Citation for published version


DOI

https://doi.org/10.3390/drones2040032

Link to record in KAR

https://kar.kent.ac.uk/69936/

Document Version

Publisher pdf

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version. Users are advised to check http://kar.kent.ac.uk for the status of the paper. Users should always cite the published version of record.

Enquiries

For any further enquiries regarding the licence status of this document, please contact: researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at http://kar.kent.ac.uk/contact.html
Review

Person Identification from Drones by Humans: Insights from Cognitive Psychology

Matthew C. Fysh * and Markus Bindemann
School of Psychology, University of Kent, Canterbury CT2 7NP, UK; m.bindemann@kent.ac.uk
* Correspondence: m.c.fysh@kent.ac.uk; Tel.: +44-01227-827374
Received: 14 August 2018; Accepted: 26 September 2018; Published: 28 September 2018

Abstract: The deployment of unmanned aerial vehicles (i.e., drones) in military and police operations implies that drones can provide footage that is of sufficient quality to enable the recognition of strategic targets, criminal suspects, and missing persons. On the contrary, evidence from Cognitive Psychology suggests that such identity judgements by humans are already difficult under ideal conditions, and are even more challenging with drone surveillance footage. In this review, we outline the psychological literature on person identification for readers who are interested in the real-world application of drones. We specifically focus on factors that are likely to affect identification performance from drone-recorded footage, such as image quality, and additional person-related information from the body and gait. Based on this work, we suggest that person identification from drones is likely to be very challenging indeed, and that performance in laboratory settings is still very likely to underestimate the difficulty of this task in real-world settings.

Keywords: drones; person identification; face matching; face recognition

1. Background

Over the last century, the study of visual processing in Psychology has been instrumental in developing an understanding of how humans parse visual information [1,2]. This review paper summarises key insights from this field into how identity is visually derived from people. All of these studies are based on human performance data in principled scientific studies typically comprising a series of related experiments. This paper provides a short literature review of key factors that affect person identification from drone footage by human observers.

Drones are routinely deployed for police and military operations that rely on the successful identification of people. In the UK, for example, drones are employed by police to track suspects [3], as well as to search for missing persons [4]. In addition, UK and US military forces use drones for the acquisition and elimination of target persons [5]. The deployment of drones for such purposes implies that drone-recorded footage can facilitate person identification. This is difficult to reconcile with reports from personnel who remotely pilot drones, which suggest that the quality of drone footage is actually very poor [6]. These reports converge with accounts of civilian casualties [7], as well as fatalities [8,9], which have been attributed to person misidentification errors based on aerial drone footage.

These real-world errors are corroborated by an extensive literature on person identification in Cognitive Psychology. This research shows that, whilst identity can be derived from someone’s body [10] and gait [11], the face is the most useful cue for facilitating identification [12,13]. Subsequent studies have demonstrated that the recognition of familiar faces, such as those belonging to a friend, family member, or famous celebrity, tends to be highly reliable [14,15], and can be facilitated even under limited viewing conditions such as when images of faces are moderately degraded [13,16]. By contrast, viewers often struggle to distinguish one unfamiliar person from another even under...
tightly controlled experimental conditions, as well as fail to recognize that two photographs depict the same unfamiliar person [17,18].

This latter process is conventionally investigated via face-matching tasks [19,20], in which participants view two side-by-side face photographs, and decide whether they depict the same person, or two different individuals [21,22]. Stimuli in these tasks typically consist of high-resolution images of faces that are matched in terms of expression, pose, and lighting (see Figure 1). In addition, pairs of photographs that depict the same person are often taken just minutes apart, to minimise the natural variability that can arise within a person’s appearance over time. Under these conditions, which are designed to maximise identification accuracy, viewers make around 20% errors [21].

Figure 1. An example pair of faces from the Glasgow Face Matching Test (GFMT) [21]. In the top row, two different images of the same person are presented (i.e., an identity match), whereas the bottom pair depicts two different people (i.e., an identity mismatch).

