ten-face array, accuracy is only at 70%[10,26]. This is found with high-quality images that depict people in a frontal view, with a neutral expression, and under good lighting. Performance remains poor when this task is reduced to a 1-to-1 comparison[27,28], or when observers match a live person to their photo[29,30], or moving video images[31]. This difficulty of unfamiliar person identification reflects the fact that, without extensive prior exposure, observers can only have limited information about how a person can vary naturally in their appearance. Consequently, attempts to identify an unfamiliar person have to rely on unsophisticated image-comparison techniques. This issue is illustrated by the fact that unfamiliar face identification is trivial across identical images[21,22], but becomes more error-prone as variability in a person's appearance increases across to-be-compared images[32,33]. Similarly, accuracy declines when lighting or viewing angle are variable across images[34], or image resolution is poor[35]. Considering that drone-captured footage is restricted by such factors, the question arises also of the extent to which unfamiliar people can be identified from such material. In this study, we investigate these questions across several tasks to examine the identification of unfamiliar (Experiment 1) and familiar people (Experiment 2 and 3), as well as the perception of a person's sex, race, and age from drone-captured footage (Experiment 4).

Experiment 1

In this experiment, observers were presented with arrays comprising a high-quality face photograph and drone-captured images of a person, and had to decide whether these materials depicted the same person or two different people. Such identity-matching tasks have been used extensively in forensic face identification[36,37], and minimize the contribution of other factors, such as memory demands, that can reduce performance[38]. In light of the expected difficulty of this task, three drone-captured images were provided for comparison with each face photograph to increase the possibility that correct identifications are made[32,39,40]. Our drone was also equipped with two different forward-facing cameras, the footage of which was compared on a between-subject basis.

Method

Participants. Forty students (34 female) from the University of Kent, with a mean age of 22.1 years (SD = 8.0), participated for course credit. All experiments reported in this paper were approved by the Ethics Committee in the School of Psychology at the University of Kent and conducted in accordance with the ethical guidelines of the British Psychological Society. In all experiments, informed consent was obtained from all participants before taking part.

Stimuli. A remote-controlled Parrot AR Drone 2.0 Power Edition, with a minimum take-off weight (MTOW) of 300 g, was employed for stimulus capture. This type of drone falls into Class I(b) with a MTOW of 200 g-2 kg as classified by NATO regulation[7], and is comparable to drones that are in use by the UK military[3] and police force[6]. The drone was equipped with two forward-facing cameras, comprising the drone's integrated HD camera with a maximum video resolution of 1280 × 720 pixels at 30 fps, and a retro-fitted GoPro Hero4 Silver with a maximum video resolution of 2704 × 1520 pixels at 30 fps. To provide stimulus footage for the experiments, this drone recorded the protagonists of a football game from pitch-side. Maximum flight-height was restricted to 15 metres of elevation using the drone's navigation software (AR.FreeFlight2.4 v2.4.22). From each camera, a total of 42 images were extracted manually with graphics software, comprising three images for each of 14 different players. This footage was synchronized across cameras, so that it captured the players at the same point in time, but varied depending on each camera's characteristics. The sets of three same-person images were arranged side-by-side, and displayed each at a size of 150 × 150 pixels for the drone camera at 72 ppi. The GoPro images were presented at a slightly smaller size of 120 × 120 at 72 ppi due to the higher resolution of this recording equipment. In addition, a high-quality full-face photograph was also taken for each player at a distance of approximately 1 m immediately prior to the drone recording. These images were then cropped to remove extraneous background and resized to 250 (W) × 340 (H) pixels at a resolution of 72 ppi.

To create the stimulus displays, the high-quality full-face images were arranged above three drone-captured video stills. For each of the 14 players, an identity match was created, in which the full-face photograph and drone-captured images depicted the same person, and an identity mismatch, in which two different people were shown. These mismatch pairings were generated by the experimenters (MB and MCF) based on the extent to which different identities were similar in terms of race, hair colour, and age. However, considering the small pool of targets, the number of possible pairings was restricted greatly (for example, the pool of targets comprised only two players of African ethnic origin). Combining the stimulus images in this way resulted in a total of 56 experimental trials, comprising 28 for each drone camera (14 identity matches and 14 mismatches). Example stimuli are illustrated in Fig. 1.

Figure 1. Illustration of an aerial view from the GoPro camera (top) with a highlighted target (red circle). The top array depicts three image stills from the drone-integrated camera and GoPro for this target, and the high-quality face photograph for the familiarity check. The bottom array depicts the corresponding images of the mismatch identity that was selected for this target. Please note that the depicted target, and all other players visible in this figure, have provided informed consent for publication of these images.

Procedure. Participants were allocated randomly to one of the two camera conditions. The experiment was run on a computer using *PsychoPy* software[41]. Each trial began with a 1-second fixation cross, which was presented in the centre of the screen. This was followed by a stimulus array, which remained onscreen until a button-press response had been registered. Participants were asked to decide as accurately as possible whether a stimulus display depicted an identity match or mismatch, by pressing one of two designated buttons on a standard computer keyboard. Each participant completed 28 trials, which were presented in a unique random order.

