

Facial image comparison by video superimposition

Facial wipes don't wash: Facial image comparison by video superimposition reduces the accuracy of face matching decisions

AILSA STRATHIE^{1, *} & ALLAN McNEILL²

¹ Institute for Infrastructure & Environment, Heriot-Watt University, Edinburgh, UK.

² Department of Psychology and Allied Health Sciences, Glasgow Caledonian University, Glasgow, UK.

* *Institute for Infrastructure & Environment, School of the Built Environment, Heriot-Watt University, Riccarton, Edinburgh, EH14 4AS, United Kingdom; Tel: +44 (0)131 451 4639; E-mail: a.strathie@open.ac.uk*

Facial wipes don't wash: Facial image comparison by video superimposition reduces the accuracy of face matching decisions

Abstract

In cases of disputed CCTV identification, expert testimony based on the results of analysis by facial image comparison may be presented to the Jury. However, many of the techniques lack empirical data to support their use. Using a within participants design, we compared the accuracy of face matching decisions when images were presented using a 'facial wipe' technique (where one image is superimposed on another, and the display gradually 'wipes' between the two), to decisions based on static images. Experiment 1 used high quality image pairs, Experiment 2 used disguised target images, and Experiment 3 used degraded target images. Across all three experiments, rather than optimising performance, facial wipes reduced accuracy relative to static presentations. Further, there is evidence that video wipes increase false positives, and therefore may increase the likelihood that images of two different people will be incorrectly judged to show the same individual.

Facial image comparison by video superimposition

Introduction

Debates surrounding the admissibility of expert evidence based on forensic science techniques are ongoing in both UK (Law Commission of England & Wales, 2009; 2011) and US jurisdictions (National Academy of Sciences (NAS), 2009). In the UK, following the ruling in *R v Turner* (1975), the law requires only that expert testimony provides information that is likely to be outside the experience or knowledge of the jury and that the witness is appropriately qualified, through either study or experience, to provide this information. The role of the judge as a gatekeeper is thus limited and does not address issues regarding the validity or reliability of the science that informs the expert's opinion (Wheate & Jamieson, 2009). This offers the advantage of allowing courts to take account of advancements in scientific knowledge (*R v Clarke*, 1995). However, it also increases the likelihood that an expert will be permitted to present opinion evidence based on untested or unreliable techniques that do not have the strong empirical basis to justify their classification as a Forensic Science (e.g. Campbell-Tiech, 2005; Edmond, Biber, Kemp & Porter 2009; Edmond et al., 2010; NAS, 2009; Wheate & Jamieson, 2009).

Legal challenges against convictions which rested on evidence based on novel scientific techniques, such as ear print comparison (*R v Dallagher*, 2002), lip reading (*R v Luttrell*, 2004), and voice recognition (*R v Robb*, 1991), have highlighted the dangers of this permissive approach to the admissibility of expert evidence in the UK courts. This has contributed to a re-evaluation of the current legislation regarding expert evidence and in proposals for reform (Law Commission, 2009). In the USA, a report by the NAS identifies that many of the difficulties with forensic science evidence stem from the lack of scientific research into the validity of the techniques used (NAS, 2009). The current paper contributes to this debate by presenting empirical evidence regarding the use of the video superimposition or facial wipe technique, which is used to establish identity based on facial

Facial image comparison by video superimposition

images captured at the crime scene. This forms part of a range of tools used in a discipline referred to as 'Facial Image Comparison' or 'Facial Mapping', which currently lacks a scientific evidence base to support its use (Campbell-Tiech, 2005; Edmond et al., 2009).

Where CCTV images captured from a crime scene are poor quality, the UK courts allow for an expert with skills in 'Facial Image Comparison' to make comparisons between images from the crime scene and images of the defendant, and to provide the jury with opinion evidence of identity based on the results of this analysis (Attorney General's Reference, 2003). Part of the difficulty in establishing a body of scientific evidence relating to this branch of the forensic sciences is that the training and background of the experts engaged to conduct Facial Image Comparison analysis is varied, and the techniques employed are not clearly defined (ACPO/NPIA, 2009; Campbell-Tiech, 2005). Empirical investigation of this discipline has so far been limited and only a small number of the individual techniques have been assessed.

Kleinberg et al. (2007) explored the use of facial anthropometry, whereby distances and angles between facial features are compared across images, as a means of matching identities from facial images. They found that the accuracy of identity decisions in a 1-in-10 face matching task was around 25% using anthropometric measures, versus 75% accuracy when decisions were based on visual inspection by humans, suggesting the method has little to offer in terms of optimising accuracy. Davis et al. (2010) used a more sophisticated, computer-assisted approach to anthropometry, but even with high quality images and a small database of faces, there were a number of failures to match correctly, leading the researchers to urge caution in the use of this technique. Further, in a review of anthropometric techniques, Davis, Valentine, & Wilkinson (2012) highlight the difficulties associated with their application, and warn against basing a conviction on this type of analysis in the absence of other corroborating evidence.

