Facial Recall and Computer Composites

Introduction

Imagine, if you will, that you are sitting quietly outside a café sipping your favourite hot beverage when someone rushes past and snatches your mobile phone, which you left on the table, as you often do. You were able to get a good look at the person's face, albeit for a short time. Your next hour is spent speaking with a police officer giving a description of what happened and what the offender looked like.

It is likely that you could describe accurately what happened. You will probably also be able to describe the perpetrator's build and clothing. There should be no trouble in saying what was the sex of the person and his or her ethnicity; you should be reasonably accurate at estimating the age, height and weight. You could probably remember some details of the person's face.

Incidents such as these are known as "volume" crime. They occur frequently, often without physical assault to the victim, and their seriousness, at least from a legal perspective, is fairly minor. Due to limitations in police resources, many perpetrators of volume crime are never caught, although time spent locating particularly prolific offenders can be worthwhile.

Crimes involving these repeat offenders, and other crimes of a more serious nature including murder, arson and rape, are generally given higher priority in police investigations. It is in these cases that eyewitnesses (witnesses or victims) may be asked to engage in a range of tasks to assist in the detection and later conviction of the offender. When the police have a *suspect*, they may be asked to take part in an identification parade. (Further details about this are the focus of a separate chapter.) Alternatively, eyewitnesses may be shown photographs of previously arrested criminals for identification, sometimes referred to as *mugshots*. In the absence of a suspect, CCTV footage or other evidence, witnesses may be called upon to externalise an offender's face. The aim is to create a visual image based on remembered information, so that it can be shown to other people for identification. Such images are known as *facial composites* and are seen in the newspapers and on TV, for example, BBC CrimeWatch. The idea is that someone who is familiar with the face will name it to the police and, in doing so, will provide new lines of enquiry.

The focus of the current chapter is on the construction and the recognition of facial composites produced by modern software systems. A separate chapter in this volume details how composites are created by sketch artists. This chapter describes and evaluates typical software programs that the police use to construct faces. It will be demonstrated that the traditional approach used with eyewitnesses is generally ineffective for producing identifiable images, and that alternatives are required if composites are to be effective in the battle against crime. Several successful developments are described. The final section looks to the future and asks what might be on the horizon for producing even more effective faces.

Introduction to face systems and cognitive processes

The traditional procedure for externalising a face from memory involves a police officer asking the eyewitness to first describe what they remember of the offender's face and then, based on that information, selecting individual facial features from a 'kit' of parts. Features include hair, face shape, eyes, nose, brows, mouth and ears. The resulting image is an assemblage or 'composition' of components, hence the term facial *composite*. Note, however, that this term also refers to images produced by sketch artists, and the newer third-generation systems described below. Examples from various systems are presented in Figure 5.1.

Figure 5.1 about here

Irrespective of the technique used to synthesise the face, one of aims of a composite is that someone who is familiar with the face will identify it. Perhaps somewhat unexpectedly, most identifications are made by police officers of repeat offenders. Some, however, emerge as part of a public appeal for information. In either case, with the aim of improving recognition rates, composites are often accompanied by details of both the crime and the offender: age, build, clothing, accent, etc.

Eyewitnesses usually see an offender once, at the time they witness the crime, and so building a composite involves the perception of an *unfamiliar* face. In contrast, identification of the image later by another person involves *familiar* face perception, which is based on an established, long-term visual memory of the face. There is a wealth of evidence to suggest that these two types of perceptual process are very different to each other (e.g. Bruce, 1982; Burton *et al*, 1999; Ellis *et al*, 1979); they are even carried out in different hemispheres of the brain (Schooler, 2002). As a consequence, one would expect that aspects of a face that are important for face construction would be different to those that are important for composite identification. This turns out to be the case.

The police ask witnesses to describe, or recall, an offender's face. Next, they see example features matching their description, decide upon the best matches (a second recall exercise), and resize and reposition each feature on the face as appropriate with the aim of creating the best likeness. Face construction in this way, by the selection of individual features, is traditionally thought to be one of information *recall*. In contrast to recall, as described in earlier chapters, face recognition is assumed to be more holistic in nature, emerging from the parallel processing of individual features and their position, or *configuration*, on the face.

The processes of face recognition and face recall are nicely illustrated in a study by Wells and Hryciw (1984). Participants first made 10 judgements about a target face. One group was asked to base these on physical attributes, such as length of nose or thickness of brows, while the other was asked to rate on perceived personality traits such as honesty and intelligence. After this target *encoding* phase, participants either constructed a composite using the American Identikit system, or attempted to identify the target from among six alternative faces. Those who made physical judgements were found to construct a better quality Identikit than those who made personality judgements, but those asked to judge personality traits were more successful at picking out the target face from among alternatives. The research demonstrates that encoding a face by its physical attributes is beneficial when the subsequent task is of a similar nature: whole face encoding, which arises from personality attribution, is best for recognition.

It turns out that the method we use to remember or encode a face is based somewhat on our expectation of what we will be asked to do subsequently. Olsson and Juslin (1999) found that a holistic encoding was preferred by 64% of participants who were unaware of an ensuing memory task, so-called *unintentional* learning, but feature encoding was preferred (62%) following *intentional* learning (Laughery *et al*, 1986). In conjunction with the above findings of Wells and Hryciw, and others (Frowd *et al*, 2007b), the implication is that composite quality is somewhat dependent on how a face is encoded in the first place. It also divides eyewitnesses into two fairly broad groups. Firstly, there are those who suppose that a description may be required of them, and deliberately use *feature encoding*; they do this by silently making 'mental notes' about the features of the face. Secondly are eyewitnesses who are surprised by the crime, such as the mobile phone theft described above, or domestic distraction burglaries, and encode the face in a more natural, holistic way.

Facial recall

Irrespective of the method used to encode the face, eyewitnesses who construct composites by selecting individual features are required to give a detailed description of the face, and with good reason. Contained within modern computerised systems are hundreds of examples per feature, far too many for an eyewitness to review in their entirety. The software operative therefore uses the given description to limit the number of features shown. In the PRO-fit system, for example, there are 219 noses in the White Male database, but a more manageable set of 24 that are classified with a 'short' length and an 'average' width. Similarly, when working with artists who construct sketched composites, witnesses are directed to example features from catalogues of reference materials that match their verbal description.

There are potential problems, however, with using verbal descriptions as part of constructing faces. Ellis, Shepherd and Davies (1980) found that when matching facial descriptions to target photographs, the longer the time between seeing the perpetrator and giving the description, the more matching errors were made. This indicates that information recall is of less value with time, an established property of human memory (Baddeley, 1990). They also found that matching accuracy was significantly less after one day, which is of forensic importance since eyewitnesses typically recall faces several days after a crime; for some, this can be even longer.