The matching task was followed by a familiarity check to eliminate stimulus identities that might have been known to a participant prior to the experiment. For this purpose, the high-quality full-face photographs were presented individually and participants indicated whether they were familiar with a target, by providing a name or uniquely-identifying semantic information.

Data Availability. The experimental stimuli and the datasets generated and analysed during the current experiments are available from the corresponding author on reasonable request.

Results

The familiarity check indicated that participants were familiar on average with 1.6 targets (SD = 1.2) in the drone camera condition and 0.3 targets (SD = 0.6) in the GoPro camera condition prior to the experiment. As each identity featured in one match and two mismatch trials, this led on average to the exclusion of 4.8 (SD = 3.6) and 0.9 (SD = 1.7) trials in these conditions, respectively. For the remaining data, the percentage accuracy for identity match and mismatch trials was calculated.

For the drone's integrated camera, match and mismatch accuracy was at 48.4% (SD = 12.9) and 73.2% (SD = 14.1), respectively. Similarly, accuracy for GoPro images was at 37.1% (SD = 16.0) for match trials and 66.7% (SD = 14.4) for mismatch trials. A 2 (camera type: drone cam vs. GoPro) × 2 (trial type: match vs. mismatch) mixed-factor ANOVA of these data revealed a main effect of camera type, $F(1,38) = 12.94$, $p < 0.001$, $\eta_p^2 = 0.25$, due to overall higher accuracy for the drone camera, and a main effect of trial type, $F(1,38) = 50.59$, $p < 0.001$, $\eta_p^2 = 0.57$, due to higher accuracy for mismatch trials. An interaction between factors was not found, $F(1,38) = 0.39$, $p = 0.53$, $\eta_p^2 = 0.01$.

As accuracy was low, this was also compared to chance performance (i.e., of 50%) via a series of one-sample t-tests (with *alpha* corrected at $p < 0.0125$ [i.e., 0.05/4] for multiple comparisons). This revealed that mismatch accuracy for the drone camera and the GoPro was above chance, $t(19) = 7.36$, $p < 0.001$ and $t(19) = 5.19$, $p < 0.001$, respectively. By contrast, match accuracy was at chance for the drone camera, $t(19) = 0.57$, $p = 0.58$, and below chance for the GoPro, $t(19) = 3.61$, $p < 0.01$.

Discussion

Observers' ability to match drone-captured images to high-quality photographs of unfamiliar faces was at or below chance, with accuracy averaging 43% across camera conditions, which indicates that positive person identifications could not be made reliably. Mismatch decisions were comparatively better but still highly error-prone, averaging at 70%. This low accuracy was obtained despite the provision of three drone-captured images for comparison with each target, which should facilitate person identification[32,39,40], and under conditions in which the mismatch stimuli were constructed from a limited number of identities.

As a small extension of this work, we also compared person identification for footage from two different camera types, comprising the drone's integrated HD camera and a retro-fitted GoPro Hero4 Silver. This revealed an advantage for the drone's integrated camera (61%) over the GoPro (52%). The difference in identification accuracy between these cameras might reflect that the drone's integrated equipment is better optimized for the viewing conditions that are incurred by aerial recordings. However, even for footage captured with the drone's integrated camera, identification accuracy was generally low. By comparison, in face-matching studies that combine high-quality face portraits from more conventional footage in 1-to-1 comparisons, and utilise more refined identity mismatches, mean accuracy is typically at 80–90%[27,28]. The current results therefore suggest that the identification of unfamiliar people from drone-captured footage is a particularly difficult task.

Experiment 2

Whereas unfamiliar face identification is error prone, recognition of familiar faces, that we have encountered many times before, is much more accurate[42,43] and proceeds even under challenging conditions, such as when poor-quality surveillance footage is employed[11]. Consequently, it is possible that people can be identified reliably from drone-captured footage when they are familiar to the observer. This was explored in Experiment 2 by assessing recognition accuracy of observers that personally knew the people in this footage. Two groups of observers were compared, comprising colleagues of the depicted targets and members of the same football group. Participants in the latter group had not been present during the recording of the drone footage, but had the additional contextual advantage of knowing who comprised the members of the football team to facilitate identification. Non-face objects can be identified in familiar contexts from images with very low resolution[44]. Experiment 2 investigates whether a similar advantage is also present during person identification from drone-captured footage.

Method

Participants. The group of colleagues comprised 17 academic staff members (eight male) at the University of Kent, with a mean age of 37.6 years (SD = 11.0), who worked alongside several of the people that were depicted in the drone-captured footage. The group of teammates consisted of ten participants (all male), with a mean age of 44.8 years (SD = 14.6), who were members of the same football group but were absent from play on the day of the drone recording.

Stimuli and Procedure. The stimuli were the same as in Experiment 1, but the face-matching task was replaced with a recognition test. Thus, the three-image arrays of drone images were now presented without the high-quality face image and participants were asked to name the depicted people directly. If participants indicated familiarity but were unable to name the target, then they were asked to provide unique semantic information to confirm identification. In this manner, all participants were presented with 28 stimulus arrays, comprising a three-photo array for each of the fourteen target identities and each of the two cameras. After completion of this task, all participants were given a familiarity check comprising the high-quality full-face photographs.