Such error rates represent a best-case scenario. In the real world, the identification of unfamiliar people from photographs is typically based on sub-optimal material such as variable ambient images, low-resolution closed-circuit television (CCTV) footage, and ID photographs that are only updated every ten years. There is now substantial evidence that under conditions such as these, the difficulty of person identification increases dramatically [13,23–25]. In addition, there is little reason to anticipate that trained experts cope with such challenges any better than novices. Some studies show, for instance, that experienced passport officers perform comparably to untrained students when comparing photographs of unfamiliar faces [26] and similarly, that police officers perform comparably to students when identifying people from poor-quality surveillance footage [13]. More recent work also suggests that even when experts do outperform novices in such tasks, substantial error rates still arise [27–29].

Such findings raise concern surrounding the reliability of person identification judgements based on drone footage, which can be heavily degraded or pixelated. The quality of such footage may also be further reduced due to unfavourable aerial views, unpredictable ambient conditions, and angular momentum (see Figure 2). In addition, the difficulty of identifying people from drone footage has
already been highlighted by drone image analysts, who are responsible for relaying visual information from drone footage to the military personnel who operate drone weaponry. These reports suggest that the camera feed from state-of-the-art military drones can be so degraded that it is difficult to distinguish a shovel from a rifle [6], men from women [7], and even adults from children [9].

Figure 2. The quality of drone-recorded footage is susceptible to a number of factors, such as the height of the drone itself, as well as ambient conditions. In addition, such footage might be further hindered by obstacles at ground level such as trees and bystanders, as well as the movement of the target themselves.

Whilst it should follow intuitively that person identification under such limited conditions is difficult, laboratory studies can provide us with a “ballpark estimate” of precisely how error-prone this task might be. Here, we review a number of studies from Cognitive Psychology that provide some insight into this question. We begin by reporting factors that influence the identification of people from the face such as, for example, image degradation, and changes in viewpoint. We then proceed to consider how identity may also be derived based on cues from the body, and whether this process can be enhanced further when viewing people in motion, as opposed to when stimuli comprise static images. Finally, we summarise a recent psychological study of person identification from aerial footage that was recorded using an aerial drone, and discuss the real-world implications of these findings.

2. Person Identification from the Face

Facial identification can be rather challenging when viewing material that is degraded or pixelated. For instance, in one investigation participants viewed a still image from poor-quality CCTV footage alongside a high-quality photograph of an unfamiliar face, and attempted to determine whether these were one person or two [30]. Error rates for this task were extremely high, with viewers incorrectly classifying almost a third of pairs as different people when they actually depicted the same person. The reverse error of this was also common, whereby nearly half of image pairs showing the same person were mistaken for different people.
This level of performance aligns with a subsequent study in which the resolution (i.e., the number of pixels) of face images was systematically reduced [31]. In this study, participants matched pairs of optimized high-resolution images (i.e., 350 pixels in width) of the same person with 90% accuracy, whilst concurrently discriminating different identities with 86% accuracy. Conversely, when viewing a high-resolution face image alongside a heavily-pixelated low-resolution counterpart (i.e., 8 pixels in width), participants were able to match same-identity face pairs with only 48% accuracy, and distinguish identity mismatches with accuracy rates of 60%.

Perhaps reassuringly, there is some suggestion that errors arising from low image resolution can be partially offset by reducing the size of moderately pixelated face images [31], or by increasing the distance between a viewer and a pixelated face image [32]. It should be noted, however, that these manipulations mitigate errors without elevating accuracy. For instance, reducing the size of pixelated face images reduces error rates from around 37% to approximately 24% [31]. In other words, under these improved conditions people could still be reliably expected to misidentify one in four pairs of faces.

How might these experimental findings translate to the identification of unfamiliar people from drones in the real world? Many laboratory studies selectively manipulate a single aspect of this task, such as image quality [31], whilst the homogeneity of other variables, such as facial orientation, remains preserved. Whilst this can be highly informative of how these specific aspects of the task influence identification performance, it remains difficult to predict how successfully drone-recorded footage might facilitate identification.

Incidentally, differences in facial pose add a further layer of difficulty to a task that is already quite challenging [17,33,34]. For instance, when comparing a frontally oriented face to one viewed from the side, participants are 10–15% more likely to mistake different identities for the same person under otherwise ideal viewing conditions than when both faces are viewed from the front [31,33]. Subsequent research has found that participants can match two frontally oriented faces as successfully as two faces that are viewed from the side [35]. This converges with earlier evidence to suggest that identification is disrupted to a greater extent by comparing two differently-oriented faces than by comparing two non-frontal faces that are viewed in the same orientation [34]. This implies that drone footage may be most effective for facilitating identification when the target person is depicted in a similar pose to a comparison image.

