Research Article

FACE RECOGNITION IN POOR-QUALITY VIDEO: Evidence From Security Surveillance

A. Mike Burton,[1] Stephen Wilson,[1] Michelle Cowan,[1] and Vicki Bruce[2]

[1]*University of Glasgow, Glasgow, United Kingdom, and* [2]*University of Stirling, Stirling, United Kingdom*

Abstract—*Security surveillance systems often produce poor-quality video, and this may be problematic in gathering forensic evidence. We examined the ability of subjects to identify target people captured by a commercially available video security device. In Experiment 1, subjects personally familiar with the targets performed very well at identifying them, but subjects unfamiliar with the targets performed very poorly. Police officers with experience in forensic identification performed as poorly as other subjects unfamiliar with the targets. In Experiment 2, we asked how familiar subjects can perform so well. Using the same video device, we edited clips to obscure the head, body, or gait of the targets. Obscuring body or gait produced a small decrement in recognition performance. Obscuring the targets' heads had a dramatic effect on subjects' ability to recognize the targets. These results imply that subjects recognized the targets' faces, even in these poor-quality images.*

The psychological study of face recognition divides into two rather different topics. First, there are projects that focus on the recognition of faces previously unfamiliar to subjects (e.g., Brown, Deffenbacher, & Sturgill, 1977; Ellis, 1975; Laughery, Alexander, & Lane, 1971; for reviews, see Clifford & Bull, 1978, and Shepherd, Ellis, & Davies, 1982). Second, there is a large literature on processes underlying recognition of familiar faces (e.g., see reviews by Bruce, 1988, and Bruce & Humphreys, 1994; and theoretical developments by Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990; Burton, Young, Bruce, Johnston, & Ellis, 1991).

Studies of unfamiliar-face recognition often have a forensic motivation. In typical experiments, subjects are shown faces of unfamiliar people and are subsequently tested using a recognition memory procedure. It has been shown on a number of occasions that recognition of previously unfamiliar faces is rather poor (e.g., Yarmey, 1979). Despite these findings, juries are said to favor eyewitness face recognition reports, and attach considerable weight to them. It is therefore very important to establish the reliability of such reports across a range of conditions, and to discover techniques for improving the reliability of recognition (Shepherd et al., 1982).

Research in recognition of both familiar and unfamiliar faces very commonly uses high-quality images of target people. However, recent developments in security surveillance pose a particular problem with image quality. Small-scale security systems based on VHS video or closed-circuit television have become very common in Europe and North America. Such systems are often installed with little attention to optimizing lighting conditions or viewing angle. This means that when an image or video sequence is needed for evidence (e.g., following a crime), it is not always easy to confirm whether the person captured in the security device is the same person accused or suspected of the crime.

In experiments reported here, we examined human face recognition in poor-quality video images. In particular, we were concerned with the effects of familiarity on recognition ability. There were two main questions of interest. First, how good is face recognition in poor-quality images? To answer this question, we used video sequences captured from a commercially available video security system. Second, if subjects are able to recognize people from these images, what is the basis for their recognition?

EXPERIMENTAL SETTING

Both experiments reported here used images from the same security device, which was chosen to be typical of many low-cost security systems. The Department of Psychology at the University of Glasgow, United Kingdom, uses a video security system installed by a local company. This is a VHS video device that is triggered when a person approaches the main entrance of the building. Each time a person enters or leaves, a security light is automatically turned on, and about 4 s of video is recorded. The video camera is located on an inside wall directed toward the main entrance door, at a height of 9 ft. The equipment includes a vista NCD 340 CCTV camera (8 mm, f1.2 lens) with a Mitsubishi HS-5424E(B) Timelapse Cassette Recorder. The video is recorded on Fuji HQ+ 180 PAL VHS videotape.

Systems of this kind are very common in the local area, and the same security company supplies many local businesses. The system was not configured in any special way for the purposes of this experiment. Informal observations suggest that the resulting image quality is rather poor, though tolerable in a low-cost system. The left panel in Figure 1 shows a still from this system.