Results

In the familiarity check, participants recognized on average 3.8 (SD = 0.5) of 14 targets in the colleagues group, equating to 27.3% (SD = 3.8) of identities, and 10.3 (SD = 3.2) of 14 targets, or 73.6% (SD = 22.8), in the teammates group. For these familiar identities, performance with the drone-captured images was analysed by grouping

Experiment 2	Colleagues			
	Hits	*Misids*	*Misses*	*False positives*
Drone camera	11.2 (16.7)	13.4 (20.3)	75.4 (29.4)	14.1 (18.0)
GoPro camera	5.6 (10.4)	19.1 (27.3)	75.3 (29.3)	16.4 (21.2)
	Teammates			
Drone camera	36.5 (11.0)	25.8 (17.9)	37.7 (12.6)	18.1 (20.9)
GoPro camera	34.6 (12.9)	28.1 (16.3)	37.3 (17.1)	20.0 (26.2)
Experiment 3	**Teammates**			
Drone camera	32.7 (16.1)	8.7 (14.0)	32.7 (16.1)	30.7 (36.0)

Table 1. Person Identification Performance for Experiment 2 and 3, by Participant Type (Colleagues versus Teammates) and Camera Type (Drone versus GoPro Camera). Standard deviations are shown in parentheses.

responses into correct identifications of a target (hits), incorrect identifications of a target as somebody else (misidentifications), and those cases in which no identifications were made (misses). In addition, performance was calculated for targets that observers indicated as unknown in the familiarity check. For these, the percentage of trials was calculated on which an identification was incorrectly made (false positives) from the drone-captured images.

The mean percentages of responses that fall into each of these categories are illustrated in Table 1 for both cameras and participant groups. These data show that recognition performance was extremely poor. For example, across both cameras in the teammates group, targets could be identified on only 36% of trials (hits). By contrast, 27% of familiar faces were misidentified as someone else, and 19% of unfamiliar faces were also falsely identified as someone familiar. This poor performance was even more marked in the colleagues group, where hits averaged across both cameras were very low, at 8%, whilst almost twice as many misidentifications (16%) were made.

To analyse these data, separate 2 (group: teammates vs. colleagues) × 2 (camera type: drone cam vs. GoPro) mixed-factor ANOVAs were conducted for each of the four measures. For hits, this analysis revealed a main effect of group, $F(1,25) = 37.51$, $p < 0.001$, $\eta_p^2 = 0.60$, due to higher recognition accuracy among teammates than colleagues. In turn, a main effect of group was also found for misses, $F(1,25) = 17.21$, $p < 0.001$, $\eta_p^2 = 0.41$, as the teammates were less likely to fail to identify a known person. For hits and misses, a main effect of camera was not found, $F(1,25) = 1.67$, $p = 0.21$, $\eta_p^2 = 0.06$ and $F(1,25) = 0.00$, $p = 0.95$, $\eta_p^2 = 0.00$, and no interaction between factors, $F(1,25) = 0.42$, $p = 0.52$, $\eta_p^2 = 0.02$ and $F(1,25) = 0.00$, $p = 0.97$, $\eta_p^2 = 0.00$, respectively. None of the main effects or interactions were significant for misidentifications and false positives, all $Fs(1,25) \leq 2.33$, all $ps \geq 0.14$, all $\eta_p^2 \leq 0.09$.

The different response categories were also compared directly to determine which identification outcome was most likely. For this analysis, the data for both cameras were combined and a series of paired-sample t-tests were conducted to compare hits, misses, misidentifications and false positives (with *alpha* corrected at $p < 0.008$ [i.e., 0.05/6] for multiple comparisons). For teammates, this analysis failed to find differences between any of the measures, all $ts(9) \leq 2.05$, all $ps \geq 0.07$. Thus, teammates were as likely to make a correct identification as an incorrect identification, or to fail to recognize a target altogether. In the colleagues group, observers recorded more misses than hits, misidentifications and false positives, all $ts(16) \geq 5.26$, all $ps < 0.001$, whereas these three measures did not differ from each other, all $ts(16) \leq 1.95$, all $ps \geq 0.07$.

Discussion

Teammate observers already knew more of the targets than colleagues prior to the experiment. They were also more likely to identify these familiar targets from the drone-captured footage, and less likely to fail to recognize a known person, indicating a context advantage that facilitated identification from low-quality images[44]. Generally, however, recognition accuracy was poor for both groups. For example, teammates only identified 36% of targets that could be recognized from the high-quality images, and recognition accuracy for colleagues was at just 8%. Thus, still images from drone-captured footage only allow for very limited recognition of familiar people. This problem is compounded by incorrect identifications, both for targets that were personally familiar and unfamiliar, which occurred as frequently as correct identifications.

Once again, we also compared person identification from footage captured by the drone's integrated camera and a retro-fitted GoPro. As in Experiment 1, an advantage in correct identifications, and a corresponding reduction in incorrect identifications, was obtained for the integrated camera. While this is consistent with the notion that the characteristics of this equipment might be more optimized for aerial footage than the GoPro, these differences were small (~4%) and not statistically reliable in Experiment 2. This might suggest that differences in recording equipment exert less of an effect on the identification of familiar than unfamiliar faces.