Laughery and colleagues (1986) similarly asked participants to describe faces from memory. Their data suggested that greater attention is generally given to the upper than the lower part of the face and, more generally, that this proceeds top-down from hair to chin. Participants in general had difficulty in describing facial features and also tended to use relative rather than absolute terms, thereby making comprehension difficult; example adjectives include: *small*, *medium*, *large*; *long*, *wide*, *short*; *broad*, *average*, *narrow*; *dark*, *light*. The authors note that the problem with describing faces from memory may simply be because we do not practice doing this in our everyday lives. The exception was hair: presumably we use this feature to describe other people; we also instruct barbers and hairdressers in our desired coiffure.

The above research findings are worrying as the adjectives used for a given face may vary from witness-to-witness, but are a fundamental component of traditional composite construction. Considerable effort, however, has been spent on improving the ability to recall such information from co-operative eyewitnesses and suspects. The work was carried out initially in the 1980s by Ron Geiselman, Ron Fisher and their colleagues (Fisher *et al*, 1989; Geiselman *et al*, 1986), resulting in the *Cognitive Interview* (CI). This is a set of interviewing techniques aimed at eliciting the most complete and accurate recall of information: details of a crime, details of the people involved, descriptions of faces, etc. The CI has been extensively evaluated and revised (See: Milne and Bull, 1999; Wells *et al*, 2007, for reviews: for a meta analysis, Kohnken *et al*, 1999).

One of the principles of the CI is that a complete description of an event, a person's appearance, or a face, is unlikely to be given in a single exhaustive attempt.

This can be improved, however, by making several such recall attempts, each of which will result in a different path being taken through our memories and, in doing so, trigger new information. Human memories are not stored in a logical order, much as we would like, but are based on factors such as their relevance or salience, event position, level of attention (arousal), type of encoding, etc. Memories tend to decay, as mentioned above, making access more difficult, but this can be improved by *context reinstatement*: asking eyewitnesses to visualise smells, sounds, personal feelings and the environment (Davies and Thomson, 1988; Vervaeke *et al*, 2002) – all potential cues that can trigger memories.

Research has also shown that recall is reduced under high levels of physical arousal (state anxiety) (Brigham *et al*, 1983; Valentine and Mesout, 2009). This is relevant to when a crime is being witnessed, but also when the person recalling the event is feeling anxious. To overcome this latter issue to some extent, the CI has an initial 'rapport-building' stage to help eyewitnesses relax.

The CI has been further adapted for investigative interviews as part of an interviewing *framework* that UK police officers follow. This is known as the PEACE CI, a mnemonic that defines each stage of the process: Planning and preparation, Engage and explain, Account, Closure and Evaluation. The interview is also combined with further techniques called *Conversational Management* that have the aim of further enhancing recall. See Dando, Wilcock and Milne (2009) for a recent review.

Computerised feature systems

Face production systems emerged in the 1970s, were non-computerised and composed of facial features printed onto rigid card (Photofit) or transparencies (Identikit). Several serious problems were identified with each, including limitations in the range, sizing and placement of facial features (Davies, 1983; Shepherd and Ellis, 1995). Modern software systems have attempted to overcome these deficiencies.

There are now many software products available to build feature composites. Police forces in the UK rely on two such systems, E-FIT and PRO-fit (Frowd *et al*, 2005b); those in the US have greater choice: FACES, Identikit 2000, ComPhotofit, CRIMES, Compusketch, CD-FIT, E-FIT and FaceKit (McQuiston-Surrett *et al*, 2006). Each system contains a large collection of individual features, normally electronically 'cut' from photographs of faces, and classified in terms of size, shape, colouring, etc. In the UK, only a sample of features is taken from each photograph, to prevent the actors' faces from being reconstructed. PRO-fit databases, for example, sample five features per face.

It is normal for composite systems to have a range of databases that span different ethnic backgrounds, gender and age. Most produce composites in greyscale, as our face perception ability is as accurate in this mode as it is in colour (Davies and Thasen, 2000; Kemp *et al*, 1996). Attempts to design a colour system seem to confirm this notion (Davies, 1986; Frowd *et al*, 2006b). Some systems, such as Identikit 2000 and Compusketch, present features in a sketch-like format, as illustrated in Figure 5.1 above.

The aim with these methods is to produce an identifiable *likeness* rather than a facsimile. This is partly due to the face being constructed from memory, but also because databases do not contain all possible combinations of feature shapes, colourings, etc. To compensate for the latter deficiency, eyewitnesses are given the opportunity for artwork to be applied to the composite face, to add shading, wrinkles, facial marks, etc. While such enhancements require expertise on the part of the police software operative, it can improve the identifiability of a composite (e.g. Gibling and Bennett, 1994). In a small (unpublished) project carried out by the author, for example, people were asked to name composites produced by Frowd and colleagues (2005b), or the same images after removal of artwork, as Figure 5.2 illustrates. People's ability to correctly name the composites halved with the artwork removed.

Figure 5.2 about here

In an attempt to increase the effectiveness of software feature systems, a 'cognitive' approach is the standard method used to build the face. This is based on considerable research suggesting that we are better able to select individual features when they are embedded in a complete face than when we see them as isolated parts (Davies and Christie, 1982; Tanaka and Farah, 1993). For example, a nose is selected with greater accuracy when seen in an intact face than when seen on its own. As we perceive faces as complete entities – that is, holistically – the cues provided by each feature help trigger other memories and improve feature selection. In modern software systems, witnesses now assess individual features as they are switched in and out of an intact face.

Composite system performance

How effective are computerised composites? This question could be answered by auditing their effectiveness in criminal investigations, perhaps via the number of arrests or criminal convictions resulting directly from composite images. While attractive, the approach suffers from the normal problems associated with field experiments: lack of control. Many factors are likely to affect both the quality of a composite and whether it is correctly identified. Example factors include the encoding of the face, the level of anxiety experienced, the number of people who see the composite (circulation) and the presence of additional information on wanted posters: modus operandi, physical descriptions of an offender, clothing, etc.

System evaluations have therefore focussed on the controlled environment of the laboratory to maintain good internal validity (Brace *et al*, 2000; Davies *et al*, 2000; Ellis *et al*, 1975; Frowd *et al*, 2005a; Koehn and Fisher, 1997). This was the approach taken by the author: asking the question, "How well can people construct composites in the laboratory when procedures are used that follow those of 'real' witnesses in the UK?"