EXPERIMENT 1: THE EFFECTS OF FAMILIARITY

In the first experiment, we examined whether personal familiarity with the targets affected the ability of subjects to recognize images from the security system. Many of the people who walk through this particular security system are lecturing staff at the University of Glasgow. It is therefore relatively easy to find subjects who are familiar with them (i.e., students who take classes in psychology). Similarly, it is relatively easy to find subjects who are unfamiliar with these targets (i.e., students who do not take classes in psychology). In this experiment, we also examined the ability of a set of police officers to recognize the targets. The police subjects were unfamiliar with the targets, but were experienced in making identification judgments.

The experiment made use of a recognition memory procedure. In the first phase, subjects were shown a set of video sequences and told they would be asked to recognize people in these clips later. In the second phase, subjects were shown a set of high-quality photos, and asked whether the person in each photo appeared in the first phase. Although

Address correspondence to Mike Burton, Department of Psychology, University of Glasgow, Glasgow G12 8QQ, United Kingdom; e-mail: mike@psy.gla.ac.uk.

Fig. 1. Images of the type used in Experiment 1: a still from a video (left) and a photograph taken in good lighting (right).

this is not a direct analogue of the usual forensic situation (in which only one target would normally be involved), it is a convenient task to use experimentally because the same procedure can be used with subjects who are both familiar and unfamiliar with the targets. Performance on recognition memory tasks typically covaries with performance on other face recognition tasks, and this procedure has been used commonly in the past to compare familiar- and unfamiliar-face recognition with the same target stimuli (e.g., Bruce, 1982).

Method

The video clips chosen showed 20 members of the lecturing staff, 10 male and 10 female. These clips were taken from the routinely collected videos of people entering the building; they were not posed, and target people were not aware at the time that their video images would form part of an experiment. Clips that contained only one person entering the building were chosen. In addition, each target person was photographed on a different day, using a high-quality digital camera, under good lighting conditions. Examples of a still from a video and a high-quality photograph are shown in Figure 1.

Sixty subjects volunteered to take part in the study. Of these, 20 were students recruited from the Department of Psychology, and each had been taught by all 20 of the target lecturers. We refer to these students as the *familiar group.* An additional 20 students were recruited from different departments throughout the university; none of these students had taken courses in the Department of Psychology. Finally, the subjects included 20 police officers attending a course at a local police training school. These were experienced officers with an average of 13.5 years of service.

Subjects were tested individually in an experimental room and were shown video clips on a standard video recorder and television. They were initially shown 10 of the 20 video clips and told they would be asked to identify these people later. Each subject was shown these clips twice, each time in a different random order. There was a short gap (2–3 s) between clips, and a rest period of 1 min after the videos were viewed. The particular subset of videos shown in this phase was counterbalanced across subjects.

There followed a test phase in which subjects were shown each of the 20 high-quality images, one at a time. They were told that they would be shown 20 faces, and that half of these people had been present in the videos. They were asked to assign a rating of 1 to 7 to each of these photos. A score of 7 indicated that the subject was sure that the person appeared in the videos; a score of 1 indicated that the person definitely did not appear in the videos.

Results and Discussion

Figure 2 shows the mean recognition scores given to the seen and unseen targets. People familiar with the targets performed well, assigning high scores to seen targets and low scores to unseen targets. Subjects in the other two groups performed less well, making a smaller discrimination between these two target groups. Formal analysis showed that all groups scored seen targets significantly higher than unseen targets, though the effect was much larger in the familiar group. There was no difference in performance between the unfamiliar student and police groups, but both performed significantly more poorly than the familiar group. Two-way analyses of variance showed no