Facial image comparison by video superimposition

Another group of techniques used by practitioners of facial image comparison involve superimposing or combining two images in order to check for correlation between the two. One such technique involves the creation of chimeric image composites, which are used to compare similarities and differences across two images, by combining the left half of the target face image with the right half of the suspect image, or vice versa. This procedure is used to assess how well the facial images match up, with the idea that aligning the images in this way draws attention to any differences or similarities between them (e.g. Oxlee, 2007). A Chimeric image of this type was presented in the case brought against the Metropolitan Police for a breach of the Health & Safety at Work etc. Act (1974), which concerned the events leading up to the tragic death of Jean Charles de Menezes. Jean Charles De Menezes was shot after police mistook him for a terrorist suspect who was sought by the police in connection with the failed 21st July London bombings. During this trial, the defence argued that the wanted man, Hussein Osman, and Jean Charles de Menezes were similar in appearance, and a chimeric image was used to demonstrate similarities between the two men. Subsequently, Strathie, McNeill & White (2012) conducted a series of experiments comparing the accuracy of face matching decisions based on full-face and chimeric image composites. Results showed that chimeric image presentations reduced Sensitivity (d') and occasioned a shift in Criterion (C) towards a liberal bias relative to full-face presentations. In short, the chimeric image composite technique appears to increase the likelihood of error by falsely categorising images of two different faces as a match. It was theorised that chimeric image composites elicited this effect because they created the perception of a new face, engaging holistic processing (Hole, 1994; Young, Hellawell & Hay, 1987), which made it more difficult to separate the constituent face halves.

Facial image comparison by video superimposition

An additional limitation of chimeric image composites is that participants see only half of each face, decreasing the facial information on which they can base their decisions. A more developed version of the composite face technique, video superimposition (or video wipes), can also be employed to demonstrate facial similarities and differences in a legal context (e.g. *R v Ciantar*, 2005, *R v Clarke*, 1995, Vanezis & Brierley, 1996, Oxlee, 2007). In this technique, one image is superimposed on top of the other, and the display gradually ‘wipes’ between the two. Video superimposition techniques offer the advantage of displaying the entirety of each of the faces under comparison, but like chimeric image composites, these techniques show the two faces under comparison as an aligned whole, raising the possibility that this presentation format may also increase errors. The use of video superimposition techniques in court has been contested as likely to mislead (*R v Clarke*, 1995), however, the appeal court upheld its use, with the judge stating that the technique is a ‘species of real evidence’ (*R v Clarke*, 1995). Facial Mapping practitioners themselves are mindful that video wipes can have a misleading effect on perception; however, they believe that with their experience and knowledge of the risks they can take account of this when using the technique to identify facial similarities or differences, (*R v Ciantar*, 2005). Whether they are correct remains to be seen. As the technique is also used to demonstrate facial similarities to the jury in court, if it does have a misleading effect on perception, jurors have no such protective experience, and as such may be particularly vulnerable to the effect. Despite legal concerns about the technique, (e.g. *R v Ciantar*, 2005, *R v Clarke*, 1995) there is currently no mandatory requirement for judges to provide a warning regarding the potentially misleading visual effect, though some have chosen to do so (e.g. *R v Ciantar*, 2005). It is this second use of the technique that the current paper seeks to address.

The video wipe display used in an applied setting (e.g., Oxlee, 2007) has features in common with presentation methods used in change detection studies (e.g. Rensink, 2002;

Facial image comparison by video superimposition

Simons & Rensink, 2005), which have found that people can fail to notice large changes to a visual display, such as an object appearing and disappearing, when a blank screen is briefly presented between the alternating images (e.g. Rensink O'Regan, & Clark, 1997). Research by David, Laloyaux, Devue, & Cleeremans (2006) demonstrates that the effect also occurs with gradual changes made to faces, even when the face is a central object of interest.

Furthermore, David et al. (2006) showed that gradual changes in facial expression were more difficult to detect than instant changes in expression that occur during a gap. These findings have some bearing on the current investigation into video wipes, as they suggest that the gradual changes that occur as the display wipes between the two images may increase the difficulty of discriminating between identities. The gradual change may operate to create visual continuity between the two faces, making the differences more difficult to detect. The resultant errors may be underpinned by the same mechanisms that result in change blindness.

However, the task of deciding if two images show the same person or two different people, is different from the task of detecting whether a change has taken place somewhere in an image, and may require different processing strategies. The facial image *always* changes in the current investigation, as all trials consist of two different images, taken by two different cameras, even if both images show the same identity. As such, a different criterion may be used, as participants must allow for a certain amount of variation on all trials. If video superimposition techniques do create visual continuity between two facial images, then these presentations are likely to evoke errors in a similar pattern to that observed with chimeric images. Therefore, this study seeks to establish the effect that video superimposition techniques have on face matching accuracy, with the aim of simulating the manner in which the technique is employed in court, and it is predicted that: (1) dynamic video superimposition displays will reduce accuracy relative to static side by side image displays and (2) image pairs viewed in video superimposition displays will elicit a bias towards same

Facial image comparison by video superimposition

responses.

Experiment 1

The first experiment sought to establish the baseline effect of facial wipe presentations on face matching accuracy, using high quality full-face images. In video superimposition, the images are aligned before one image is superimposed on the other, so the display consists of one entire image at the beginning and end of a transition. As the display slowly wipes back and forth between the two images, at each intermediary stage only one face appears on screen, made up of each of the two facial images in varying proportions. As each of the intermediary stages consists of a chimeric image created by a combination of the two photos, it is predicted that exposure to these composite faces will influence face matching decisions in a similar manner to static chimeric images (Strathie et al., 2012).

