Catching Even More Offenders with EvoFIT Facial Composites

Charlie D. Frowd, Melanie Pitchford, Faye Skelton, Anna Petkovic
University of Central Lancashire
cfrowd@uclan.ac.uk, mpitchford2@uclan.ac.uk, fcskelton@uclan.ac.uk, petkovic.anna@gmail.com

Colin Prosser and Brian Coates
Humberside police
Colin.Prosser@humbeside.pnn.police.uk and Brian.Coates@humberside.pnn.police.uk

Abstract
Facial composites are an investigative tool used by police to identify suspects of crime. Unfortunately, traditional methods to construct the face have rather low success rates. We have been developing a new recognition-based method called EvoFIT that requires eyewitnesses to select whole faces from arrays of alternatives. Both published laboratory research and existing police field-trials have found that EvoFIT produces images that are more identifiable than images from traditional systems. In the current paper, we present an evaluation of a more recent version of EvoFIT: in 2010, EvoFIT was deployed in 35 criminal investigations by Humberside police and these images directly led to identification of 21 suspects, equating to 60% success—quadruple the performance of the previous system used within the force. The evaluation also showed that identification of a suspect led to conviction in 29% of investigations (6 out of 21). Overall, a conviction occurred in 17% of cases involving use of an EvoFIT (6 out of 35). We also outline more recent developments which indicate that an arrest is now likely in three out of every four cases in which EvoFIT is used, and a conviction rate of one in five.

1. Introduction

Sometimes, physical evidence is available in police investigations to allow suspects to be identified: CCTV footage, fingerprints, DNA, footprints, etc. At other times, however, the evidence is less tangible, being contained in the memory of an eyewitness. In this latter situation, recovering an accurate account of what happened and the people involved are important for solving the case. For serious crime—rape, burglary and murder—investigating officers normally consider construction of a facial composite in order to identify a suspect. Here, witnesses and victims who have clearly seen the offender’s face describe the appearance of the face and work with a trained composite officer to construct a facial likeness. The resulting image is circulated within a police force or in the media for identification—to generate further lines of enquiry and, ideally, to locate the offender.

Many techniques have been devised to construct faces from memory. Initially, these were manual-type systems such as sketch, Identikit (in the US) and Photofit (UK). Later, in the 1990s, software programs were created including CD-FIT, E-FIT and PRO-fit; more recently there are US systems such as FACES and Identikit 2000. All of these methods require witnesses to select individual facial features (eyes, hair, nose, mouth, etc.) to create a facial composition or ‘composite’ [1].

Considerable research has established that these ‘feature’ methods are ineffective when used to construct a face from memory two or more days after a target face has been seen [1]. For example, in one of our projects [2] using police-like procedures and a 2-day delay, E-FIT and PRO-fit produced composites that were named by other people with an average (using the mean) of less than 5% correct. This finding has been replicated (e.g. [3][20]) and found to apply to other feature systems [5][6].

It is for this reason that alternative methods have been developed to construct facial likenesses [7][8][9]. These are based on selecting complete faces, a task that should be better aligned to natural (holistic) face recognition [12], and potentially easier for a witness to do than selecting individual features. We have been developing EvoFIT ([1][3][6][9][10][11][13][14][18][19]), in which, witnesses are shown screens of complete faces and select those that resemble the offender. Selected faces are combined together (similar to genetic ‘breeding’) to produce more faces for selection and, when repeated a few times, a composite is ‘evolved’. While the approach has intuitive appeal, about 10 years of intensive research were required to develop a system that could produce more identifiable composites than other systems [1].