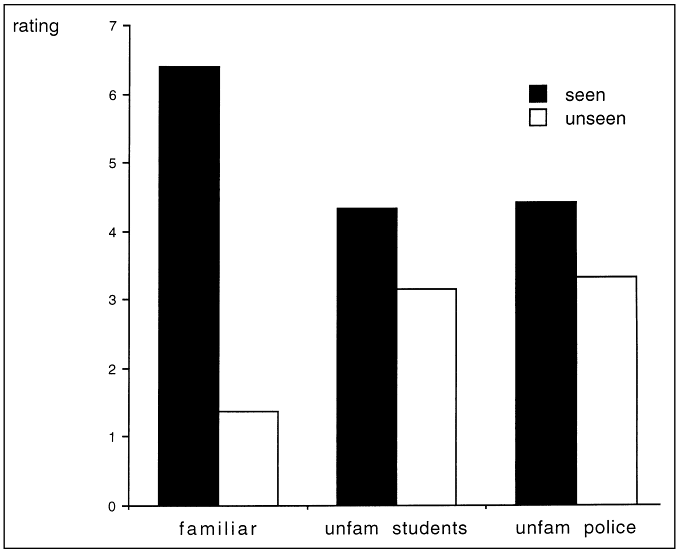

Fig. 2. Accuracy of identification in Experiment 1. Subjects who were familiar and unfamiliar ("unfam") with the targets rated whether they had seen the targets in previously viewed videotapes. A rating of 7 indicates certainty that the target had appeared in the videos, and a rating of 1 indicates certainty that the target had not appeared.

main effect of subject group, $F(2, 57) < 1$; a significant effect of seen versus unseen target, $F(2, 57) = 324$, $p < .001$; and a highly significant interaction, $F(2, 57) = 92$, $p < .001$. (Full details of the analyses are available from the authors.)

These data show a very marked benefit for people personally familiar with the targets. The use of the ends of the rating scale was common in the familiar group, and subjects were very accurate indeed in making the seen/unseen decisions. Subjects unfamiliar with the targets performed very poorly, regardless of whether they were students or police officers. Although there were reliable differences on the judgments for targets that had and had not been shown in the videos, these differences were comparatively small for these two groups. These results seem particularly important for the issue of security surveillance. If images of this quality are to be used as legal evidence, it is important to establish which characteristics of the image and of the viewer lead to accurate recognition. From this study, it seems that only personal familiarity will provide a good basis for accuracy of judgments.

What is the basis for the high scores of the familiar subjects? People familiar with the targets may have recognized a number of characteristics. For example, perhaps they recognized the clothes of their lecturers, or their body shape or gait. It seems reasonable to propose that subjects used any cue available in order to make the identifications. The very low resolution of the information carried in the face (very few scan lines on the video) led us to hypothesize that it was not faces that subjects were recognizing in this study, but whole bodies, and we examined this notion further in Experiment 2.

EXPERIMENT 2: THE BASIS OF THE FAMILIARITY ADVANTAGE

In this study, clips from the same video security device were used. However, only subjects familiar with the targets were recruited. In order to examine the basis for the familiarity advantage, we selectively disrupted aspects of the videos by obscuring the head, body, or gait of the targets. As this experiment used only familiar subjects, we used a simple identification task, rather than the recognition memory task used in Experiment 1.

Method

Video clips of 15 target people were selected. Ten of these people were lecturing staff (6 male, 4 female) who would be familiar to all subjects. The remaining 5 people were visitors (3 male, 2 female) who would not be familiar to subjects. In contrast to Experiment 1, video clips were not taken from naturally occurring incidents on the surveillance video, but target people were asked to walk into the building on a prescribed route through the door and toward the camera until they passed out of its range. All clips were gathered on the same day. All clips were edited to last for 3 s. The left panel in Figure 3 shows a still from one of these videos.