Frowd and colleagues (2005b) evaluated the performance of five composite systems. These included E-FIT, PRO-fit and Photofit. There was also a sketch artist, who drew the face by hand using pencils, and a system in development called EvoFIT, which is discussed in detail in a separate section below. In the research, participant-witnesses looked at a good quality colour photograph of an unfamiliar target for one minute. They did so in the knowledge that a composite would be required, to encourage a potentially optimal (feature) encoding strategy. Three to four hours later, they met with an experienced composite operator, in this case one of four experimenters, or a police sketch artist, who worked with them to construct a composite. Each person received a cognitive interview (CI), to prompt recall of the face's appearance, were shown appropriate features that matched this description (except EvoFIT) and attempted to produce the best likeness possible in their own time. Systems were used as stipulated by the manufacturers, and the use of artwork techniques was offered to each person to enhance the likeness of the face.

The targets were 10 celebrity faces, half of which had been previously rated by other people as being distinctive in appearance, and the other half as average. Each person made a single composite of one of these targets with one of the five systems. Fifty images were produced in total. See Figure 5.3 for examples.

Figure 5.3 about here

Another group of participants evaluated the composites, primarily by naming (but a matching task was also used, as discussed later). Images from E-FIT and PRO-fit were named equivalently with a mean at 18.0%, compared with sketches at 9.2%, Photofit at 6.2% and EvoFIT at 1.5%. Thus, naming was fairly good for the two computerised feature systems currently used in the UK, and both were better than the Photofit system that they had replaced; however, naming from sketch and EvoFIT was surprisingly low. The study found a large effect of target distinctiveness: composites of an unusual face were named about three times more successfully overall than those of a more average appearance. This facial distinctiveness effect is observed more generally, when *recognising* unfamiliar faces (Shapiro and Penrod, 1986), and here suggests that offenders will be much more identifiable from a composite if their face has an unusual appearance.

Other research projects have found a similar naming rate for the software feature systems when the delay-to-construction is short (Brace *et al*, 2000; Bruce *et al*, 2002; Davies *et al*, 2000; Frowd *et al*, 2004; 2007a; 2007b; 2008a). A different story emerges, however, when the target delay is much longer. Frowd and colleagues (2005a) followed the same basic design, with the same operators controlling E-FIT, PRO-fit and sketch, but the delay was two days, which is the norm for real witnesses. Examples are presented in Figure 5.1. This time, participant-witnesses made composites that were correctly named at only 3.2% overall, with sketch emerging as the best method, at 8.1%. A sorting task was used as an alternative (proxy) to naming, involving additional participants matching composites to target photographs. There is a tendency when completing the task to compare individual features between composites and targets, and so it provides a broad measure of *feature* quality in the composites. Based on the rate of successful matching, sketches also emerged as the best method of face construction. Taken together, the Frowd *et al* studies suggest that while sketching is unable to produce composites as effectively as the software feature systems when the delay is short, it is more effective when the memory is older and weaker.

The above research employed faces of celebrities as targets, which may not be representative of faces of criminals. Follow-up work using a similar design, including a long delay but non-famous targets, found that while a good quality face was produced occasionally, naming rates remained low overall for the computerized feature systems (Frowd *et al*, 2005c; 2007b; 2010). An evaluation of other methods – FACES 3.0 and Identikit 2000 – has been carried out, with similarly disappointing results (Frowd *et al*, 2007d). Other researchers tell a similar story (Kohen and Fisher, 1997; Kovera *et al*, 1997).

It is perhaps worth mentioning that the participant-witnesses in the above projects were generally given an excellent opportunity to remember the face, engaged in techniques believed to construct the face in the best way, and yet their composites were rarely identified successfully by other people. Such a result is worrying since these techniques are used to detect offenders. What then appears to be the problem?

What could be going wrong?

Frowd and colleagues (2007a) evaluated the quality of the internal and external features of composites constructed in Frowd and colleagues (2005a) – i.e. those produced after a two day delay. Examples are presented in Figure 5.4. Two proxies to naming were used, including the sorting task described above, with participants inspecting composites of internal features, external features or unaltered images. The study revealed similar matching for external and complete composites, but both were superior to internal composites. A follow-up experiment replicated this internal composite *disadvantage*; it also indicated that hair was the most important exterior feature.

Figure 5.4 about here

The study indicates a general ineffectiveness of composite systems in constructing the internal features, but it is this region of the face that is important for later recognition by another person. Also, that there is an emphasis on the exterior part, in particular the hair, during construction. These findings are consistent with the above research, indicating a bias towards verbal recall for the exterior region. Further research by Frowd and Hepton (2009), which focussed on the EvoFIT system described below, has also demonstrated the importance of the internal features in a recognition (rather than a matching) task, a forensically more valid measure. This work has also shown that the amount of *identifying* information in the external features is rather low, as is the case with photographic (non-composite) faces (Ellis *et al*, 1979).

There is a more fundamental problem: the tasks of face description and feature selection are simply contrary to the way faces are seen, as wholes. This observation dates back 30 years, and was noted long before the emergence of the modern systems (Davies *et al*, 1978). It is clear that advances have been made, since a more identifiable face can be constructed using today's technology (Frowd *et al*, 2005b), but performance remains poor when construction take place following long retention intervals (as is the case with witnesses to crime).

There is a hint from sketch production about how to improve the software systems. While sketching is still based on the selection of individual features, the initial focus is on *configural* information – the placement of features on the face – and then on increasing the detail in groups of features. This procedure would appear to encourage more natural, holistic face perception (Davies and Little, 1990; Frowd *et al*, 2005a), an approach similar to modern software systems that require feature selection in the context of an complete face. In fact, a holistic approach to face production has been successfully applied to each stage of the process: to the initial interview, system and finished composite. Some of these are discussed below; each is in UK police use.

Improving computer composites

Morphing

Bruce and colleagues (2002) carried out one of the first projects to successfully improve the performance of a modern face system. The research mirrored the practical situation where an offender had been seen by multiple eyewitnesses, and tried to answer the question as to which of these observers should construct a composite. The problem is that no test currently exists to reliably predict who would produce the best quality image. Their solution was to ask each person to construct a face, and then to average the individual attempts into a single 'morphed' image. Bruce and colleagues argued that as the composites were created from different people's memories, any errors therein would not be correlated and so would cancel in the production of a morph. They demonstrated that the morphed composite was more identifiable than the average individual image, and was sometimes better than the best individual instance.