With Wipe presentations, participants are exposed to a complete view of both faces, and the display alters along both vertical and horizontal axes. This means there are likely to be more areas of difference between the two images than when only half of each face is presented. By increasing the number of available cues on which participants can base unfamiliar face matching decisions, the Wipe presentations may counteract any biasing effect associated with viewing a series of chimeric images as the display transitions between the two face images. However, in the current investigation it was predicted that the movement would help maintain visual continuity, and as with chimeric images, this would make it difficult for participants to separate the images into two separate identities.

Facial image comparison by video superimposition

Method

Participants

Twenty-eight students at Glasgow Caledonian University, seven of whom were male, participated in the experiment and received payment for their time. Participants were aged between 18 and 47 ($M = 24.5$, $SD = 7.1$), and all had normal, or corrected-to-normal, vision.

Design

The experiment utilised a 2 x 2 within participants design. The first independent variable was presentation format, with two levels, Still (two static faces), and Wipe (dynamic image presentation wiping between two faces). The second independent variable was Trial Type with two levels, Same or Different. The dependent variable was the accuracy of the same/different responses, which was measured in the form of sensitivity (d') and criterion (C) scores.

Materials

Thirty-two target face images were selected from the Glasgow Unfamiliar Face Database (GUFDB), (Burton, White, & McNeill, 2010) which contains images of 304 individuals between the ages of 17 and 60 years old. Within this database there are several images of each individual, captured using three different cameras. To eliminate potentially helpful transient cues such as lighting and camera focal length, image pairs were constructed using images from different cameras (for details see, Burton et al., 2010). For the mismatched pairs, the target image was paired with a photo of a different person who had been judged to have a similar appearance. The similarity ratings were obtained using a card-sorting procedure developed by Bruce, Henderson, Greenwood, Hancock, Burton, & Miller (1999). Full details of the database and of this procedure, are provided in Burton et al. (2010). For each of the targets in the current study, two face pairs were selected: a matching pair, consisting of two

Facial image comparison by video superimposition

different images of the target person, and a mismatched pair, where the target image was paired with an image of a different person.

To create the stimuli for the experimental conditions the original full-face images were modified as follows: Gridlines were placed so that the two images could be aligned and scaled to roughly the same proportions, and the two images were then placed side-by-side to form the pairs for the Still condition. To produce the pairs for the video wipe condition, the images were layered one on top of the other, and animated using Final Cut Pro software, so that the display gradually wiped back and forward between the two images on both horizontal and vertical axes. The animation took 60 seconds to complete. During this time, the video transitioned between the two faces on both vertical and horizontal planes, following the pattern A-B-A-B-A-B, with each transition lasting 12 seconds. The use of video superimposition techniques varies between practitioners and there are no prescribed guidelines on this, but as the idea of the technique is to allow for comparison of similarities and differences, the speed of the display was chosen as it enabled people sufficient time to view the transition between the two images clearly. All stimuli were presented in greyscale and were presented with each of the faces measuring approximately 400 pixels high. An illustration of the wipe stimuli, showing frames captured from the transitions in a wipe presentation can be seen in Figure 1.

Figure 1 about here please

Figure 1: Example of frames from a mismatched video wipe transition in Experiment 1.

Facial image comparison by video superimposition

Procedure

The experiment was administered via a Macintosh laptop (screen resolution 1280 x 800 pixels) running Psyscope experimental software (Cohen, MacWhinney, Flatt, & Provost, 1993). Participants sat approximately 20" away from the screen. Each trial commenced with a fixation cross that was displayed on screen for 1 second before stimuli onset. In each trial, participants were instructed to decide whether the two facial images showed the same person, or two different people, and to respond via a keypress. Each participant viewed 8 trials in each condition, making a total of 32 trials. The stimuli sets were counterbalanced across participants so that each image pair appeared in each condition with the same frequency. The video wipe trials took 60 seconds to display, and participants waited until the cycle was complete before they made their decision. For parity, it had been planned that the still images should also be displayed for 60 seconds, but pilot testing suggested that forcing participants to wait 60 seconds to respond caused frustration, therefore in the control condition participants were allowed to respond before the full time had elapsed. The shorter presentation time here for still images is justified because this occurs when making such identification decisions in real-life settings. The trials were blocked by presentation format (Wipe/Still), and the blocks were presented in random order.

Results

Table 1 summarises mean percentage accuracy broken down by presentation format and trial type. The responses across same and different conditions were processed using signal detection theory (SDT) (Macmillan & Creelman, 2005; Green & Swets, 1966) to form a measure of Sensitivity (d') and criterion (C) for each participant, at each level of presentation format. Sensitivity (d') and criterion (C) were compared across levels of presentation format using two separate paired t-tests. For sensitivity, there was a significant effect of presentation

Facial image comparison by video superimposition

format, $t(27) = 3.65, p < .05, d = 0.693$ with significantly higher mean sensitivity in the still condition ($M = 3.58, SD = 1.01$) than in the wipe condition ($M = 2.7, SD = 1.16$).

Table 1 about here please

A paired t-test conducted on criterion scores revealed a significant difference in response bias between Still ($M = 0.10, SD = 0.69$), and Wipe ($M = -0.17, SD = 0.55$) conditions, $t(27) = 2.01, p < .05, d = 0.38$. The slight positive criterion value in the Still condition ($M = 0.10$) suggests participants exhibit a small bias towards 'different' responses in this condition, while the negative criterion value obtained in the Wipe condition ($M = -0.17$) suggests participants have a tendency towards 'same' responses in this condition. To explore this further, two separate, one sample, t- tests were conducted to compare criterion scores against a chance score of zero (i.e. no bias). Relative to zero, no response bias was observed in either the Still, $t(27) = .74, p > .05$, or Wipe, $t(27) = -1.68, p > .05$, conditions.