Copies of the resulting 15 video clips were edited, using digital video editing equipment, in each of the following ways:

- *Body obscured.* A black rectangle was positioned over the body, scaled to fit the body but not to obscure the background or the head of the person. The rectangle tracked the person through the video sequence, changing shape as necessary (i.e., growing as the subject approached the camera). A still from this condition is shown in the center panel of Figure 3.
- *Face obscured.* A black rectangle was positioned over the head, scaled to obscure the head but not the body of the person. Again, this rectangle tracked the head through the sequence, changing size as necessary. A still from this condition is shown in the right panel of Figure 3.
- *Gait obscured.* To disrupt gait information, we sampled the video frames at seven equal intervals through the 3-s period. Instead of all frames (and hence continual motion) being shown, only seven still frames were shown, each for an equal period. Their duration added up to 3 s. This manipulation destroyed the apparent motion of the video. The viewer saw seven snapshots rather than a moving display, and this made it very difficult to perceive the gait of the target.

The editing procedure resulted in 60 different clips, 15 people × 4 conditions (body obscured, head obscured, gait obscured, unedited). Five different stimulus tapes were prepared in the following way. On each tape, a randomly ordered sequence of the 45 edited clips (i.e., all clips except the original, unedited version) appeared first. The 15 unedited clips then appeared in a randomly ordered sequence. Thus, the edited clips were not presented in blocks, condition by condition, but in mixed order. However, these all preceded the unedited clips.

Twenty-five volunteer subjects were recruited from students studying in the Department of Psychology, University of Glasgow. None had taken part in Experiment 1. Subjects were asked to identify the target in each of the 60 clips in turn. The five different tapes (containing different random orders) were counterbalanced. Subjects were tested individually. They were told that they would see a series of

Fig. 3. Stills from video sequences used in Experiment 2. From left to right, these stills show examples of the unedited, body-obscured, and face-obscured conditions.

videos and that some would contain people familiar to them. After each 3-s clip, the experimenter asked whether they recognized the person in the clip, and if so, to identify the person by name or other distinguishing information. There was no time limit for responses, and subjects were told that they should concentrate on the accuracy of their judgments.

Results

Overall accuracy was high. Across all stimuli, subjects correctly identified 73% of the familiar targets, and correctly rejected 92% of the unfamiliar targets.

Responses to the familiar targets were analyzed in two ways. First, we analyzed the data as though all four conditions were presented in random order, taking subjects' average accuracy score in each of the four conditions (body obscured, face obscured, gait obscured, unedited). However, there are two potential problems with that method. First, the unedited condition was not presented in random order, but always last. Therefore, recognition rates for that condition might be artificially high because of subjects having become familiar with the stimuli through exposure to the edited stimuli. Second, a potentially more serious problem is that recognition in any condition could be affected by prior exposure to a target person in a different condition. So, subjects might recognize a person in the face-obscured condition because they had recently seen that person in the gait-obscured condition. For this reason, in a second analysis, only the data for the first view of each target person were used. In this analysis, each subject contributed only 10 data points, one for each familiar target person. The condition in which this person was first seen provided the only data to enter into the analysis.

Figure 4 summarizes all the data by condition. "Hit" refers to a correct identification of a target familiar to the subject, "miss" refers to a failure to recognize a familiar person as familiar, and "incorrect" refers to an error in which a familiar person was identified as familiar, but mistaken for another familiar person.

Details of all statistical analyses are available from the authors, but can be summarized as follows. Analysis of the hit scores revealed a highly significant effect of condition, $F(3, 72) = 233$, $p < .001$. Tukey HSD (honestly significant difference) tests showed that the unedited condition produced significantly more correct identifications than any other condition; the gait-obscured and body-obscured conditions did not differ, but both produced reliably more correct identifications than the face-obscured condition. Analysis of the miss scores showed the identical pattern of results. Finally, incorrect errors were very infrequent, and were not analyzed further.