Experiment 3

Whilst the identification of familiar people is difficult from drone-captured still images, identification can be enhanced when moving images are provided[45], particularly under difficult viewing conditions[16,46]. Experiment 3 therefore investigated the recognition accuracy of familiar people from moving drone-captured footage, by replacing the image arrays of Experiment 2 with 10-second video recordings from which these still images were originally taken. Due to the comparable performance across camera types in Experiment 2, only footage from the drone's integrated camera was employed in Experiment 3.

Method

Participants. The participants consisted of 16 males who were members of the football group that was recorded by the drone. Seven of these participants are depicted in the drone footage, whereas the other nine were absent from play on the day of the drone recording. One participant failed to record their age. The remaining participants had a mean age of 42.7 years (SD = 10.7). Data collection was conducted online and participants were invited to take part via a football members email list.

Stimuli and Procedure. The stimuli and procedure were identical to Experiment 2, except that the stimulus arrays comprising three drone-captured images were replaced with the video footage from which these still images had been taken. This footage was displayed in an online browser using Qualtrics survey software. In total, 14 video clips were shown, comprising a 10-second recording from the integrated drone camera for each of the 14 target identities. As the recording captured a game of football, several targets were visible in each video clip. The first second of each video therefore displayed a still image in which the to-be-identified target identity was highlighted with a red circle, followed by nine seconds of moving footage that followed on naturally from the still. Following each video, participants were asked to name the target or to provide unique semantic information for identification. The videos were shown in a random order. After completion of the video task, all participants were given the familiarity check comprising the high-quality full-face photographs.

Results

The familiarity check indicated that participants recognized on average 10.5 (SD = 2.5) of the 14 targets, equating to 75.0% (SD = 18.1) of identities. For these familiar identities, performance with the drone-captured footage was broken down into hits, misidentifications and misses (see Table 1). In addition, performance for unfamiliar targets was also converted into false positives. Note that two of the 16 observers recognized all of the target identities in the familiarity check. Analysis of false positives is therefore based on N = 14.

To determine which identification outcome was most likely, the different response categories were compared directly via a series of paired-sample t-tests (with *alpha* corrected at $p < 0.008$ [i.e., 0.05/6] for multiple comparisons). This analysis revealed that more hits than misidentifications of familiar targets were made, $t(15) = 4.71$, $p < 0.001$. By contrast, the percentage of hits was comparable to false positive identifications of unfamiliar targets, $t(13) = 0.21$, $p = 0.84$. Target misses exceeded misidentifications, $t(15) = 5.99$, $p < 0.001$. Misses also exceeded hits and false positives, but these differences were not significant, $t(15) = 2.87$, $p = 0.01$ and $t(13) = 1.89$, $p = 0.08$, respectively. Misidentifications did not differ reliably from false positives, $t(13) = 2.27$, $p = 0.04$.

To examine the potential benefit of moving footage for identification directly, these data were compared with the teammates' performance with still images from the drone camera in Experiment 2 via a series of independent-samples t-tests (with *alpha* corrected at $p < 0.013$ [i.e., 0.05/4] for multiple comparisons). This revealed that hits were comparable for still images and moving footage, $t(24) = 0.65$, $p = 0.52$, as were false positives, $t(22) = 0.99$, $p = 0.33$. By contrast, still images gave rise to more misidentifications, $t(24) = 2.73$, $p < 0.013$, and fewer misses, $t(24) = 2.71$, $p < 0.013$.

Due to the restricted subject pool that teammates provide, seven of the participants of Experiment 3 also appeared in the stimulus footage as football players. As a final step of the analysis, this allowed us to probe self-recognition from the drone footage. All recognized themselves in the familiarity check, but only three of these seven participants (41.9%) recognized themselves in the drone video. Of the remaining four participants, one misidentified themselves as another person (14.2%) and three could not make an identification (41.9%).

Discussion

The percentage of correct identifications from moving drone-captured footage in this experiment was comparable to the static drone footage of Experiment 2. Still images gave rise to more misidentifications of familiar people and fewer cases in which no identification was made. This suggests that moving drone footage might lead participants to exert more caution in committing to an identification. At the same time, false identifications, of targets that were not known to participants prior to the experiment, were comparable across both types of footage.

Overall, these data confirm that person identification from drone footage is highly error-prone. This appears to be the case under conditions that typically facilitate identification, namely when recognition of familiar people is examined[11,42,43], context limits the number of possible answers[47], and moving footage is supplied[16,46]. In addition, the stimuli also provided body information, which can aid identification further[48,49]. The difficulty of this task is illustrated further by the observation that only three of seven participants who were featured as stimuli could identify themselves from the drone footage.

Experiment 4

Considering that identification is highly error-prone, the question arises of whether other person information can be gleaned from drone-captured footage. In Experiment 4, observers unfamiliar with the targets depicted in the drone-captured footage were asked to judge the sex, race and age of these persons. This information is typically extracted accurately from high-quality images[50–54]. It is unknown to which extent this is possible from drone-captured footage.