As a result of their work, the UK police guidelines on facial identification have been modified to permit construction of multiple composites of the same offender, for the purpose of producing a morph for public appeals (ACPO, 2009). Alternative approaches to morphing are required, however, for situations involving a single eyewitness. While it is possible for an observer to create multiple composites of the same offender (see Frowd *et al*, 2006c, for such a case) the norm is to create a single image. Alternative approaches are presented below for this more general case.

Interviewing

As mentioned above, considerable effort has improved the effectiveness of the CI, to recover the most accurate and complete description of a crime, a criminal's face, etc. The interview concerns information recall, which is why it is of value at the start of face construction (Frowd *et al*, 2005c): to enable subsets of features to be located within a composite system. Next, witnesses identify the best matching features. When doing this, they are also engaging in *face recognition* processes. Improving these processes should therefore improve the accuracy of feature selection and the overall identifiability of a composite.

An established procedure to improve unfamiliar face recognition is to attribute holistic judgements at encoding, as described above (Wells and Hryciw, 1984). Another is to make these judgements prior to a recognition attempt (Berman and Cutler, 1989). This latter method was developed into a 'holistic' CI, or H-CI, for face construction (Frowd *et al*, 2005c; 2007b; 2008a). The procedure commences as normal with a CI. Next, witnesses are asked to think about the personality of the face silently for one minute and make seven whole-face judgements (e.g., intelligence, friendliness, honesty, pleasantness, athleticism, trustworthiness and distinctiveness) about the face, rating each on a three-point Likert scale (low/medium/high) before constructing the face as normal.

Frowd and colleagues (2008a) made composites following a CI, which were correctly named at 9%, while those created after an H-CI were very much better named, at 41%. The research indicated that specific holistic judgements do not appear to be important for the interview to be effective, allowing police operatives to select items appropriate for each investigation. It would be inappropriate, for example, to ask victims of sexual assault about the "friendliness" of an offender's face, but this may be acceptable for confidence crimes. A list of suitable adjectives may be found online at HCI (2009).

Caricature

One reliable finding in the literature is that faces of an unusual appearance are recognised more accurately than more average faces (Shapiro and Penrod, 1986). As mentioned above, such target distinctiveness effects extend to face construction for both manual and computerised systems (Frowd *et al*, 2005b). Frowd and colleagues (2007c) argued that composites tend to be quite bland in appearance and so their recognition might be improved by artificially inflating the level of distinctiveness. While artists do this effectively, to produce very recognisable renditions, caricaturing involves considerable skill and is somewhat idiosyncratic (Goldman and Hagen, 1978). Various commercial software programs, however, are now available that produce more consistent results. The PRO-fit system itself, for example, includes such a utility. Each works by comparing features shapes and relationships in a facial image – be it a photograph of a face or a composite – with respect to an average face (Benson and Perrett, 1991) and then exaggerating any differences to produce a positive caricature: they can also de-emphasise differences to produce a representation that is more similar to the reference, a negative caricature.

In a series of experiments, a fixed level of positive caricature, such any of those illustrated in Figure 5.5, only slightly improved participants' ability to identify the composite face (Frowd *et al*, 2007c). In follow-up work, another group of people adjusted the level of caricature, positive to negative, to produce the most identifiable image. There were two surprising results. Firstly was a general preference for a *negative* caricature for the three types of systems tested. This emerged as the composites contained errors, which became reduced, so rendering the image more acceptable, when the face was made to appear more average. The outcome is somewhat similar to a morphed composite: a reduction in error and a face that better resembles the intended target. Secondly, there were large individual differences in preferences: some people preferred a moderate positive caricature, while others a slight negative one.

Figure 5.5 about here

These results led to an important finding: naming rose by about 15% overall when the composite was seen to change in small (5%) steps from -50% to +50% caricature. Also, the sequences were effective for the three types of system tested, but the best improvement emerged for images that were poorly named initially – typical of those produced in criminal investigations. More specifically, 13 of the composites were poorly named, with a mean between 0 and 10%, an overall mean of 3.0%, but this substantially rose to 31.2% when people were presented with the 21 frame sequence. The sequences therefore increased correct identification ten-fold!

The most convenient format for publishing these composites sequences is via an animated GIF image; they can be used on wanted person's web pages, or on TV. An example is available for viewing online at www.EvoFIT.co.uk.

System

The above techniques all suffer from one problem: face construction is problematic if an eyewitness is unable to recall an offender's face in detail. Without good recall, a subset of individual features cannot be identified and so there will be too many items for a witness to inspect. In the UK, for this reason, about 70% of eyewitnesses are denied the opportunity of constructing a composite from a feature system or sketch.

One obvious way forward is to break the dependence on verbal descriptions. However, if a single face is still used, eyewitnesses may still engage in recall to comment upon the accuracy, size and position of individual features. The solution taken by several system designers has been to present multiple faces concurrently and ask witnesses to base selections on the *overall* appearance. The face construction task should then become one of face recognition. Doing so should have the advantage of more stable performance over time, compared to systems based on face recall (Davies, 1983; Ellis *et al*, 1980; Shepherd, 1983), offering the possibility of accurate face construction even from weak memories.

Several third-generation or 'recognition' systems have now emerged that are premised in this way. In the UK, there are EvoFIT (Frowd *et al*, 2004) and E-FIT-V (Gibson *et al*, 2003), and in South Africa, ID (Tredoux *et al*, 2006); there is also a German system in development (Blanz *et al*, 2006). The systems have a software module that is able to generate a large number of faces, each with a specified set of face coefficients. The basic approach is to present witnesses with screens of randomly-generated faces to select. Selected items are 'bred' together, by combining the underlying coefficients, to produce more items for selection. Repeating this procedure a few more times allows the face set to become more similar to each other and more similar to offender's face in the memory of the eyewitness. The best likeness is ultimately taken as the 'composite'. The underlying mechanism for generating faces using PCA models, and the typical method for combining items using an evolutionary algorithm (EA), is the subject of a separate chapter by Chris Solomon.

All composite systems are essentially engaged in a search problem: to locate a specific identity. For the recognition types, the search is within a high dimensional face 'space' and each presented face represents a potential 'solution'. In principle, the more faces on which witnesses give feedback, the more thorough the search will be and the greater the chance will be of locating an identifiable face. (See Frowd *et al*, 2004, for an evaluation of 'population' size to support this idea.) But, system designers need to be careful to allow convergence on an appropriate likeness before witnesses suffer fatigue.

In the following section, EFIT-V is outlined, followed by EvoFIT and the developments that have been necessary to render it effective. As with the feature types, the software is controlled by experienced personnel.