One further way in which facial comparison tasks in experimental settings are designed to maximise performance is to allow participants unlimited time to compare photographs of unfamiliar faces [21,36,37]. However, it is conceivable that when attempting to identify someone from drone-recorded material, the available time for making an identification may be limited for a number of practical reasons, such as if a suspected militant is only briefly exposed whilst moving from one location to another. Constraining the amount of time for which participants view faces in the laboratory exerts some intriguing effects on facial comparison performance. Studies suggest that participants require between one and two seconds to decide whether two faces depict one person or two [38,39]. Shorter display durations appear to specifically reduce observers’ capacity to distinguish different identities, by around 10% [38,40]. Similar effects are observed following the administration of time pressure, whereby compelling participants to make identity judgements more quickly also leads to an escalation in the number of different identities that are mistaken for being the same person [41,42].

In the context of drones, these represent errors of the worst kind, whereby, for example, a civilian may be wrongfully identified as a militant. If drone image analysts also experience time pressure to identify people from recorded footage, then this finding highlights yet further capacity for identification errors based on drone-recorded footage.

3. Person Identification from the Body

Person identification may not always be based on facial information alone. The distance from which people are observed may reduce the utility of the face for identification purposes considerably,
for example [43]. In addition, similarities in facial appearance may result in two different people being classified as the same person when decisions are based solely on the face. Conversely, such errors may be offset by analyzing physical characteristics of the body, such as height, weight, and build.

There is some evidence that the body may support identification decisions under such conditions. For example, in one study participants attempted to identify a person from video footage filmed at far, moderate, and close distances [43]. To isolate the contribution of the face and the body for such decisions, the videos were edited to show either the whole person, the person’s face without the body, or the person’s body without the face. Across all distance conditions, to-be-identified persons were more accurately identified based on the face than on the body. Identity judgements based on the whole person (i.e. the face and the body) were also comparable in accuracy to those based on the face alone at moderate and close distances. More importantly, whole-person judgements were more accurate than for the face or the body alone when the target person was furthest away. In summary, these results reflect that when attempting to identify someone from far away, people integrate information from both the body and the face to make an identification. Conversely, as the distance between a viewer and target narrows, identity decisions become primarily driven by information from the face.

This intriguing finding converges with earlier evidence that people utilise information from the body in identification tasks under limiting conditions [10], but without consciously being aware of doing so [44]. For instance, in one study viewers rated facial features (e.g., eyes and nose) as being more useful for identification than body features (e.g., the shoulders). This was despite being more successful at this task when both the face and the body were available for analysis, as opposed to when only the face was presented [44]. Another study suggests that viewers utilise the body under more adverse conditions. For example, when comparing images of different people that happen to look very similar, the inclusion of the body appears to improve participants’ ability to distinguish one person from another. Likewise, the inclusion of the body in identity photographs also seems to aid performance when viewing very dissimilar images of the same person, as opposed to when comparing images of just the face [10]. In the context of drones, these findings therefore suggest that under adverse conditions that preclude identification from the face alone, information from the body may be utilised to enhance accuracy.

4. Person Identification from Motion

Research has also considered the role of motion in person identification. In one study, participants more accurately identified pixelated familiar faces from video footage than when these were presented as still images [32]. In addition, students in another study could accurately identify their lecturers from poor-quality surveillance footage [13]. It is worth noting, however, that in this latter study, obscuring the gait of people in video footage reduced identification accuracy slightly, by around 5%, whilst obscuring the person’s face reduced accuracy enormously, by around 60%. That is to say, the removal of gait—a motion-based cue—was less detrimental to identification performance than the removal of the face—a non-motion-based cue—when viewing degraded footage. These findings therefore suggest that when attempting to identify familiar people, gait should perhaps not be considered a crucial factor for facilitating recognition.