Method

Participants. A total of 60 participants (33 female) volunteered to participate in this experiment. Two participants did not record their age. The remaining participants had a mean age of 26.4 years (SD = 15.0). All reported normal (or corrected-to-normal) vision. Data collection was conducted online and participants were invited to take part via an email list for research volunteers.

Stimuli and Procedure. For each target, a drone-captured still image was selected from the stimulus arrays of Experiment 2, which depicted the person in frontal or near-frontal face view, as well as the high-quality photographs. The drone-captured images were presented at a size of 150 × 150 pixels, whilst digital photographs were presented at a size of 400 × 300 pixels at a resolution of 72 ppi. In the experiment, these stimuli were displayed in a web browser using Qualtrics software on a between-subject basis. Thus, half of the participants viewed the drone images, whilst the other half the viewed high-quality face photographs. For each target, observers were required to make sex, race, and age judgements, which were presented in multiple-choice format. For sex, the response options consisted of "male" and "female". For race, these choices comprised "White", "Black", "Asian", "Mediterranean", "Indian", and "Hispanic", to reflect the ethnicities of the depicted football players, as well as "Middle-Eastern", and "Mixed-Ethnicity". In addition, observers were permitted to enter an alternative classification in text. For age, each option covered a period of ten years, with 10–19 years and 60 + years being the youngest and oldest possible responses, respectively. In addition, each question included "*Cannot tell*" as a possible response option. Finally, as a familiarity check, participants were also asked to name targets, or to provide unique semantic information, for any targets that were recognized.

Results

The familiarity check indicated that none of the participants recognized any of the 14 targets. The mean percentage of correct responses was then calculated for the drone and high-quality images for the sex, race, and age decisions. An independent-samples t-test showed that the accuracy of sex decisions was better for high-quality face photographs at 98.6% (SD = 3.5) than the drone images at 62.6% (SD = 13.6), $t(58) = 14.03$, $p < 0.001$. A similar advantage for high-quality photographs was observed for race decisions, at 74.3% (SD = 14.8) versus 42.4% (SD = 10.3), $t(58) = 9.68$, $p < 0.001$, and age decisions at 46.4% (SD = 12.4) versus 26.9% (SD = 13.5), $t(58) = 5.84$, $p < 0.001$.

Discussion

This experiment provides broader evidence that drone-captured footage forms an unreliable basis for person perception. Sex information, for instance, was extracted poorly from drone stills, for which only 63% of targets were classified correctly. We did not plan to examine sex categorisation when we initiated this series of experiments, but were led to examine this question by the poor identification accuracy in Experiments 1 to 3. Consequently, all of the targets in the drone-captured footage were men, rather than a mixture of males and females. This one-sided sample could have affected observers' responses from drone-captured footage, by leading to some female-sex decisions on the basis that observers might have expected a proportion of such stimuli in a sex categorization task. However, this was clearly not the case for the high-quality face photographs, for which performance was at ceiling. This contrast demonstrates that the low accuracy of sex categorization is reflective of the drone-captured footage, rather than the composition of targets' sexes in this experiment. Accuracy for race and age decisions from drone-captured footage was even lower, at 42% and 27%, respectively. Thus, these decisions were more likely to be incorrect than correct under the current conditions. By contrast, performance with the face photographs here, as well as previous research, demonstrates that sex, race and age information are consistently extracted with much better accuracy from high-quality images[50–54].

General Discussion

This study explored the extent to which people can be identified from aerial footage recorded by a remote-controlled drone. The identities of unfamiliar (Experiment 1) and familiar target people (Experiment 2 and 3), and their sex, age and race (Experiment 4) were difficult to extract from drone-captured footage. This suggests that such footage provides a challenging substrate for person classification. In an extension of this work, we also compared identification of unfamiliar (Experiment 1) and familiar targets (Experiment 2) for footage from two different camera types, comprising the drone's integrated HD camera and a retro-fitted GoPro. Whilst this revealed a small advantage for the drone's integrated camera, identification accuracy was generally low for both camera types.

Many factors could account for these results. The movement speed and trajectory of the drone, its flight stability, as well as distance-to-target and its high vantage point are likely to degrade the available information for person identification. On the other hand, we employed high-definition recording equipment with good image stabilisation, flight height was limited to only 15 meters, targets were recorded from pitch-side, against a uniform background (the green pitch), and the number of (familiar) target identities was limited.

In this context, the current data have important implications. Drones are already employed routinely in military and police operations, for example, in searches for missing persons[1], crowd control[2], and military reconnaissance and lethal strikes[3,4]. The drone of the current study provides a limited proxy for military aircraft drones. However, some of the smaller drones employed by military and police are comparable to the equipment of this study[6]. For example, some police-employed drones operate at altitudes between ground level and 400 ft and at speeds from 0 to 38 kts[3,6,7]. Moreover, some of these drones carry recording equipment with a resolution that is substantially less than the cameras employed here (e.g., only 640 × 512 pixels)[55]. By comparison, the drone in the current study recorded targets from a maximum altitude of 49 ft in the experiment and was equipped with two cameras with considerably higher resolution (e.g., 1280 × 720 pixels for the drone's integrated HD camera). In addition, the targets' distance and orientation to the drone varied naturally during the recordings, thereby providing multiple perspectives to facilitate identification. These advantages did not appear to offset the difficulty of the task, however. Consequently, the finding that it is extremely difficult to identify people, or even just their sex, race and age from drone-recorded footage, such as that provided in the current study, raises concerns about their application for person perception in police and military operations.