The technical details of EvoFIT have been described in many research papers and system reviews. See Frowd et al. [9] for a thorough technical review of its development, computer simulations and initial testing.
with users, and [18] and [19] for more recent laboratory evaluations. We provide a brief technical account here. At the heart of EvoFIT is a face generator built using Principal Components Analysis (PCA), to capture variation in shape and texture for a database of about 70 facial images (of a given race, gender and age range). Initial faces presented to witnesses are generated from this computer model with random PCA parameters, in sets of 18. Figure 1 is an example set produced from the white male database covering ages of 30 to 40 years. Witnesses select from screens such as these, and the parameters underlying these selections are combined together to produce more faces for witnesses to select. Technically speaking, this ‘breeding’ process involves uniform crossover and selection with parameters occasionally mutated (to increase variability within the population of faces, improving performance). Witnesses also select a single face that represents the best overall likeness and this preferential item is given greater weightage: twice the number of breeding opportunities in the Genetic Algorithm; it is also transferred unaltered to the next generation as part of an ‘elitist’ approach. The process of selection and breeding is repeated, to ‘evolve’ a face.

In 2007, volunteers produced composites using EvoFIT with a mean naming of 25% correct [3]. This software contained techniques to help witnesses focus on the central part of the face—the region that is important for identification by another person [4]. It also contained ‘holistic’ tools to manipulate the evolved face: to adjust perceived age, masculinity, health and other overall properties of the face; further scales are available to add various types of facial hair, and lighten and/or darken eyes, brows, etc. Using the same construction procedures, including a 48-hour delay to interview, composites from the PRO-fit feature system were named at 5% [3]. At this time, then, identification was about five times higher from EvoFIT than from a leading feature system. Even better performance is now possible (see below and section 3.1).

From 2007 to 2009, field trials were carried out in Lancashire and Derbyshire police forces [13]. In spite of the difficulties involved in running field studies, discussed in detail in [13], methods were established for assessing performance. Police officers were trained within these forces and used EvoFIT for a nominal period of six months. EvoFIT composites directly lead to identification and arrest of a suspect in 20% of cases—see Figure 2 for an example case. In contrast, the previous E-FIT system did not provide any identifications1.

Figure 2. An EvoFIT (left) constructed by a victim during the Lancashire police field trials. The image was named as Ross Gleave (right) and led to his arrest and conviction of attempted rape of a child under 13 years.

From summer 2009 to spring 2010, further field trials were run by Devon and Cornwall police, and police in Romania. These evaluations were similar in nature to the above but used a simplified procedure in the face-recall interview with witnesses. Previously, witnesses described the appearance of the offender’s face, known as free recall, and were then prompted to remember more information about it using a standard question-and-answer format, cued recall. However, based on research suggesting potential problems with this latter technique

1For the 2007-10 field trials [13][14], EvoFIT was used 126 times and identified 32 suspects; in a comparable audit of E-FIT, no identifications resulted from 20 composites. This increase in identification from E-FIT to EvoFIT is statistically reliable (at the standard probability level of 0.05) using a Fisher Exact Probability Test (p = .027). (N.B. A Chi-Square test is not appropriate here as one of the expected frequencies is less than five.)
[20], these later field trials did not involve cued recall. Identification was found to be markedly better, with a suspect identified in 38% of cases in Romania and 40% in Devon and Cornwall. See Figure 3 for an example case.

Figure 3. An EvoFIT constructed of a violent robber (who also used a weapon) and a photograph (right) of the person convicted of this and a second similar offence.

The field trials also made use of a new technique for police to make public appeals on TV, wanted-persons web-pages and online newspapers. The technique presents a composite as a series of progressively more-exaggerated caricatures, similar to the way that an artist caricatures a face. When observing such a sequence, identification increases substantially [15]. It is a standard animation format and is produced by a composite operator as a secondary exhibit for enhancement (and a statement is produced to this effect). An example can be seen online at http://www.uclan.ac.uk/animate compostes.

Research continued to seek ways to improve EvoFIT, and in 2010, we developed an even more effective method [19]. This involved witnesses first constructing the internal facial features, the region including the eyes, brows, nose and mouth, and which are important for successful recognition of the face [4]; having created this region, witnesses then select hair, ears and neck. Laboratory tests indicated that EvoFITs produced in this way were twice as identifiable, at 45% correct, than the previous method. An updated version of the system was given to police users in 2010 to pilot test and formally audit.