System: EFIT-V

In the E-FIT-V system, a database of race and gender is first selected, to reflect the background of the offender as remembered by the eyewitness. The system is flexible in use, but witnesses generally start by selecting an appropriate facial shape – round, square, oval and so forth – and a hairstyle, before being presented with screens of nine faces that change in appearance by both 'shape' (feature shapes and placement on the face) and 'texture' (pixel colourings of the eyes, brows, mouth, etc.) Witnesses select examples that resemble an offender, or reject others, and an EA breeds the relevant items together. The process is repetitive, and there are software tools for manipulating individual features and whole-face attributes such as age and masculinity. A paint program is available to add lines, wrinkles, shading, etc. See Chris Solomon's chapter for a more detailed description of EFIT-V.

System: EvoFIT

To construct an EvoFIT, an appropriate database for age, gender and race is first selected. Witnesses then choose an appropriate set of external features, in particular hair, and are presented with arrays of complete faces, 18 per page. During the early stages of development, users preferred to view facial shape and facial texture separately (Frowd *et al*, 2004) and so the interface was developed to present information in this way. In practice, they are shown four screens of shape and select the two best likenesses per page up to a maximum of six, then similarly for texture. Also, to assist with selecting the latter, the textures themselves are presented on a specified face shape, one that the user believes to be the best shape at this stage. To assist in the conversion of a good likeness, witnesses next choose from the best combination of selected shape and texture, and identify a so-called 'best' face, which is subsequently given twice the number of breeding opportunities in the EA. EvoFIT tended to converge on a good facial type after three breeding cycles – the initial generation plus two breeding cycles.

Frowd *et al* (2007b) evaluated the effectiveness of this version of EvoFIT using the 'gold standard' described above: unfamiliar targets to the people constructing the composites, a two day delay, use of a CI, etc. Under target feature encoding, composites were named better from EvoFIT (11%) than from a typical modern 'feature' system (4%). While this represents an improvement in identifiability relative to the traditional, overall performance was not impressive.

Two further developments have been effective. The aim of the first was to address the issue raised previously: the external features have high importance for unfamiliar face perception – composite construction here – but it is the internal features that are important for recognition – composite naming. In EvoFIT, a Gaussian or 'blur' filter is applied to the external features after they had been chosen. This image filtering allows selections to be based on the central part of the face, as Figure 5.6 illustrates. After evolving, blurring is disabled. Two projects have shown that the selective blurring of faces in this way promotes a more identifiable composite when the target delay is short (Frowd *et al*, 2008b) and long (Frowd *et al*, 2010).

Figure 5.6 about here

The second development allowed better convergence of age and other holistic properties of a face. This was achieved in part by limiting the age capability of the face generators, specifically by building (PCA) models from faces of a specified age range. There are now separate models built from people in their twenties, thirties, etc. However, an incorrect aged face was still produced sometimes (although the age match was now closer than before). The basic problem here is the complexity of the task: witnesses may select faces that are accurate in one aspect (e.g. honest-looking) but not another (attractiveness), and these choices will be reflected with evolution. This issue was addressed by allowing users to change their evolved face along a number of psychologically-sensible 'holistic' dimensions. These include age, face weight, masculinity, attractiveness, threatening and honesty. The method used to create these scales is described in Frowd, Bruce, McIntyre and colleagues (2006a); an example manipulation is shown in Figure 5.7.

Figure 5.7 about here

Example composites evolved with EvoFIT and these two developments are presented in Figure 5.8. Using the gold standard construction procedure, which is described above, including a two day delay, the system's effectiveness was recently assessed (Frowd *et al*, 2010). The enhanced software was used, except that witnesses also selected a facial *aspect ratio*, an appropriate face width and height, after choosing the hair. Selecting a facial aspect ratio allows items to be shown with the ratios constrained, with the aim of helping face selection and the production of an identifiable likeness. Second, due to improvements brought about by blurring, holistic tools and facial aspect ratios, two rather than three breeding cycles were used. This has the obvious advantage of presenting witnesses with fewer faces to choose from. Results of the evaluation were that blurring and holistic scales were effective on their own, but best performance emerged when used in conjunction with each other: mean naming was 25% correct from EvoFIT and 5% from a 'feature' system.

Figure 5.8 about here

EvoFIT has turned out to be something of an enigma, and attempts to improve its effectiveness have not always been successful (Frowd *et al*, 2007b; 2008b). While the third-generation systems are supposed to be based on face recognition rather than recall, this idea is too simplistic (at least there is evidence of this from the development of EvoFIT). For example, one might expect that the holistic component of the holistic CI (H-CI) would be effective on its own, to improve a user's face recognition ability and thus his or her accuracy in selecting whole faces from the presented arrays. In recent work (Frowd *et al*, submitted), this holistic-attribution component actually promoted worse quality composites than the face recall part of the CI; curiously, correct naming of EvoFITs from H-CI (40%) was superior to those from CI (25%). Face recall therefore allows selections to be made with more accurate features, which is beneficial to the evolution process in the long run. But, detailed descriptions produce over-emphasis on individual features, and so holistic-attribution after the CI provides a shift towards whole-face selection. That said, there is a twist-in-the-tale, as a previous EvoFIT procedure with feature manipulations early-on gave inferior results for the H-CI. So, the type of interview administered and the procedure used to evolve the face have a somewhat complex and unexpected relationship.

The future

One might ask whether facial composites really have a role to play in identifying offenders. About five years ago, indications were that modern computer feature systems consistently failed to produce good quality images when the target delay was fairly long (Frowd *et al*, 2005a). This is particularly worrying as many police forces rely on this type of technology to detect offenders. It is perhaps interesting to note the historical pattern that has emerged with composite systems: a system is produced, adopted by police forces and then found to be ineffective. This was the case with the non-computerised systems, Photofit and Identikit, and likewise with their modern computerised decedents. When developing EvoFIT, we were keen to avoid making the same mistake; about 10 years of intensive development were required to reach satisfactory performance (Frowd *et al*, 2010).

At least one of the above developments has general benefits: caricature animation is not only effective for feature systems, but also for sketches, which the evidence would suggest are hard to recognise even when the target delay is short (Frowd *et al*, 2005b); for EvoFIT, indications are that the animation boosts mean naming levels by a further 15%. In addition, while the H-CI is beneficial to the latest EvoFIT procedure, and perhaps to all recognition-based systems, it will be interesting to see whether the new interview is of value to sketch artists; indications are that it should be, given that sketching involves the selection of individual features, which should similarly be improved following improvements to an observer's face recognition ability. It is also possible that the H-CI may be even more effective here due to the inherently holistic nature of sketch production (Davies and Little, 1990).