Discussion

As predicted, participants performed better with Still presentations than with Wipe presentations. The significant effect of presentation format (Still/Wipe) on sensitivity (d') confirms the prediction that video wipes reduce, rather than optimise, the accuracy of face matching decisions. Presentation format also evoked a different pattern of response bias. Video wipes elicited a more liberal response bias (more 'same' responses) than still presentations. However, although they differed from each other, neither condition had a response bias that differed from zero.

Looking first at the effect of presentation format, participants typically viewed the trials for longer in the Wipe condition than in the Still condition, and it might be expected that this longer exposure time would confer an advantage. However, the results suggest otherwise as, despite the longer exposure times, sensitivity was higher with Still

Facial image comparison by video superimposition

presentations. If the same processes used in unfamiliar face memory tasks also underpin unfamiliar face matching, forcing participants to view the images for a longer time period before responding may have encouraged them to adopt a different processing strategy, overriding their initial, more accurate responses. For example, there is some evidence from face memory research that given sufficient time, participants can deliberately adopt a featural processing strategy at encoding, and that this feature-based strategy results in lower rates of recognition than an holistic processing strategy (e.g. Wells & Hryciw, 1984).

Further, research by Megreya & Burton (2006, 2007) suggests that when matching unfamiliar faces, participants rely more on featural, rather than holistic, processing. Based on their results, if longer viewing times do elicit a featural strategy this should be beneficial in a matching task. Additionally, during a wipe presentation, the whole of the first facial image is seen only at the beginning of each transition, and the entirety of the second facial image is seen only at the end of a transition. Consequently, participants are exposed to each of these whole faces for only short periods. For much of the trial, participants view a face that consists of portions of each of the two images under comparison in varying proportions. As such, it may be difficult to concentrate on an individual facial feature in order to compare the faces feature-by-feature. The purpose of the wipe presentation is to allow people to examine how well the faces align, with the line moving across the image and encouraging viewers to focus on the point of intersection as the line moves across the face. This does not seem compatible with a feature based matching strategy. Therefore, it seems most likely that the reduction in accuracy with wipe presentations results from forcing participants to process the faces holistically, and impeding their ability to use the feature-based approach that previous research (e.g. Megreya & Burton, 2006, 2007) indicates may be a more effective strategy in an unfamiliar face matching task.

Facial image comparison by video superimposition

Importantly, the pattern of responding obtained in the current experiment, with video wipes, is similar to that found with chimeric images (Strathie et al., 2012), suggesting that both techniques evoke a similar processing strategy and that this impedes, rather than optimises, face matching decisions.

Experiment 2

The results of Experiment 1 established that Wipe presentations reduce matching accuracy relative to Still image presentations with high quality target images. However, Facial Image Comparison techniques are most likely to be used in court when the culprit's facial appearance is unclear in the CCTV footage, due to low image quality or culprit disguise, and it may be that Wipe presentations are advantageous in these conditions. To address this issue Experiment 2 uses more forensically realistic stimuli. The target (culprit) faces simulated the images that might be captured from a crime scene and showed the target wearing a disguise. The photographs of the suspect simulated the images that might be obtained of a suspect in custody and were undisguised.

In live person to photo matching, Davis & Valentine (2009) found participants were more accurate with unfamiliar face matching decisions when the target faces were shown wearing glasses than when they were shown undisguised. This finding may be driven by the external feature matching strategy that is thought to exemplify unfamiliar face matching (Clutterbuck & Johnston, 2002; Young et al., 1985). By occluding some of the internal features of the face, a disguise may focus attention on the external features of the face, thereby underlining the use of this strategy. The disguised faces used in the current experiment were wearing glasses *and* a hood, so both external and internal features were occluded. As such, an external feature strategy would not be useful in the current task.

The provision of expert evidence on Facial Image Comparison where the image of the target is degraded or disguised is founded on the premise that these techniques are

Facial image comparison by video superimposition

particularly useful under these conditions. Therefore, under the circumstances in which facial wipe techniques are employed in an applied setting, with disguised target images, wipe presentations would be expected to increase matching accuracy relative to still image presentations. However, based on the results of the previous experiment, the opposite result is expected. Furthermore, compared to undisguised target images, the disguise is predicted to evoke more false positive responses. The rationale for this prediction is that with less of the perpetrator image visible, there are likely to be fewer points of conflict with the appearance of the suspect image, regardless of whether they show the same person or not.

Method

Participants

Twenty-eight students at Glasgow Caledonian University, nine of whom were male, participated in the experiment, and received payment for their time. Participants were aged between 19 and 51 ($M = 31.9$, $SD = 10.7$), and all had normal or corrected-to-normal vision.

Design & Materials

The design was the same as for Experiment 1. Thirty-two target face images were captured from video clips contained in the GUFID (Burton et al., 2010). The images show a frontal view of the target's face wearing a hood and dark glasses. Thirty of these were the same target identities used in Experiment 1, but as a disguised image was unavailable for two of the original target identities, these two were replaced with alternative identities. As in the previous experiment, two face pairs were created for each target, a same pair and a different pair. The high quality still images that were used to create the pairings were undisguised.