In drawing these conclusions, we note also that this is only the first study to explore person identification with drone-captured footage. Our trial count was restricted by the number of targets that could be recorded, whilst sample size was limited by participants who were familiar with these targets. Moreover, it is presently unknown how these findings generalize across different drone types and viewing conditions. It is possible, for example, that the poor accuracy in person identification that was observed here can be offset when flight height is lowered or drone-to-target distance is reduced, though operational requirements may not allow this to safeguard those on the ground[56] or to avoid detection of a drone during covert deployment[2]. Similarly, it is possible that person identification from drone-captured footage might be improved by magnification equipment, such as optical zoom, though this may also increase the difficulty of target tracking. In addition, we note that the identification of unfamiliar people can be difficult even with high-quality face portraits[8,26,28,42]. Thus, the extent to which person identification is possible from drone-captured footage under viewing conditions that are optimised further is an open question.

Finally, whilst the aim of the current study was to examine person identification from drone-captured footage by human observers, similar studies in computer vision are now beginning to emerge[57]. This raises the question of how the accuracy of human observers and machine algorithms in person identification might compare. Face-matching studies with more conventional footage suggest that machine algorithms outperform human observers under conditions of moderate difficulty[58,59], and perform at least to a similar level with challenging face pairs, such as images in which illumination and a person's day-to-day appearance are variable[58,59]. However, the face images that were employed in these studies are of substantially higher quality than the drone-footage under investigation here, making it difficult to draw direct comparisons at this point in time.

In conclusion, the current study suggests that a person's identity, sex, race and age are difficult to extract from drone-captured footage. However, more extensive studies are clearly needed to investigate unfamiliar and familiar person identification and categorization from drone-captured footage, with more stimuli and greater sample sizes, utilising more drone types, a greater range of image-capture and magnification devices, and with footage recorded under a much wider range of viewing conditions. Comparisons of human observers and machine algorithms in person identification from drone-captured footage are also required to advance the field.

References

1. Porter, T. Surveillance cameras on drones. *Gov. UK*. at https://www.gov.uk/government/case-studies/surveillance-cameras-on-drones (2016).
2. Omand, D. *et al*. The security impact of drones: challenges and opportunities. *Birmingham Policy Commission* at. http://www.birmingham.ac.uk/Documents/research/policycommission/remote-warfare/final-report-october-2014.pdf (2014).
3. Brooke-Holland, L. Overview of military drones used by the UK armed forces. *House of Commons Library*. at http://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN06493#fullreport (2015).
4. Ministry of Defence. The uk approach to unmanned aircraft. *Joint Doctrine Note 2/11* at https://www.gov.uk/government/publications/jdn-2-11-the-uk-approach-to-unmanned-aircraft-systems (2011).
5. Proxy Dynamics. PD-100 PRS. *FLIR Unmanned Aerial Systems* at http://www.proxdynamics.com/products/pd-100-black-hornet-prs.
6. Camber, R. Take off for police drones air force: Remote-controlled 'flying squad' to chase criminals and hunt for missing people. *Daily Mail* at http://www.dailymail.co.uk/news/article-4329714/Remote-controlled-flying-squad-chase-criminals.html (2017).
7. Ministry of Defence and Military Aviation Authority. Regulatory Article (RA) 1600: Remotely piloted air systems. *Gov. UK*. at https://www.gov.uk/government/publications/regulatory-article-ra-1600-remotely-piloted-air-systems-rpas (2015).
8. White, D., Kemp, R. I., Jenkins, R., Matheson, M. & Burton, A. M. Passport officers' errors in face matching. *PLoS One* **9**, e103510, https://doi.org/10.1371/journal.pone.0103510 (2014).
9. White, D., Dunn, J. D., Schmid, A. C. & Kemp, R. I. Error rates in users of automatic face recognition software. *PLoS One* **10**, e0139827, https://doi.org/10.1371/journal.pone.0139827 (2015).
10. Bruce, V., Henderson, Z., Newman, C. & Burton, A. M. Matching identities of familiar and unfamiliar faces caught on CCTV images. *J. Exp. Psychol. Appl.* **7**, 207–218, https://doi.org/10.1037//1076-898X.7.3.207 (2001).
11. Burton, A. M., Wilson, S., Cowan, M. & Bruce, V. Face recognition in poor-quality video: Evidence from security surveillance. *Psychol. Sci.* **10**, 243–249, https://doi.org/10.1111/1467-9280.00144 (1999).
12. Bindemann, M., Brown, C., Koyas, T. & Russ, A. Individual differences in face identification postdict eyewitness accuracy. *J. Appl. Res. Mem. Cogn.* **1**, 96–103, https://doi.org/10.1016/j.jarmac.2012.02.001 (2012).
13. Memon, A., Havard, C., Clifford, B. & Gabbert, F. A field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. *Psychol. Crime Law* **17**, 711–729, https://doi.org/10.1080/10683160903524333 (2011).
14. Bruce, V., Carson, D., Burton, A. M. & Kelly, S. Prime time advertisements: Repetition priming from faces seen on subject recruitment posters. *Mem. Cognit.* **26**, 502–515, https://doi.org/10.3758/BF03201159 (1998).
15. Bruce, V. & Valentine, T. Identity priming in the recognition of familiar faces. *Br. J. Psychol.* **76**, 373–383, https://doi.org/10.1111/j.2044-8295.1985.tb01960.x (1985).
16. Lander, K. & Bruce, V. Recognizing famous faces: Exploring the benefits of facial motion. *Ecol. Psychol.* **12**, 259–272, https://doi.org/10.1207/S15326969ECO1204_01 (2000).
17. Bindemann, M., Burton, A. M. & Jenkins, R. Capacity limits for face processing. *Cognition* **98**, 177–197, https://doi.org/10.1016/j.cognition.2004.11.004 (2005).
18. Brunas, J., Young, A. W. & Ellis, A. W. Repetition priming from incomplete faces: Evidence for part to whole completion. *Br. J. Psychol.* **81**, 43–56, https://doi.org/10.1111/j.2044-8295.1990.tb02344.x (1990).
19. Johnston, R. A. Incomplete faces don't show the whole picture: Repetition priming from jumbled faces. *Q. J. Exp. Psychol. Sect. A* **49**, 596–615, https://doi.org/10.1080/027249896392513 (1996).
20. Troje, N. F. & Kersten, D. Viewpoint-dependent recognition of familiar faces. *Perception* **28**, 483–487, https://doi.org/10.1068/p2901 (1999).
21. Burton, A. M. Why has research in face recognition progressed so slowly? The importance of variability. *Q. J. Exp. Psychol.* **66**, 1467–1485, https://doi.org/10.1080/17470218.2013.800125 (2013).
22. Jenkins, R. & Burton, A. M. Stable face representations. *Philos. Trans. R. Soc. Lond. B.* **366**, 1671–1683, https://doi.org/10.1098/rstb.2010.0379 (2011).
23. Jenkins, R., White, D., Van Montfort, X. & Burton, A. M. Variability in photos of the same face. *Cognition* **121**, 313–323, https://doi.org/10.1016/j.cognition.2011.08.001 (2011).
24. Burton, A. M., Jenkins, R., Hancock, P. J. B. & White, D. Robust representations for face recognition: The power of averages. *Cogn. Psychol.* **51**, 256–284, https://doi.org/10.1016/j.cogpsych.2005.06.003 (2005).