Sketches tend to be qualitatively different to the other systems on one respect: they contain less shading information – see Figures 5.1 and 5.3. This would appear to result from our inability to recall the texture of the face in sufficient detail for accurate rendering on the page, and therefore major regions are often left blank, or with minor shading. One curious possibility is that a sketched face may, in some circumstances, be more accurate overall by virtue of there being less incorrect information!

In a small project, the effect of reducing visual detail was explored (Frowd *et al*, 2008b). This was done by simply increasing the brightness level in a set of composites. Results found significantly better naming for such enhanced images relative to veridical. What this demonstrates is that information reduction can be useful, as some of the inaccurate information will be removed, to allow a perceiver's cognitive system to 'fill in the gaps'. Examples are presented in Figure 5.9. While a less detailed representation does not help face *construction* by individual features (Frowd *et al*, 2005a; 2007d), it does for EvoFIT (Frowd *et al*, 2008b). Indeed, a database of this type may be useful in situations where the race of the face is unfamiliar to the eyewitness and recognition abilities are challenged even further (Meissner and Brigham, 2001). It may also be valuable for observers who have viewed an offender with *unintentional* encoding, perhaps due to the sudden nature of the crime, and so have fairly limited detail of the face on which to draw.

Figure 5.9 about here

Software designers are still producing a 'one-size fits all' solution: a single system for all eyewitnesses. This simply may not be optimal. While it is normal for witnesses to produce a poor likeness using feature selection following long delays, they sometimes produce a very good one – Figure 5.9's PRO-fit of Nicholas Cage is a case in point; the same idea applies to sketched composites (Frowd *et al*, 2005a). Perhaps one of the challenges facing psychologists today is to understand individual differences between observers for the purpose of matching them to the face construction technique. This turns out to be a particularly hard task, one that I myself have attempted. Given the role that face recall and recognition play in face construction, which also appears to change according to the system used, perhaps a combined recall-recognition diagnostic test might be a fruitful avenue of research.

Concluding comments

About five years ago, face construction by the selection of individual features was shown to be a generally ineffective method, especially when involving long delays. Considerable research effort, spanning some three decades, has led to important developments to the system, interview and image presentation format. It is now possible that a face built under forensically-relevant conditions is well-recognised and thus of value to law enforcement, whether this is from a modern feature system or from the newer EvoFIT (and perhaps even one of the other recognition types). The future for facial compositing is promising: there is clearly headroom for improvement and some interesting new avenues for research and development.

Answers

The identities shown in Figure 5.3 are of Brad Pitt (actor), Robbie Williams (popstar), David Beckham (footballer), Noel Gallagher (popstar) and Michael Owen (footballer). In Figure 5.8, they are of Simon Cowell (pop manager), George W. Bush (former US president), David Tennant (actor) and Noel Gallagher (popstar).

(6898 words not including Chapter Introduction)

References

ACPO(2009). Facial Identification Guidance. Produced by the National Policing Improvement Agency. Unpublished document.

Baddeley, A. (1990). *Human memory: theory and practice*. LEA.

Benson, P.J., and Perrett, D.I. (1991). Perception and recognition of photographic quality caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, **3**, 105-135.

Berman, G.L., and Cutler, B.L. (1998). The influence of processing instructions at encoding and retrieval on face recognition accuracy. *Psychology, Crime and Law*, **4**, 89-106.

Blanz, V., Albrecht, I. Haber, J., and Seidel, H.P. (2006). Creating Face Models from Vague Mental Images. *Eurographics,* **25**. Electronic document.

Brace, N., Pike, G.E., Allen, P., and Kemp, R. (2006). Identifying composites of famous faces: investigating memory, language and system issues. *Psychology, Crime and Law*, **12**, 351-366.

Brigham, J.C., Maass, A., Martinez, D., and Whittenberge, G. (1983). The Effect of Arousal on Facial Recognition. *Basic and Applied Social Psychology*, **4**, 279-293.

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

Bruce, V., Ness, H., Hancock, P.J.B, Newman, C., and Rarity, J. (2002). Four heads are better than one. Combining face composites yields improvements in face likeness. *Journal of Applied Psychology*, **87**, 894-902.

Burton, A.M., Wilson, S., Cowan, M., and Bruce, V. (1999). Face recognition in poor quality video: evidence from security surveillance. *Psychological Science*, **10**, 243-248.

Dando, C. J., Wilcock, R. and Milne, R. (2009). The Cognitive Interview: The efficacy of a modified mental reinstatement of context procedure for frontline police investigators. *Applied Cognitive Psychology*, **15**, 679 – 696.

Davies, G.M. (1983). Forensic face recall: the role of visual and verbal information. In S.M.A. Lloyd-Bostock and B.R. Clifford (Eds.). *Evaluating witness evidence*. Chichester: Wiley, Chapter 6.

Davies, G.M. (1986). Recall and reconstruction of faces; Implications for theory and practice. In Ellis, H. D., Jeaves, M., Newcome, F. and Young, A. (Eds.) Aspects of Face Processing, Dodrecht, The Netherlands, Martinus Nijhoff.

Davies, G.M., and Christie, D. (1982). Face recall: an examination of some factors limiting composite production accuracy. *Journal of Applied Psychology*, **67**, 103-109.

Davies, G.M., and Little, M. (1990). Drawing on memory: Exploring the expertise of a police artist. *Medical Science and the Law*, **30**, 345-354.

Davies, G.M., Shepherd, J., and Ellis, H. (1978). Remembering faces: acknowledging our limitations. *Journal of Forensic Science*, **18**, 19-24.

Davies, G.M., and Thasen, S. (2000). Closed-circuit television: How effective an identification aid? *British Journal of Psychology*, **91**, 411-426.

Davies, G.M., and Thomson, D.M. (1988). *Memory in context: context in memory*. London: Wiley.

Davies, G.M., van der Willik, P., and Morrison, L.J. (2000). Facial Composite Production: A Comparison of Mechanical and Computer-Driven Systems. *Journal of Applied Psychology*, **85**, 119-124.

Ellis, H.D. Shepherd, J.W., and Davies, G.M. (1975). Use of photo-fit for recalling faces, *British Journal of Psychology*, **66**, 29-37.

Ellis, H.D., Shepherd, J.W., and Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, **8**, 431-439.

Ellis, H.D., Shepherd, J.W., and Davies, G.M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration*, **8**, 101-106.

Fisher, R. P., Geiselman, R. E., and Amador. M. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witnesses of crime. *Journal of Applied Psychology*, **74**, 722-727.