Indirect support for this proposal comes from other work, which found that accuracy for the identification of familiar people is similar between video and image format [45]. Perhaps more importantly, this study also found that the identification of unfamiliar people seems to be worse when viewing degraded footage, compared to when viewing a single “best” static image extracted from the source material. In the context of drones being deployed to record footage of people on the ground, this work suggests that the selection of one useful static image may enhance identification accuracy over viewing an extended video clip.

At the same time, unfamiliar people might be identified more reliably through high-quality video footage than high-quality photographs [11,46]. Yet, whilst such findings communicate that viewing dynamic versus static people can benefit identification accuracy, these investigations are limited in
what they can tell us about identifying people from drone-recorded footage; the quality of which currently appears to preclude the discrimination of men from women [7].

5. Person Identification from Aerial Footage Collected by a Drone

Based on the studies described so far, it seems reasonable to assert that identifying people from drone footage is a difficult task. However, it is tricky to establish from these studies alone just how difficult this task is. Consider, for example, that accuracy is around 50% when matching heavily-pixelated images of the same unfamiliar person [31]. This level of performance represents people’s ability to compare a pixelated face to a high-resolution counterpart under conditions that are otherwise designed to maximise face-matching accuracy. Performance in the real world becomes substantially more difficult to predict when considering additional factors, such as the inclusion of the body which might improve accuracy [43], but also variations in height, position, and vantage point, which might increase errors. In other words, the accuracy of person identification from drones cannot be easily inferred based on material that was not obtained via a drone.

To date, only one study has directly investigated the extent to which person identification by humans is possible from still images and video footage that were gathered using an aerial drone [47]. In this study, a Parrot AR drone with a minimum take-off weight (MTOW) of 300 g was used to record 14 male adults playing a game of football (soccer). According to NATO taxonomy, this type of drone would fall into Class I (b), which describes minidrones that are deployed for person surveillance and target acquisition [5]. An aerial view from the perspective of the drone can be viewed in Figure 3. In addition, a close-up high-quality digital face photograph was obtained of each person on the same day.

Figure 3. An aerial view from the perspective of the drone employed by Bindemann et al. [47].

Across several experiments, person identification from this drone-recorded footage was tested and was found to be poor. In one experiment, which was designed to provide optimized conditions to study person identification from drone-captured footage, observers viewed three still images of an unknown person that had been extracted from drone-captured video, and which were presented alongside a high-quality face photograph (see Figure 4). Participants were then asked to decide whether the person in the drone stills was the same as the person depicted in the digital face photograph. Accuracy for matching three drone images to a high-quality photograph of the same person was at 48%,
whilst people could distinguish different identities with accuracy rates of 73%. A further experiment suggested that viewers could recognize familiar targets from 10-s segments of drone-captured video footage with only 33% accuracy. This is perhaps surprising, given research showing that familiar-face identification can be reliably facilitated under impoverished conditions [13,45]. Yet other studies show that even familiar faces can be misidentified when such decisions are based on photographs that are highly pixelated [16,32,48], and that a resolution “cut-off” exists at which point faces can no longer be reliably identified. Consequently, it is conceivable that the drone footage recorded by Bindemann et al. [47] was of insufficient quality to facilitate accurate recognition even of familiar people. Finally, observers who were unfamiliar with the targets in Bindemann et al.’s study were also asked to judge the sex, race, and age of targets from drone-captured images and could only do so with an accuracy of 63%, 42%, and 27%, respectively. Together, these findings illustrate that identification of both unfamiliar and familiar people from drone footage, as well as the perception of people generally, represents a very difficult task.

**Figure 4.** Example stimuli used by Bindemann et al. [47]. The left panel represents an identity match, whereby the high-quality digital photograph shows the same person as depicted in the three images extracted from the drone camera that are shown underneath. Conversely, the right panel depicts an identity mismatch, whereby the high-quality digital photograph depicts a different person to the one shown in the three images below.