Figure 5 shows mean identification scores for the first time each target was encountered. Each of the different orders (videotapes) presented to subjects differed slightly in the number of targets appearing for the first time in each condition, so the data are expressed as proportions. Note that the targets in the unedited condition were always shown last, and so do not appear in Figure 5. Analysis of the hit scores revealed a highly significant effect of condition, $F(2, 48) = 107$, $p < .001$. Tukey HSD tests revealed that the face-obscured condition gave rise to significantly fewer hits than either of the other conditions and that the body-obscured and gait-obscured conditions did not differ significantly. The miss scores showed the same pattern of poorer performance for the face-obscured condition.

Discussion

The data from this experiment strongly suggest that subjects were using information from the face to identify people in these videos. There was a small (but reliable) reduction in accuracy when a person's gait or body was concealed. It is evident from Figure 4 that the face-obscured condition produced much worse performance than all the

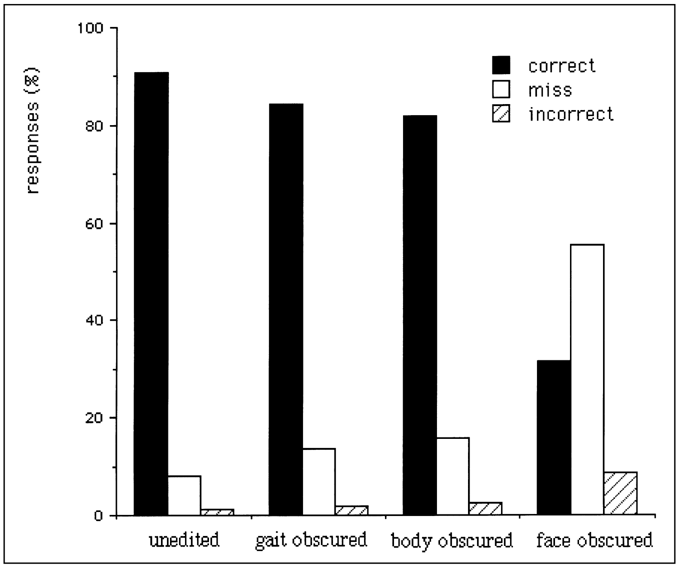

Fig. 4. Recognition accuracy in each video condition in Experiment 2.

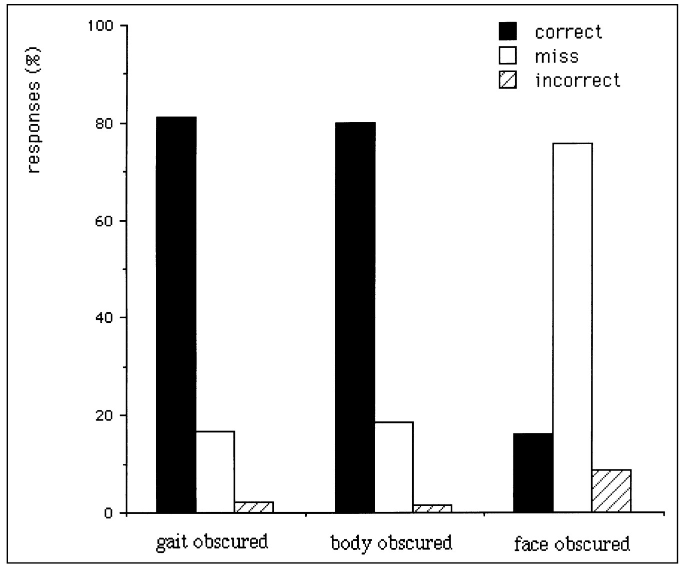

Fig. 5. Recognition accuracy for the first presentation of each target person in Experiment 2.

others. This is even more apparent in Figure 5, which shows that when these images were seen for the first time, people were extremely inaccurate at recognizing them. It was in this condition that subjects had to rely on body shape, gait, and knowledge of the people's clothes. However, it seems that they were unable to make good use of these cues to identify the target people.

GENERAL DISCUSSION

The pattern of data described here can be summarized as follows. When viewing poor-quality videos, people are very good at recognizing familiar targets, and very poor at recognizing unfamiliar targets. The advantage given by familiarity appears to be largely due to recognition of the face itself, rather than recognition of other cues such as gait, body shape, or clothing.