The stimuli were created in the same way as in experiment 1, with two exceptions. To ensure that the movement in the wipe condition was not responsible for any difference in

Facial image comparison by video superimposition

performance, the control condition was modified from the format of Experiment 1. After the images in the pair were placed side-by side, Final Cut Pro was used to add a thin black line that moved across the display in the same pattern as in the wipe condition, without alteration to the image displayed. Also, animations for both the control and video wipe conditions took 32 seconds to display to ensure parity across conditions. The wipe trials were shortened by reducing the number of transitions to four (two vertical and two horizontal), and altering the speed of the transitions so that each lasted 8 seconds. Pilot testing suggested that these shorter transitions still allowed participants time to view the two images under comparison. All stimuli were presented in greyscale, and were presented with each of the faces measuring approximately 400 pixels high. An example showing still frames captured from a moving wipe trial can be seen in Figure 2.

Fig 2 about here please

Figure 2: Example of frames from a mismatched wipe pair from Experiment 2

Procedure

The experimental procedure was the same as for Experiment 1, save that all trials in both conditions took 32 seconds to display and participants waited until this cycle was complete before they made a decision.

Results

Table 2 summarises mean percentage accuracy broken down by presentation format and trial type. The responses across same and different conditions were processed using SDT to form a measure of Sensitivity (d') and criterion (C) for each participant, at each level of

Facial image comparison by video superimposition

presentation format. Sensitivity (d') and criterion (C) were compared across levels of presentation format using two separate paired t-tests. For sensitivity, a significant effect of presentation format was found, $t(27) = 3.15$, $p < .05$, $d = 0.61$, with higher mean sensitivity in the Still condition ($M = 2.89$, $SD = 1.26$) than in the Wipe condition ($M = 2.32$, $SD = 1.07$).

Table 2 about here please

Criterion scores were also examined, and a significant effect of presentation format was found, $t(27) = 2.26$, $p < .05$, $d = 0.44$, revealing a difference in response bias between Still ($M = 0.12$, $SD = 0.49$), and Wipe ($M = -0.10$, $SD = 0.61$) conditions. The positive criterion value in the Still condition ($M = 0.12$) suggests that participants exhibit a bias towards 'different' responses in this condition, while the negative criterion value obtained in the Wipe condition ($M = -0.10$) suggests participants are biased towards responding 'same' in this condition. Two separate, one sample, t-tests were conducted to compare criterion scores against a chance score of zero (i.e. no bias). Relative to zero, no response bias was observed in either the Still, $t(27) = 1.34$, $p > .05$, or Wipe, $t(27) = -.88$, $p > .05$ conditions.

Discussion

The significant effect of presentation format on sensitivity suggests that matching performance in the Wipe condition was significantly poorer than in the Still presentation condition. The pattern of results is similar to that obtained in Experiment 1 with undisguised target images, suggesting that wipe presentations also impair matching accuracy relative to still image presentations with disguised target images. Although the difference in response bias suggests that participants were more liberal with Wipe trials than with Still trials, the bias did not differ from zero in either condition.

Facial image comparison by video superimposition

Participants in the current experiment viewed trials in each of the two conditions for equal lengths of time. This confirms that longer exposure times do not explain the reduction in accuracy on different trials observed with Wipe presentations in Experiment 1. This adds more weight to the view that it is the processing mode evoked by video wipes that reduces the accuracy of matching decisions, by encouraging participants to view the moving image as a single entity (thereby increasing 'same' responses). This explanation is in keeping with Wilford & Wells' (2010) assertion that holistic processing for faces reduces change-localisation abilities. Although presentation times in this experiment are longer than those in Wilford & Wells' (2010) study, the way in which video wipe presentations display the faces may still encourage participants to rely on holistic processing, thus making them poorer at localising differences between two faces.

Alternatively, it might be argued that in the same identity Wipe trials, when one of the faces is disguised, the overall similarity between the two images is reduced, thereby interfering with holistic processing. This would result in a tendency towards false negative responses (saying two images of the same person are different). However, the reverse pattern was observed in the current study, with analysis of criterion scores demonstrating participants were more liberal (e.g. more likely to respond 'same' with Wipe trials than with Still trials). This is the same pattern of responses observed with undisguised images in Experiment 1.

It should be noted that these findings contrast with those of Davis & Valentine (2009), who found increased accuracy when participants made matching decisions with disguised target images. In the current experiment, the disguised images did not facilitate performance. Accuracy in the Still condition (88%) was comparable to that in Experiment 1, and to performance in previous full face matching studies that used the same 2AFC design (e.g. Burton et al., 2010; Megreya & Burton, 2006, 2007). However, the benefit observed by

Facial image comparison by video superimposition

Davis & Valentine was for images where only the eyes were covered and this may account for the failure to find any benefit here.

In summary, the results of the current experiment extend upon the results of Experiment 1 to show that Wipe presentations reduce the accuracy of unfamiliar face matching decisions when target faces are disguised, as well as when they are undisguised. Facial Image Comparison techniques are most likely to be employed where the perpetrator image is disguised or the image is of low quality, so the finding that video wipes reduce performance with disguised, as well as, undisguised targets, is of applied importance.

Experiment 3

In Experiments 1 and 2, using high quality and disguised images, face matching accuracy was lower with video wipe presentations than with still image presentations. In a court setting, the services of a Facial Image Comparison practitioner are most likely to be engaged where the images from the crime scene do not yield a clear view of the offender's face, due to either disguise or poor image quality. The use of disguised target images in Experiment 2 replicated the first of the scenarios in which a Facial Comparison practitioner is likely to be employed to present evidence in a criminal case, and the current experiment explores the effectiveness of the technique in the second scenario, by using degraded quality target images. The aim here was to simulate poor quality CCTV images and images were pixelated to achieve this effect (Bindemann, Attard, Leach, & Johnston, 2013; Demanet, Dhont, Notebaert, Pattyn, & Vandierendonck, 2007).