Even these error rates may represent a best-case scenario for person identification from drones in the real world. For example, the drone that was employed by Bindemann et al. [47] recorded people from a maximum height of 15 m. By contrast, Class I NATO drones that are deployed for surveillance and reconnaissance, for example, operate at altitudes ranging from ground-level to 15,000 ft, and at speeds of up to 80 kts [5]. In addition, weaponised Class III drones have a maximum operating altitude of 50,000 ft, and can travel at up to 250 kts [5]. Even smaller drones, such as those deployed by the police, operate at speeds of up to 38 kts and altitudes up to 400 ft [49]. Considering this range in operational parameters, we conclude that it would be unsurprising if the level of performance observed by Bindemann et al. [47] still substantially underestimates the true difficulty of identifying people from drone-recorded footage. Of course, the accuracy of this process should expectedly improve following further developments in technology that enhance the quality of recorded footage, as well as
the stability of drones when airborne. However, we reiterate that person identification by humans remains difficult even under optimal viewing conditions [21,50].

6. Possible Solutions

One potential solution to this problem might be the development of person-recognition algorithms. Recent work has made progress in developing systems that are capable of tracking [51], detecting [52] and identifying individuals in drone footage [53]. In addition, automated recognition systems have demonstrated near-perfect performance in some benchmark tests [54,55]. Yet, such results encounter the same problem as with benchmark tests of human ability to identify faces, namely that such tests represent a limited proxy for the real world. Indeed, algorithms have also been found to perform substantially worse than humans in identification tests that incorporate relevant challenges from the real world, such as problematic lighting and nonfrontal poses [44,56,57]. Perhaps for this reason, these systems continue to be monitored in practical settings by humans who are responsible for verifying correct decisions made by these systems, whilst simultaneously overruling cases where the system has made an incorrect judgement [28]. Current research suggests that human observers cannot reliably detect instances where the system has made an inaccurate identification [58], implying that algorithms bias the identity judgements of humans. This means that for the foreseeable future, the final identification decision in real world settings will continue to reside with the human observer.

An alternative strategy for reducing identification errors in humans might be to recruit drone image analysts who are already highly proficient at facial comparison. A great deal of research in the last decade has focused on individual differences in facial identification [22,59–62], and it is now established that the ability to identify faces varies considerably from one person to another. For example, even under optimized conditions, some participants perform at chance level (i.e., 50%) when matching photographs of unfamiliar faces, whilst others demonstrate perfect accuracy [21]. Recent work has also identified a number of individuals—sometimes referred to as “super-recognizers”—who are remarkably good at recognizing faces even under adverse conditions [48,62]. For instance, one recent study showed that, despite heavy image pixelation, super-recognizers could distinguish images of celebrity faces from lookalikes with 93% accuracy. By contrast, student control participants could do this with only 73% accuracy [48].

Currently it is unclear as to why most super-recognizers are better at face recognition. There is some evidence that high proficiency in this task can be trained [28,56,57]. On the other hand, trained professionals have also demonstrated novice-level performance in other identification tasks [13,26,28]. In the context of person identification from drones, therefore, a possible strategy for the immediate future might be to recruit image analysts based on their performance in benchmark tests of person identification. Similar strategies are already being advocated for other settings that rely heavily on person identification, such as passport control [63], the police [64], and in banks [65].

7. Conclusions

The aim of this review was to highlight some of the challenges for human observers associated with identifying people from drones. Many obstacles to successfully identifying someone from drone-recorded footage can be anticipated based on Cognitive Psychology research. Although at present, there is no clear solution to mitigating the difficulties that can arise when attempting to make an identification using footage obtained from a drone, there are several important points that emerge from current research. The first of these is that it is already difficult to identify unfamiliar people based on optimized digital photographs of faces [21]. This means that even if one could obtain high-resolution imagery of to-be-identified persons from drone-recorded footage, some errors could still be expected to occur. A second, related point is that if drone footage produces high-quality footage, additional factors such as time pressure and changes in facial orientation must nonetheless be accounted for. Finally, identification performance in laboratory settings is likely to underestimate identification performance in the real world. Whilst experimental stimuli provide a useful basis for
establishing how poor people are under controlled conditions, these cannot account for the range of conditions that arise in nature.

Author Contributions: M.C.F. drafted the initial version of this manuscript. This was then revised by M.B., who provided edits and suggestions, and contributed to writing the final version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The individual components for Figure 2 were retrieved from https://pixabay.com/.

Conflicts of Interest: The authors declare no conflicts of interest.

References


© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).