25. Young, A. W. & Burton, A. M. Recognizing Faces. *Curr. Dir. Psychol. Sci.* **26**, 212–217, https://doi.org/10.1177/0963721416688114 (2017).
26. Bruce, V. *et al*. Verification of face identities from images captured on video. *J. Exp. Psychol. Appl.* **5**, 339–360, https://doi.org/10.1037//1076-898X.5.4.339 (1999).
27. Bindemann, M., Avetisyan, M. & Rakow, T. Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *J. Exp. Psychol. Appl.* **18**, 277–291, https://doi.org/10.1037/a0029635 (2012).
28. Burton, A. M., White, D. & McNeill, A. The Glasgow Face Matching Test. *Behav. Res. Methods.* **42**, 286–291, https://doi.org/10.3758/BRM.42.1.286 (2010).
29. Kemp, R. I., Towell, N. & Pike, G. When seeing should not be believing: Photographs, credit cards and fraud. *Appl. Cogn. Psychol.* **11**, 211–222, doi:10.1002/(SICI)1099-0720(199706)11:3<211::AID-ACP430>3.0.CO;2-O (1997).
30. Megreya, A. M. & Burton, A. M. Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *J. Exp. Psychol. Appl.* **14**, 364–372, https://doi.org/10.1037/a0013464 (2008).
31. Davis, J. P. & Valentine, T. CCTV on trial: Matching video images with the defendant in the dock. *Appl. Cogn. Psychol.* **505**, 482–505, https://doi.org/10.1002/acp.1490 (2009).
32. Ritchie, K. L. & Burton, A. M. Learning faces from variability. *Q. J. Exp. Psychol.* **70**, 897–905, https://doi.org/10.1080/17470218.2015.1136656 (2017).
33. Dowsett, A. J., Sandford, A. & Burton, A. M. Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *Q. J. Exp. Psychol.* **69**, 1–10, https://doi.org/10.1111/bjop.12103 (2016).
34. Longmore, C., Liu, C. H. & Young, A. W. Learning faces from photographs. *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 77–100, https://doi.org/10.1037/0096-1523.34.1.77 (2008).
35. Bindemann, M., Attard, J., Leach, A. & Johnston, R. A. The effect of image pixelation on unfamiliar-face matching. *Appl. Cogn. Psychol.* **27**, 707–717, https://doi.org/10.1002/acp.2970 (2013).
36. Fysh, M. C. & Bindemann, M. Forensic face matching: A review. In Bindemann, M. & Megreya, A. M. (eds) *Face processing: Systems, disorders and cultural differences*. (pp. 1–20) New York: Nova Science Publishing, Inc (2017).
37. Johnston, R. A. & Bindemann, M. Introduction to forensic face matching. *Appl. Cogn. Psychol.* **27**, 697–699, https://doi.org/10.1002/acp2963 (2013).
38. Estudillo, A. J. & Bindemann, M. Generalization across view in face memory and face matching. *i-Perception.* **5**, 589–601, https://doi.org/10.1068/i0669 (2014).
39. Bindemann, M. & Sandford, A. Me, myself, and I: Different recognition rates for three photo-IDs of the same person. *Perception* **40**, 625–627, https://doi.org/10.1068/p7008 (2011).
40. White, D., Burton, A. M., Jenkins, R. & Kemp, R. I. Redesigning photo-ID to improve unfamiliar face matching performance. *J. Exp. Psychol. Appl.* **20**, 166–73, https://doi.org/10.1037/xap0000009 (2014).
41. Peirce, J. W. PsychoPy - Psychophysics software in python. *J. Neurosci. Methods.* **162**, 8–13, https://doi.org/10.1016/j.jneumeth.2006.11.017 (2007).
42. Megreya, A. M. & Burton, A. M. Unfamiliar faces are not faces: Evidence from a matching task. *Mem. Cognit.* **34**, 865–876, https://doi.org/10.3758/BF03193433 (2006).