These results have a number of important implications, for both theoretical and applied research in face recognition. Psychologists concerned with familiar-face recognition have routinely sought to discover the building blocks of the recognition process. Faces can be parameterized in a number of different ways. For example, some researchers trying to automate the recognition process have tried to characterize faces by a list of two-dimensional distances in the picture plane, and relations between such measures (e.g., Burton, Bruce, & Dench, 1993; Kanade, 1977; Sakai, Nagao, & Kanade, 1972). More recently, others have used image-based tools relying on patterns of light and dark across the whole image (Burton, Bruce, & Hancock, in press; Kirby & Sirovich, 1990; Turk & Pentland, 1991). It seems from these results that facial identities are available in relatively low resolutions, and this is consistent with previous research on the spatial scale at which information about identity is available (Bachmann, 1991; Harmon & Julesz, 1973). However, videos, which are seen as sequences of frames, provide much more information than any individual frame at this resolution. These issues of resolution are likely to be important to theoretical developments in face recognition.

The implications for forensic practice are also very important. In particular, it seems that identification based on these types of video sequences is very unreliable, unless the viewer happens to know the target person. There have been some other recent findings which suggest that matching unfamiliar faces is difficult, even in the context of high-quality images. For example, Kemp, Towell, and Pike (1997) studied the ability of supermarket cashiers to verify the identity of shoppers from a small (2-cm-square) photograph printed onto a credit card. Kemp et al. found a high error rate in this setting. Cashiers correctly detected fraudulent identity cards on only 36% of trials when foils were chosen to resemble the card bearers. Even when foils bore no particular resemblance to the bearer, only 66% of frauds were detected.

Some recent work in our own laboratory underlines the difficulty of matching unfamiliar faces (Bruce et al., in press). In a series of three experiments, we showed subjects pictures of unfamiliar targets taken on very high-quality video, and asked them to pick out the same targets from an array of high-quality photographs. The videos and photographs of the targets had been taken in good lighting conditions and on the same day, so superficial aspects of the faces (hairstyle, weight, etc.) remained constant. Even in these apparently very favorable conditions, there was a high error rate. When we used stills taken from the videos, errors were highest when there was a pose difference between the target's face on the video and in the photo. However, even in a 10-alternative forced-choice condition, with no time pressure, simultaneous presentation of target and array, and unaltered pose, errors in the order of 25% were observed. Finally, we tested subjects' ability to match a moving, high-quality video clip with an item from a simultaneously presented array of photographs. Once again, errors were unexpectedly high, in the order of 30%. These results, coupled with the results from the present study, suggest that face recognition for unfamiliar people is dominated by pictorial codes, capturing image-specific details. Recognition of familiar people, however, is much more flexible, and appears to be mediated by more abstract representations, capable of generalization over significant changes in image properties.

There are several issues that need to be resolved as a result of this work. First, it will be important to establish exactly the range of video material over which results such as these hold. The particular security system used here was only one example of a commercially available system, and it may be that systems with better image quality support better identification by unfamiliar viewers (though the study by Bruce et al., in press, suggests that improved quality will never eliminate completely the disadvantages observed for unfamiliar viewers). Furthermore, the particular setting of this experiment gave considerable contextual help to viewers familiar with the targets. All subjects familiar with the targets in these experiments knew that the setting was the psychology department in their university, and that the people they were likely to see would be local academics. The help given by context and expectation needs to be quantified. For example, we do not yet know whether subjects would recognize a famous television personality, should one happen to have passed unexpectedly through this video context. Similarly, it is not clear how accurate they would have been in recognizing their lecturers if the lecturers had been presented in an unexpected context, such as a security recording of a crime. These are empirical questions, and it seems that there is a need for full exploration of the various parameters in order to guide good practice in the security industry. Second, these results show rather poor recognition of moving bodies, even by those subjects personally familiar with the target people. Again, this finding needs to be explored further. It seems intuitively reasonable to suppose that observers do use gait and body-shape information to discriminate among people, but this intuition is not supported in the data.