Method

Participants

Thirty participants who were staff or students at Glasgow Caledonian University, took part in

Facial image comparison by video superimposition

the experiment, and received payment in return for their time. The participants were aged between 17 and 50 ($M = 26.5$, $SD = 10.8$), and eight of the thirty were male. All had normal or corrected-to-normal vision.

Design & Materials

The design was the same as in Experiments 1 and 2. The 64 image pairs (32 same, 32 different) used in Experiment 2 (image size 400 x 600 pixels, 100dpi) were modified to provide the stimuli for this experiment. Using Adobe Photoshop, the quality of the target image in each face pair was degraded in a standardised fashion, by increasing the pixel size to cell size 5 with the mosaic setting, and applying a 10 pixel motion blur at an angle of 5° . The high quality images that were paired with the target images remained unaltered, as these represent the images that would be obtained of a suspect in custody. Final Cut Pro software was used to animate the images. The format of the video wipes replicated those in Experiment 2, with the animation taking 32 seconds to gradually wipe back and forward between the two images on both horizontal and vertical axes. The still images were animated as in Experiment 2. All stimuli were presented in greyscale, and were presented with each of the faces measuring approximately 400 pixels high. An illustration of frames from a wipe trial can be seen in Figure 3.

Procedure

The procedure was the same as for Experiment 2.

Figure 3 about here please

Figure 3: Example of frames from a matching wipe trial from Experiment 3.

Results

Table 3 summarises mean percentage accuracy broken down by presentation format and trial type. The responses across same and different conditions were processed using SDT to form a measure of Sensitivity (d') and criterion (C) for each participant, at each level of presentation format. Sensitivity (d') and criterion (C) were compared across levels of presentation format using two separate paired t-tests. Results show that mean sensitivity in the Still condition ($M = 2.5$, $SD = 1.06$) is similar to sensitivity in the Wipe condition ($M = 2.25$, $SD = 1.05$), and the effect of presentation format is not significant, $t(29) = 0.89$, $p = 0.19$, $d = 0.45$.

Table 3 about here please

A paired t-test conducted on criterion scores revealed a significant effect of presentation format, $t(29) = 2.75$, $p < .05$, $d = 0.5$ with a difference in response bias between Still ($M = 0.43$, $SD = 0.54$), and Wipe ($M = 0.09$, $SD = 0.6$) conditions. The small positive criterion value in the Wipe condition, (0.09) suggests that participants exhibit a very slight tendency towards 'different' responses in this condition, while the larger positive criterion value obtained in the Still condition, (0.43) suggests participants showed a greater bias towards

Facial image comparison by video superimposition

responding 'different' in this condition. Two separate, one sample, t-tests were conducted to compare criterion scores against a chance score of zero (i.e. no bias). Relative to zero, no response bias was observed in the Wipe condition, $t(29) = .82, p > .05$, but there was a significant bias towards 'different' responses in the Still condition, $t(29) = 4.31, p < .05$.

Discussion

There was no significant effect of presentation format (Wipe/Still) on sensitivity (d') in the current experiment, with similar values across both conditions. However, the criterion scores show that there is a more conservative response bias (a greater tendency to make 'different' responses) with Still presentations than with Wipe presentations.

In this experiment, face matching accuracy in the Still condition (approximately 83%) appears slightly lower than with high quality, full face pairs in the equivalent condition in Experiments 1 (92%) and 2 (88%), which is consistent with the lower quality of the images used here. However, there are large individual differences in unfamiliar face matching performance and, accuracy in each experiment is within the range identified in previous research studies, (e.g. Burton et al., 2010) supporting the assertion that this task is highly error-prone (e.g. Bruce et al., 1999; Burton et al., 2010; Kemp, Towell & Pike, 1997; Megreya & Burton, 2006, 2007) especially where the target images are poor quality (Henderson et al., 2001).

The prediction that, as in Experiments 1 and 2, video wipe presentations would reduce matching accuracy relative to still image matching was not supported by the data. There was no significant difference in matching accuracy across these two conditions, and the pattern of results in this experiment differed from those observed in Experiments 1 and 2. However, although Wipe presentations were no poorer than Still presentations, they once again failed to offer an advantage over Still presentations, which is the intended purpose of this technique. Therefore, the results of this experiment support the conclusions drawn in Experiments 1 and

Facial image comparison by video superimposition

2, in so much as they fail to offer any support for the use of this presentation technique in court.

This pattern of results may suggest that participants are particularly cautious with Still image presentations when the images are degraded. This interpretation is supported by the criterion scores, which showed a strong conservative bias in the Still condition. It is possible that the reduced information provided by the degraded images encouraged participants to adopt a more conservative response strategy overall with criterion shifted towards conservative responding across conditions. In Experiment 1, Wipe presentations with high quality target images evoked a liberal bias, and Still presentations a very slight conservative bias relative to each other. In the current experiment, the mean criterion scores showed a conservative bias for both conditions (though only the Still condition differed from zero), and this may reflect an increase in ‘different’ responses across presentation formats due to the nature of the degraded stimuli.