43. Ritchie, K. L. *et al*. Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox. *Cognition.* **141**, 161–169, https://doi.org/10.1016/j.cognition.2015.05.002 (2015).
44. Barenholtz, E. Quantifying the role of context in visual object recognition. *Vis. Cogn.* **22**, 30–56, https://doi.org/10.1080/13506285.2013.865694 (2017).
45. Lander, K. & Bruce, V. Repetition priming from moving faces. *Mem. Cognit.* **32**, 640–647, https://doi.org/10.3758/BF03195855 (2004).
46. Lander, K., Christie, F. & Bruce, V. The role of movement in the recognition of famous faces. *Mem. Cogn.* **27**, 974–985, https://doi.org/10.3758/BF03201228 (1999).
47. Davies, G. & Milne, A. Recognizing faces in and out of context. *Curr. Psychol. Rev.* **2**, 235–246, https://doi.org/10.1007/BF02684516 (1982).
48. O'Toole, A. J. *et al*. Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Vision Res.* **51**, 74–83, https://doi.org/10.1016/j.visres.2010.09.035 (2011).
49. Rice, A., Phillips, P. J. & O'Toole, A. J. The role of the face and body in unfamiliar person identification. *Appl. Cogn. Psychol.* **27**, 761–768, https://doi.org/10.1002/acp.2969 (2013).
50. Megreya, A. M. The effects of a culturally gender-specifying peripheral cue (headscarf) on the categorization of faces by gender on the categorization of faces by gender. *Acta Psychol.* **158**, 19–25, https://doi.org/10.1016/j.actpsy.2015.03.009 (2015).
51. Moyse, E. & Brédart, S. An own-age bias in age estimation of faces. *Rev. Eur. Psychol. Appl.* **62**, 3–7, https://doi.org/10.1016/j.erap.2011.12.002 (2012).
52. Rossion, B. Is sex categorization from faces really parallel to face recognition? *Vis. Cogn.* **9**, 1003–1020 (2002).
53. Wild, H. A. *et al*. Recognition and sex categorization of adults' and children's faces: Examining performance in the absence of sex-stereotyped cues. *J. Exp. Child Psychol.* **77**, 269–291, https://doi.org/10.1006/jecp.1999.2554 (2000).
54. Zhao, L. & Bentin, S. Own- and other-race categorization of faces by race, gender, and age. *Psychon. Bull. Rev.* **15**, 1093–1099, https://doi.org/10.3758/PBR.15.6.1093 (2008).
55. DJI FLIR Zenmuse XT 336x256 30Hz Thermal Imaging Camera. at http://firecam.com/dji-flir-zenmuse-xt-336x256-30hz-thermal-imaging-camera/.
56. Civil Aviation Authority (CAA). Unmanned Aircraft System Operations in UK Airspace – Guidance. *Cival Aviation Publication*. **722** at https://publicapps.caa.co.uk/modalapplication.aspx?appid=11&mode=detail&id=415 (2015).
57. Layne, R., Hospedales, T. M. & Gong, S. Investigating open-world person re-identication using a drone. *Computer Vision - ECCV 2014 Workshops.* **3**, 225–240, https://doi.org/10.1007/978-3-319-16199-0_16 (2014).
58. O'Toole, A. J., An, X., Dunlop, J., Natu, V. & Phillips, P. J. Comparing face recognition algorithms to humans on challenging tasks. *ACM Trans. Appl. Percept.* **9**, 1–13, https://doi.org/10.1145/2355598.2355599 (2012).
59. O'Toole, A. J. *et al*. Face recognition algorithms surpass humans matching faces over changes in illumination. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 1642–1646, https://doi.org/10.1109/tpami.2007.1107 (2007).

Author Contributions

M.B. and M.C.F. designed the experiments. M.C.F., S.S.K.S., K.D. and H.M.T. collected the data. M.B. wrote the first draft, and M.C.F. contributed in writing this report.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.