The current paper describes a more recent audit of EvoFIT’s performance, specifically for a 12-month period in Humberside police starting January 2010. It also indicates performance using the above ‘internals-first’ method of construction. In the final part of the paper, an overview is provided of more recent enhancements and system characteristics.

2. Humberside police

Humberside police is one of 43 forces in England and Wales. They are based in North-East England and comprise areas in the counties of Lincolnshire and Yorkshire, with major towns of Kingston upon Hull, Grimsby and Scunthorpe. To police this region, the force has four Basic Command Units, or Divisions.

2.1. Humberside’s previous composite system

Humberside police used the E-FIT feature system in investigations from 2007 to 2009. During this period, seven police officers worked with witnesses and victims to produce 56 composites for cases of murder, rape, sexual assault and burglary. These 56 E-FITs led to eight identifications in total—a name put forward that, in conjunction with other evidence, produced eight suspects. The level of suspect identification was therefore 14% (8 out of 56). In 2009, no identifications were produced and so the force considered alternative composite systems to make better use of police time, effort and money.

2.2. EvoFIT evaluation and field trial

In 2010, Humberside police conducted an initial evaluation of EvoFIT. Very-good performance resulted and EvoFIT was rolled out across the force; it was also deployed about twice as often annually than was the previous system. Identification of suspects was 60%, as described below.

2.2.1. Initial evaluation. It is normal for UK police forces to evaluate products before purchase, and facial-composite software normally follows this procedure. Two existing composite operators were trained on EvoFIT in January 2010 and evaluated the system for four months. During this time, six composites were constructed for burglaries, and one each for a sexual touching case and to locate a missing person. Five identifications (arrests) were made from names put forward from these images.

2.2.2. Operationalizing EvoFIT. Based on the success of the initial evaluation, eight further police officers were trained, to give appropriate cover in the four Divisions. Five of these officers had already been trained on E-FIT and so were given the same training as the first two officers. This was a 2-day course covering interviewing methods for face construction with EvoFIT (e.g. including free but not cued recall), instruction on EvoFIT procedures and holistic- and feature-manipulation tools, with in-class practice. Another training course was run for three further officers without previous composite-operator training. These trainees attended a 5 day course that also gave instruction on history of composites, UK
national police guidelines for composite production, exhibiting composites as evidence in court and use of appropriate artwork techniques—for adding extra detail such as moles, scars, unusual facial hair and tattoos.

2.2.3. Audit of performance. EvoFIT was deployed 35 times during the 12-month evaluation. Figure 4 presents the distribution of offences for composites constructed. Burglary in general represented about two-thirds (63%) of cases; the category of ‘Other burglary’ includes burglary with grievous bodily harm (GBH, one case), aggravated (involving use of a weapon, one case) and (attempted) distraction burglary (two cases). Robbery and assault (involving Actual Bodily Harm) equate to 11% each and the remaining cases (14%, ‘Other’) are for a Section 18 offence (causing GBH with intent), attempted rape, sexual touching, and for locating a missing person.

An audit was carried out by the force during the evaluation period. Out of 35 EvoFITs constructed, several composites were not used for identification (e.g. suspects subsequently identified by other means). In total, 21 suspects were directly identified from the EvoFITs. Identification of suspects by type of crime is shown in Table 1. Assault (with ABH) resulted in the highest successful identification (4 out of 4); the ‘Other’ category contained a single success, identifying a suspect in an attempted-rape case. For the largest group of composites constructed, for all types of burglary, the average identification was 59% (13 out of 22).

![Figure 4. Distribution of 35 crimes where an EvoFIT was constructed in the trial. Data are arranged by type of offence. See text for definitions and more details.](image)

Table 1. Identification of EvoFIT composites (expressed in percentage of total per category) as a function of type of crime. The categories are the same as those in Figure 4.

<table>
<thead>
<tr>
<th>Type</th>
<th>Burglary</th>
<th>Distraction Burglary</th>
<th>Other Burglary</th>
<th>Robbery</th>
<th>Assault</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>63%</td>
<td>50%</td>
<td>75%</td>
<td>75%</td>
<td>100%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Overall identification was 60% (21 out of 35). Using Chi-square statistics, this level of performance is significantly higher than the combined performance of EvoFIT (8 out of 56) used in force during the previous 3 years [$X^2(1) = 8.9$, $p = .003$]; the Odds Ratio (a standard measure of effect size) for this comparison is 4.2, meaning that an EvoFIT was 4.2 times more likely to lead to an identification than was an E-FIT.