Although sensitivity was similar across conditions, the results of this experiment provide further evidence that Wipe presentations do not increase the accuracy of matching decisions. Instead, they appear to alter criterion, evoking a less conservative response bias than still image presentations.

Main Discussion

The aim of this study was to explore the effect of video superimposition techniques on face matching accuracy. Experiment 1 investigated the baseline effect of video wipe presentations, where one image is superimposed on a second and the display gradually wipes back and forth between the two, on the accuracy of unfamiliar face matching decisions. The results demonstrated that video wipes reduced sensitivity and induced a liberal bias relative to static full face image matching.

Facial image comparison by video superimposition

Experiment 2 replicated this effect with more forensically realistic stimuli, which showed the target face wearing a disguise. Facial Image Comparison techniques are more likely to be employed within an applied setting to facilitate matching decisions with more challenging target images, yet the results showed that video wipes also resulted in lower sensitivity than still image presentations with disguised target images. The pattern of responses obtained was similar to that obtained in Experiment 1, with wipe presentations producing a more liberal response bias than still image presentations.

Experiment 3 extended upon these results by examining the effect of video wipes with degraded quality target images. The results of this experiment differed from the previous two in that there was no significant effect of presentation format (Wipe/Still) on sensitivity. There was a significant effect on criterion scores, with still images evoking a more conservative response bias. Although Experiment 3 did not find the same advantage for sensitivity with still presentations observed in Experiments 1 and 2, the aim of using video wipes in court is to increase matching accuracy relative to static presentations. As there was no benefit for wipe presentations relative to still presentations in any of the three experiments, the results fail to offer any support for the continued use of this technique in court.

Overall accuracy for unfamiliar face matching with still images across the three experiments in this study (between 83 - 94 %) was comparable with accuracy levels in previous face matching studies (e.g. Burton et al., 2010; Megreya & Burton, 2006, 2007). This reinforces the claim that even in optimal conditions people make a high proportion of errors in a simple unfamiliar face-matching task. Importantly, the results of these three experiments failed to find any evidence that the use of video superimposition techniques reduces this error rate. Instead, in two of the three experiments the video presentation techniques actually reduced sensitivity. Furthermore, criterion values suggest that wipe presentations evoke a less conservative response bias than still image presentations.

Facial image comparison by video superimposition

Theoretically, the results support the view that, like static chimeric images (Strathie et al., 2012), video superimposition techniques that alternate two images without visual disruption create a visual continuity that increases 'same' responses in an identity matching task. Research on change blindness has established that gradual changes to visual displays are difficult to detect (e.g. David et al., 2006). Similarly, accuracy in the Wipe experiments, where the display transitions slowly between two images, was impaired relative to accuracy with Still presentations. However, in wipe experiments the image changes on every trial, regardless of whether the identity changes, and the task of deciding if two images show the same person differs markedly from the task of simply detecting if any change has occurred.

In sum, the results do not support the use of video superimposition techniques as a means of optimising identification from images, and across three experiments the techniques failed to produce any increase in accuracy relative to static presentations. In two of the three experiments, the techniques produced a bias towards 'same' responses, which has potentially serious consequences in an applied context as this suggests that the technique increases the likelihood that a video wipe consisting of images of two different people will be judged as showing the same person. These results provide empirical evidence that video superimposition techniques do not provide an advantage when making identity matching decisions and should not be provided as identification evidence in court. Additionally, the research contributes to wider debates about the admissibility of expert evidence based on novel forensic techniques which lack a strong base of empirical evidence (e.g. Law Commission, 2009; 2011; NAS, 2009; Wheate & Jamieson, 2009), highlighting the importance of testing the science that informs forensic opinion, in order to ensure that the information it provides is scientifically sound, and that testimony based on such techniques qualifies as expertise.

Facial image comparison by video superimposition

References

- Association of Chief Police Officers (ACPO)/National Police Improvement Agency (NPIA), (2003). *Facial Identification Guidance*. Retrieved May 30 2014, from: <http://www.acpo.police.uk/documents/crime/2009/200911CRIFIG01.pdf>
- Attorney General's Reference. [no.2 of 2002], (2003). 1 CR. App. R. 321, England.
- Bindemann, M., Attard, J., Leach, A., & Johnston, R. (2013). The effect of image pixelation on unfamiliar face matching. *Applied Cognitive Psychology*, 27, 707-717. doi:10.1002/acp.2970.
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P., Burton, A.M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5, 339-360.
- Burton, A.M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, 42, 286-291.
- Campbell-Tiech, A. (2005) "Stockwell" revisited: the unhappy state of facial mapping. *Archbold News*, 6, 24 June, 4-6.
- Clutterbuck, R., & Johnston, R.A. (2002). Exploring Levels of Face Familiarity by Using an Indirect Face-Matching Measure. *Perception*, 31 (8). 985-994.
- Cohen J.D., MacWhinney, B., Flatt M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers*, 25, 257-271.
- David, E., Laloyaux, C., Devue, C., & Cleeremans, A. (2006). Change blindness to gradual changes in facial expressions. *Psychologica Belgica*, 46, 253-268.
- Davis, J.P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, 23(4), 482-505. DOI: 10.1002/acp.1490
- Davis, J.P., Valentine, T. & Davis, R.E. (2010). Computer assisted photo-anthropometric analyses of full-face and profile facial images. *Forensic Science International*, 200, 165-176. DOI: <http://dx.doi.org/10.1016/j.forsciint.2010.04.012>
- Davis, J.P., Valentine, T. & Wilkinson, C. (2012). Facial Image comparison. In: C. Wilkinson & C. Ryan (2012) *Craniofacial identification*. (pp. 136-153).Cambridge: Cambridge University Press.