Finally, those people relying on video security surveillance systems need to examine the potential of biometric procedures for identification. In the particular case of poor-quality video and targets unlikely to be familiar to viewers, one needs to establish a procedure for automatically deriving matches between targets and suspects. This will be a particularly difficult job. In the case of familiar-face recognition, there are no existing systems that can outperform human recognition. However, in the case of unfamiliar-face recognition, it is clear that automatic procedures that outperform human abilities by a very large margin are needed. Automatic recognition systems that have been developed and tested against one another do generally show good performance, routinely achieving greater than 90% accuracy in standardized tests (e.g., Phillips, Moon, Rauss, & Rizvi, 1997). However, all these systems perform their analyses on high-quality images. The challenge for the next generation of automatic face recognition devices is to outperform human levels of performance matching unfamiliar faces in low-quality images.

Acknowledgments—This work was funded by a research project from the Economic and Social Research Council (Reference No. R000236688) to Vicki Bruce and Mike Burton.

REFERENCES

Bachmann, T. (1991). Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity. *European Journal of Cognitive Psychology, 3,* 87–103.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology, 62,* 311–318.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73,* 105–116.

Bruce, V. (1988). *Recognising faces.* London: Erlbaum.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P.J.B., Burton, A.M., & Miller, P. (in press). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied.*

Bruce, V., & Humphreys, G.W. (1994). Recognizing objects and faces. *Visual Cognition, 1,* 141–180.

Bruce, V., & Young, A.W. (1986). Understanding face recognition. *British Journal of Psychology, 77,* 305–327.

Burton, A.M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception, 22,* 153–176.

Burton, A.M., Bruce, V., & Hancock, P.J.B. (in press). From pixels to people: A model of familiar face recognition. *Cognitive Science.*

Burton, A.M., Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81,* 361–380.

Burton, A.M., Young, A.W., Bruce, V., Johnston, R.A., & Ellis, A.W. (1991). Understanding covert recognition. *Cognition, 39,* 129–166.

Clifford, B.R., & Bull, R. (1978). *The psychology of person identification.* London: Routledge & Kegan Paul.

Ellis, H.D. (1975). Recognising faces. *British Journal of Psychology, 66,* 409–426.

Harmon, L.D., & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science, 180,* 1194–1197.

Kanade, T. (1977). *Computer recognition of human faces.* Basel, Switzerland: Birkhauser Verlag.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology, 11,* 211–222.

Kirby, M., & Sirovich, L. (1990). Applications of the Karhunen-Loeve procedure for the characterisation of human face. *IEEE: Transactions on Pattern Analysis and Machine Intelligence, 12,* 103–108.

Laughery, K.R., Alexander, J.F., & Lane, A.B. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position and type of photograph. *Journal of Applied Psychology, 55,* 477–483.

Phillips, P.J., Moon, H., Rauss, P., & Rizvi, S. (1997). The FERET evaluation methodology for face recognition algorithms. In *Proceedings of Computer Vision and Pattern Recognition 97* (pp. 137–143). Los Alamitos, CA: IEEE.

Sakai, T., Nagao, M., & Kanade, T. (1972). Computer analysis and classification of photographs of human faces. In *Proceedings of the first USA-Japan Computer Conference* (pp. 55–62). Los Alamitos, CA: IEEE.

Shepherd, J.W., Ellis, H.D., & Davies, G.M. (1982). *Identification evidence: A psychological evaluation.* Aberdeen, Scotland: University of Aberdeen Press.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3,* 71–86.

Yarmey, A.D. (1979). *The psychology of eye-witness testimony.* New York: Free Press.

(Received 7/2/98; Accepted 10/28/98)