Six identifications led to conviction—three were for distraction burglary, and one each for aggravated burglary, burglary in a dwelling and assault. The result is that 29% of identifications (6 out of 21) led to conviction. So, for all 35 EvoFIT composites produced in the evaluation, 17% of them (6 out of 35) led to conviction.

From the available data, the shortest interval from onset-of-crime to construction-of-EvoFIT was 1 day, and this occurred four times (13%). The longest was for a witness interviewed after two months (who produced two composites)—such an interval occurs occasionally in criminal investigations; for the following analysis, to avoid skewing the data, this outlying case is removed. For the remaining EvoFITs (Total, $N = 30$), the average interval to construction was 3.8 days (Standard Deviation, SD = 2.4 days). The average was very similar for the 21 cases involving identification of a suspect (Mean, $M = 3.8$ days, SD = 2.5 days) and for the 6 cases that led to conviction ($M = 3.7$ days, SD = 1.8 days). Intervals are not statistically different for cases with or without identification [$t(28) = 0.03$, $p = .97$], nor for identification with and without conviction [$t(19) = 0.2$, $p = .87$]. Overall, 50% of EvoFITs were constructed within two to three days of the crime, and 93% of them were constructed within a week.

2.2.4. Case studies. In this section, we consider cases which provide the strongest evidence that EvoFIT had correctly identified the perpetrator of a crime—that is, for investigations leading to conviction. For example, one case relates to a serious assault on the owner of a hotel. Here, one of the two suspects was identified through publication of the EvoFIT, and the enquiry revealed that the victim’s business partner had organised the attack.

Figure 5 illustrates composites constructed from three separate investigations. The EvoFIT is shown on the left and the person convicted in that case is shown on the right. The EvoFIT in the first row was produced by a female victim of distraction-burglary. In the second row, the EvoFIT was constructed by a male victim of aggravated-burglary. In this crime, the victim returned home to find a male burglar in his house. The burglar made attempts to assault him with an axe before making off. Both EvoFITs in the third and fourth row were produced by a female victim of distraction burglary.

---

2 At the time of writing, we were awaiting confirmation of dates from three cases that did not lead to identification of a suspect, and so our analyses for construction interval relate to 32 out of 35 investigations.

3 We use the established method [17] of removal based on 2 standard deviations from the mean (for non-identification cases).
3. Discussion

In the absence of suitable leads, it can be of vital importance for effective policing to obtain an identifiable image of an offender’s face from eyewitness memory. Systems for constructing composites have come and gone, but none have been capable of doing the one thing for which they were intended: to produce an image that another person could correctly identify (e.g. [2][3][5][14]). We have been attempting to achieve this objective since 1998. The result is a software system which synthesises arrays of faces from computer-generated models. Witnesses repeatedly select faces from such arrays, with ‘breeding’ of selected items, to allow a composite to be ‘evolved’. Afterwards, there are ‘holistic’ tools available to improve the overall likeness of the face and to manipulate individual features on request. In addition, the initial focus is on the important central part of the face, the region that is important for recognition of the composite by another person later.

In 2007, we developed an EvoFIT system that could produce composites that other people named at 25% correct [3]. Field trials of this system [13] revealed a similar figure for identifications leading to the arrest of a suspect; Figure 2 illustrates a high profile case that led to conviction at the time. The process used with EvoFIT has been improving [1], and a further field trial [14] found an arrest rate of about 40% when police interviewed witnesses using the simplified face-recall procedure outlined earlier.