Facial image comparison by video superimposition

- Edmond G., Biber K., Kemp R., Porter G., (2009). Laws Looking Glass: Expert Identification Evidence Derived from Photographic and Video Images. *Current Issues in Criminal Justice*, 20:3, 337-377.
- Demant, J., Dhont, K., Notebaert, L., Pattyn, S., & Vandierendonck, A. (2007). Pixelating familiar people in the media: Should masking be taken at face value? *Psychologica Belgica*, 47, 261–276.
- Edmond, G., Kemp, R., Porter, G., Hamer, D., Burton, M., Biber, K., & San Roque, M. (2010). Atkins v The Emperor: The cautious use of unreliable expert opinion. *The International Journal of Evidence and Proof*, 14, 146-166.
- Henderson, Z., Bruce, V., & Burton, A.M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, 15, 445-464. DOI: 10.1002/acp.718
- Hole, G.J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception* 23, 65-74.
- Kemp, R., Towell, N. & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*, 11, 211-222.
- Kleinberg, K.F., Vanezis, P., & Burton, A.M. (2007). Failure of anthropometry as a facial identification technique using high-quality photographs. *Journal of Forensic Sciences*, 52(4), 779-783.
- Law Commission (2011). *Expert Evidence in Criminal Proceedings in England and Wales Report No.325*. Retrieved 10 June 2013 from:
<http://www.justice.gov.uk/lawcommission/publications/expert-evidence.htm>
- Law Commission, (2009). *The admissibility of expert evidence in criminal proceedings in England & Wales: A new approach to the determination of evidentiary reliability, Consultation Paper No.190*. Retrieved 10 June 2013 from:
http://www.justice.gov.uk/lawcommission/docs/cp190_Expert_Evidence_Consultation.pdf
- Green D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York Wiley.
- Macmillan, N.A., Creelman, C.D. (2005). *Detection theory: A user's guide* (2nd edition). Mahwah (NJ): Erlbaum.
- Megreya, A.M. & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a

Facial image comparison by video superimposition

- matching task. *Memory & Cognition* 34(4), 865-876.
- Megreya, A.M. & Burton, A.M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception and Psychophysics*, 69, 1175-1184.
- National Academy of Sciences, (2009). *Strengthening Forensic Science in the United States: A Path Forward*. National Academies Press, Washington, D.C.
- Oxlee, G., (2007). Facial Recognition and Imagery Analysis. In: T. Thompson & S.M. Black (Eds.), *Forensic Human Identification: An Introduction*. Taylor and Francis: Florida, 257-270.
- R v Ciantar* (2005). EWCA Crim 3559.
- R v Clarke*, (1995). Cr. App. R. 425.
- R v Dallagher*, (2002). EWCA Crim 1903.
- R v Luttrell*, (2004). EWCA. Crim 1344.
- R v Robb* (1991) 93 Cr. App. R. 161.
- R v Turner* (1975) QB 834.
- Rensink R.A. (2002). Change detection. *Annual Review of Psychology*, 53, 245–277.
- Rensink R.A., O'Regan, J.K., and Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368-373.
- Simons, D.J., & Rensink, R.A. (2005). Change blindness: past, present, and future. *Trends in Cognitive Sciences*, 9(1), 16-20.
- Strathie, A., McNeill, A., & White, D. (2012). In the dock: Chimeric image composites reduce identification accuracy. *Applied Cognitive Psychology*, 26(1) 140-148.
DOI: 10.1002/acp.1806
- Vanezis, P., & Brierley, C. (1996). Facial image comparison of crime suspects using video superimposition. *Science and Justice* 36, 27-34.
- Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory and Cognition*, 12, 338-344.
- Wheate, R.M., & Jamieson, A. (2009). A Tale of Two Approaches – The NAS Report and the Law Commission Consultation Paper on Forensic Science, *International Commentary on Evidence*, 7(2:3), 1-25. DOI: 10.2202/1554-4567.1110
- Wilford, M. M., & Wells, G.L. (2010). Does facial processing prioritize change detection? Change-blindness illustrates costs and benefits of holistic processing. *Psychological Science*, 21, 1611-1615. DOI: 10.1177/0956797610385952
- Young, A.W., Hay, D.C., McWeeny, K.H., Flude, B.M. and Ellis, A.W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14, 737-

Facial image comparison by video superimposition

746.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configurational information in face perception. *Perception, 16*, 747–759.

Table 1

Mean percentage accuracy with standard deviations for Experiment 1.

<i>Presentation Format</i>	<i>Trial Type</i>	<i>Mean</i>	<i>SD</i>
Still	Same	90	15
	Different	92	18
Wipe	Same	89	12
	Different	82	18

Table 2

Mean percentage accuracy with standard deviations by conditions for Experiment 2.

<i>Presentation Format</i>	<i>Trial Type</i>	<i>Mean</i>	<i>SD</i>
Still	Same	86	13
	Different	89	16
Wipe	Same	85	12
	Different	79	20

Table 3

Mean percentage accuracy and standard deviations by condition for Experiment 3.

<i>Presentation Format</i>	<i>Trial Type</i>	<i>Mean</i>	<i>SD</i>
Still	Same	74	19
	Different	91	11
Wipe	Same	80	18
	Different	83	15

Facial image comparison by video superimposition