Frowd, C.D., Bruce, V., McIntyre, A., and Hancock, P.J.B. (2007a). The relative importance of external and internal features of facial composites. *British Journal of Psychology*, **98**, 61-77.

Frowd, C.D., Bruce, V., McIntyre, A., Ross, D., Fields, S., Plenderleith, Y., and Hancock, P.J.B. (2006a). Implementing holistic dimensions for a facial composite system. *Journal of Multimedia*, **1**, 42-51.

Frowd, C.D., Bruce, V., Ness, H., Bowie, L., Thomson-Bogner, C., Paterson, J., McIntyre, A., and Hancock, P.J.B. (2007b). Parallel approaches to composite production. *Ergonomics*, **50**, 562-585.

Frowd, C.D., Bruce, V., Plenderleith, Y., and Hancock, P.J.B. (2006b). Improving target identification using pairs of composite faces constructed by the same person. *IEE Conference on Crime and Security* (pp. 386-395). London:IET.

Frowd, C.D., Bruce, V., Ross, D., McIntyre, A. and Hancock, P.J.B. (2007c). An application of caricature: how to improve the recognition of facial composites. *Visual Cognition*, **15**, 954-984.

Frowd, C.D., Bruce, V., Smith, A., and Hancock, P.J.B. (2008a). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied*, **14**, 276 – 287.

Frowd, C.D., Bruce, V., Storås, K., Spick, P. and Hancock, P.J.B. (2006c). *An evaluation of morphed composites constructed in a criminal investigation*. Proceedings of the 16th Conference of the European Association of Psychology and Law, London: IP-PA Publishing, pp. 59-66.

Frowd, C.D., Carson, D., Ness, H., McQuiston, D., Richardson, J., Baldwin, H., and Hancock, P.J.B. (2005a). Contemporary Composite Techniques: the impact of a forensically-relevant target delay. *Legal and Criminological Psychology*, **10**, 63-81.

Frowd, C.D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S., and Hancock, P.J.B. (2005b). A forensically valid comparison of facial composite systems. *Psychology, Crime and Law*, **11**, 33-52.

Frowd, C.D., Hancock, P.J.B., and Carson, D. (2004). EvoFIT: A holistic, evolutionary facial imaging technique for creating composites. *ACM Transactions on Applied Psychology (TAP)*, **1**, 1-21.

Frowd, C.D., and Hepton, G. (2009). The benefit of hair for the construction of facial composite images. *British Journal of Forensic Practice*, **11**, 15-25.

Frowd, C.D., McQuiston-Surrett, D., Anandaciva, S., Ireland, C.E., and Hancock, P.J.B. (2007d). An evaluation of US systems for facial composite production. *Ergonomics*, **50**, 1987–1998.

Frowd, C.D., McQuiston-Surrett, D., Kirkland, I., and Hancock, P.J.B. (2005c). The process of facial composite production. In A. Czerederecka, T. Jaskiewicz-Obydzinska, R. Roesch and J. Wojcikiewicz (Eds.). *Forensic Psychology and Law* (pp. 140-152). Krakow: Institute of Forensic Research Publishers.

Frowd, C.D., Park., J., McIntyre, A., Bruce, V., Pitchford, M., Fields, S., Kenirons, M. and Hancock, P.J.B. (2008b). Effecting an improvement to the fitness function. How to evolve a more identifiable face. In A. Stoica, T. Arslan, D. Howard, T. Higuchi, and A. El-Rayis (Eds.) *2008 ECSIS Symposium on Bio-inspired, Learning, and Intelligent Systems for Security* (pp. 3-10). NJ: CPS. (Edinburgh).

Frowd, C.D., Pitchford, M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., McIntyre, A.H., and Hancock, P.J.B. (2010). The psychology of face construction: giving evolution a helping hand. *Applied Cognitive Psychology*. DOI: 10.1002/acp.1662

Frowd, C.D., Nelson, L., Skelton, F., Noyce, R., et al. Interviewing techniques for Darwinian facial composite systems. Submitted to *Journal of Experimental Psychology: Applied*.

Geiselman, R.E., Fisher, R.P., MacKinnon, D.P., and Holland, H.L. (1986). Eyewitness memory enhancement with the cognitive interview. *American Journal of Psychology*, **99**, 385-401.

Gibling, F., and Bennett, P. (1994). Artistic enhancement in the production of photofit likeness: an examination of its effectiveness in leading to suspect identification. *Psychology, Crime and Law*, **1**, 93-100.

Gibson., S.J., Solomon, C.J., and Pallares-Bejarano, A. (2003). Synthesis of photographic quality facial composites using evolutionary algorithms. In R. Harvey and J.A. Bangham (Eds.) *Proceedings of the British Machine Vision Conference* (pp. 221-230).

Goldman, M., and Hagen, M. A. (1978). The forms of caricature: Physiognomy and political bias. *Studies in the Anthropology of Visual Communication*, **5**, 30-36.

HCI (2009). Holistic-Cognitive Interview. www.uclan.ac.uk/scitech/psychology/research/evofit/holistic_cognitive_interview.php

Kemp, R., Pike, G., White, P., and Musselman, A. (1996). Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues. *Perception*, **25**, 37-52.

Koehn, C.E., and Fisher R.P. (1997). Constructing facial composites with the Mac-a-Mug Pro system. *Psychology, Crime and Law*, **3**, 215-224.

Köhnken, G., Milne, R., Memon, A., and Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime and Law*, **5**, 3-27.

Kovera, M.B., Penrod, S.D., Pappas, C., and Thill, D.L. (1997). Identification of computer generated facial composites. *Journal of Applied Psychology*, **82**, 235-246.

Laughery, K.R., Duval, C., and Wogalter, M.S. (1986). Dynamics of facial recall, in Ellis, H.D., Jeeves, M.A., Newcombe, F., and Young, A. (Eds.). *Aspects of face processing* (pp. 373-387). Dordrecht: Martinus Nijhoff.

McQuiston-Surrett, D., Topp, L. D., and Malpass, R. S. (2006). Use of facial composite systems in U.S. law enforcement agencies. *Psychology, Crime and Law*, **12**, 505–517.

Meissner, C.A., and Brigham, J.C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, **7**, 3–35.

Milne, R. and Bull, R. (1999). Investigative interviewing: Psychology and practice. Chichester: John Wiley and Sons, Ltd.

Olsson, N. and Juslin, P. (1999). Can self-reported encoding strategy and recognition skill be diagnostic of performance in eyewitness identification? *Journal of Applied Psychology*, **84**, 42-49.

Schooler, J.W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology*, **16**, 989-997.