In the current paper, we describe an evaluation of a more-recent version of EvoFIT. Over a 12-month period in 2010, 35 composites were constructed by Humberside police with witnesses and victims of mostly serious crime, about two-thirds of which were for burglaries. Twenty-one suspects were identified, or 60% of all EvoFITs constructed in the trial. This performance is over four times higher than composites from the previous system (E-FIT) used by the force. It was also found that six cases led to conviction, meaning that 29% of identifications led to conviction, or 17% of all EvoFITs. These latter figures reflect the situation that many cases do not have sufficient corroborative and physical evidence (DNA, fingerprints, fibres, etc.) to ensure a reliable conviction, and therefore do not proceed to court.

3.1. Further research and development

Further research has resulted in two notable developments for EvoFIT. The first is an improvement to the initial interview used immediately prior to face construction. It is normal for an operator to chat informally with witnesses, to build rapport, and then to ask them to think back to the crime and visualize the face, context reinstatement; next, if witnesses are able, they provide a free-recall description of the face. As
mentioned above, operators may also prompt witnesses to attempt further recall on individual features, cued recall—although this technique seems to be detrimental to face construction [14]. However, face recall in general may encourage witnesses to focus on individual features during face construction, which is unlikely to be an optimal strategy since an identifiable image requires accurate features along with accurate placement of features on the face—to give an overall (holistic) impression. The interview was enhanced to overcome this issue: after face recall, witnesses now focus on the overall properties or character of the face. This ‘holistic’ cognitive interview (H-CI) is straightforward to learn and administer, but allows a witness to produce a substantially more identifiable EvoFIT [18].

The second improvement is a new method for publishing composites in the media as part of a public appeal for information. The reader may recollect that animated composites are useful for TV and internet outlets [15], and were used in the current evaluation. The new technique is an additional secondary exhibit that can be used for the newspapers. It is called ‘perceptual-stretch’ and simply asks observers to look at the composite from the (left- or right-hand) side. When this is done, the face appears to be longer-and-thinner than normal—in other words, it is stretched vertically. Our cognitive system appears to deal with this situation by normalising the face and, in doing so, is less sensitive to some of the errors in the face, improving recognition.

In a recent laboratory test, the current version of EvoFIT was evaluated in conjunction with the H-CI and perceptual-stretch technique for identification. Forty volunteers looked at a short video clip of an unknown face, and 24 hours later constructed an EvoFIT composite of it either after the normal face-recall interview, or following an H-CI. As found before [3], EvoFITs were correctly named at 25% using the older techniques, but naming increased to 56% using H-CI, and then again to 74% using both H-CI and perceptual stretch.

3.2. Current version of EvoFIT

Laboratory research and police field trials have enabled creation of a system capable of producing very identifiable composites. Further field trials are currently in progress to measure the extent of the above and newest developments, but based on indications so far (previous paragraph), an arrest should now result in three out of every four cases where EvoFITs are produced; also, since 29% of these cases successfully proceed further (Section 2.2.3), this means that one in five EvoFITs should result in conviction (or 21%, calculated as 29% convictions for 74% identifications).

It would appear then that EvoFIT makes excellent use of police time and resources. There are now 60 individual face databases, for production of different offenders: male and female of different ages and races—White, Black, Asian, Eastern European, Chinese, Hispanic and various mixed-race combinations. A common procedure for face construction is used with each database using a holistic cognitive interview and holistic and feature tools; finished images can be published in the media as animated and/or perceptual-stretch composites.

4. Summary

About 15 years of intensive laboratory research and police fieldwork has been dedicated to developing EvoFIT. The designers have sought, using procedures as used in policework, to produce identifiable composites. The EvoFIT team has provided police training and after-sales (technical) support since 2007.

Here, we described results of the latest field trial with Humberside police in 2010. The force constructed 35 EvoFIT composites with witnesses in a range of serious crimes and identification of a suspect (an arrest) directly resulted from these images in 60% of cases; 29% of these arrests led to conviction. Put another way, for all composites produced in the evaluation, EvoFIT led to identification and then conviction once in every six cases. Further developments have been made to improve system performance further and field trials are under way to assess the current level of suspect identification and conviction. The project is an excellent example of the benefit to be gained by careful laboratory research coupled with collaborative formal field trials to develop an effective tool for policing.

5. References