Shapiro, P. N., and Penrod, S.D. (1986). Meta-analysis of facial identification rates. *Psychological Bulletin*, **100**, 139-156.

Shepherd, J.W. (1983). Identification after long delays. In S.M.A. Lloyd-Bostock, and B.R. Clifford (eds.), *Evaluating witness evidence* (pp. 173–187). Chichester: Wiley.

Shepherd, J.W., and Ellis, H.D. (1995). Face recall - methods and problems. In S.L. Sporer, R.S. Malpass and G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 87-115). Hillsdale, NJ: Lawrence Erlbaum.

Tanaka, J.W., and Farah, M.J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **46A**, 225-245.

Tredoux, C.G., Nunez, D.T., Oxtoby, O., and Prag, B. (2006). An evaluation of ID: an eigenface based construction system. *South African Computer Journal*, **37**, 1-9.

Valentine, T., and Mesout, J. (2009). Eyewitness identification under stress in the London Dungeon. *Applied Cognitive Psychology*, 23, 151-161.

Vervaeke, G., Vanderhallen, M., Van Koppen, P., and Goethals, J. (2002). *Much Ado About Crime: Chapters on psychology and law*. Brussel, Belgium: Uitgeverij Politeia NV.

Wells, G.L., and Hryciw, B. (1984). Memory for faces: encoding and retrieval operations. *Memory and Cognition*, **12**, 338-344.

Wells, G., Memon, A., and Penrod, S.D. (2007). Eyewitness evidence: improving its probative value. *Psychological sciences in the public interest*, **7**, 45-75.

Acknowledgement
The author would like to thank Dr Chris Atherton at the University of Central Lancashire, Preston, for her insightful comments on a draft version of Chapter 5, and Emma Walker for proof reading a near-final version.

List of figures
(All of these figures are available as individual image files.)

Figure 5.1. Examples composites from (left to right) E-FIT, PRO-fit, Sketch, EvoFIT, FACES 3.0 and Identikit 2000. These were constructed from different people's memories of the actor, Ben Affleck (Frowd *et al*, 2005a, 2007d).

Figure 5.2. Artistic enhancement: the image on the left was constructed using E-FIT of the pop singer, Noel Gallagher; the final image, after artwork was applied, is shown on the right.

Figure 5.3. Composites constructed by Frowd *et al* (2005b) using (left-to-right): E-FIT, PRO-fit, Sketch, Photofit and EvoFIT. Can you name the celebrities? They are listed at the end of the chapter.

Figure 5.4. Internal features, external features and complete composite of the actor, Ben Affleck used in Frowd *et al* (2007a).

Figure 5.5. A composite of the former UK Prime Minister Tony Blair that has been progressively caricatured as part of Frowd and colleagues (2007c): -50%, -25%, 0% (veridical), +25% and +50% caricature. Naming of the target identity was found to substantially improve when participants saw a sequence of 21 such images.

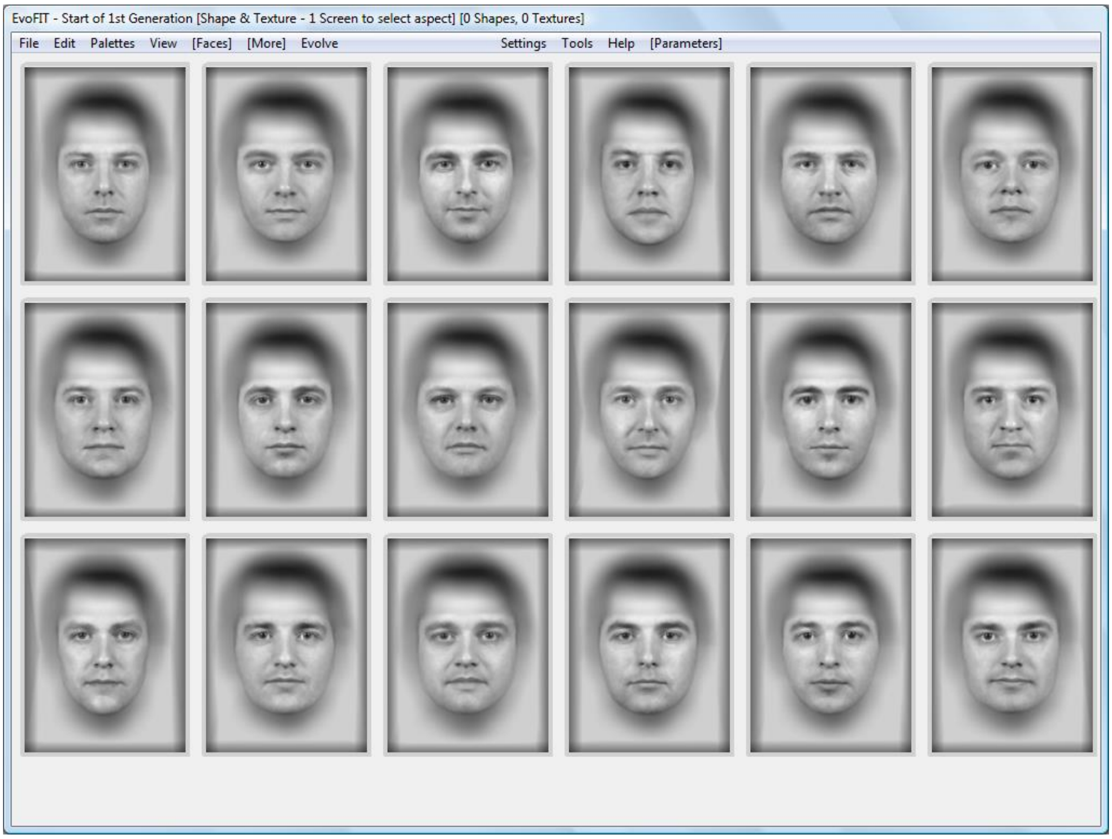

Figure 5.6. An example EvoFIT screen. The external features of the faces are 'blurred', to allow eyewitnesses to focus on the central part of the face. At the end of evolving, the blur is removed to allow the final image to be clearly seen.

Figure 5.7. The face on the left was evolved of the footballer David Beckham from memory in Frowd and colleagues (2006a); on the right, after holistic tool use. In this case, the perceived level of health and attractiveness were increased.

Figure 5.8. Example EvoFITs constructed using the system plus recent developments. The celebrity identities shown are listed at the end of the chapter.

Figure 5.9. Reducing the detail of a composite by increasing brightness levels. The left pair are of Tony Blair, while the right pair are of the actor Nicholas Cage.