Interactive Evolutionary Generation of Facial Composites for Locating Suspects in Criminal Investigations

Christopher J. Solomon, Stuart J. Gibson*, Joseph J. Mist

School of Physical Sciences, Ingram Building, University of Kent, Canterbury, Kent, United Kingdom, CT2 7NH

Abstract

Statistical appearance models have previously been used for computer face recognition applications in which an image patch is synthesized and morphed to match a target face image using an automated iterative fitting algorithm. Here we describe an alternative use for appearance models, namely for producing facial composite images (sometimes referred to as E-FIT or PhotoFIT images). This application poses an interesting real-world optimization problem because the target face exists in the mind of the witness and not in a tangible form such as a digital image. To solve this problem we employ an interactive evolutionary algorithm that allows the witness to evolve a likeness to the target face. A system based on our approach, called EFIT-V, is used frequently by three quarters of UK police constabularies.

Keywords: Facial Composite, Interactive Evolutionary Algorithm, Statistical Appearance Model, Facial Identification, EFIT-V, Interactive Evolution Strategy

1. Introduction

In the event of a crime, police officers often rely on an eyewitness to provide a comprehensive account of the incident. In some circumstances, the witness has to convey a description of the suspect based only on a brief encounter. The pertinent question is how to accurately convey the suspect's face when the image only exists as a memory in the witness' mind? The accepted procedure is for a trained operator, usually a police officer, to work with the eyewitness in an attempt to produce a facial likeness to the suspect. A facial composite system is a software tool used by the operator that allows the expression of the facial appearance retained in the witness' memory as a digital image. The desired outcome is that the generated image be of sufficient accuracy that subsequent display to members of the public will result in recognition followed by the apprehension of the suspect. The probability of recognition is enhanced by providing additional supporting information such as the time, location and nature of the incident. Composite images have previously been matched to individuals contained in a database using a face recognition system [15]. Currently, this practice has not been widely adopted by users of facial composite systems.

Although there are many differences of detail between the various commercial composite systems available, until six years ago all commercial systems operated with the same feature-based philosophy. The approach was to select the individual features of the face from databases of examples which have been suitably labelled and categorized. The commercially available systems Electronic-Facial Identification Technique (E-FIT), Pro-FIT, Identikit, Comphotofit and Faces all fall into this category [7]. The effectiveness of the feature-based approach is fundamentally constrained by the witness' limited ability to provide detailed descriptions of facial appearance [14]. Further, there is strong evidence to suggest that faces are primarily encoded in memory as whole objects or as gestalts rather than as a collection of isolated features [24, 23]. One of the earliest innovations, which aimed to address some of the limitations associated with a finite database of candidate features, was the experimental system

*Corresponding author

Email addresses: c.j.solomon@kent.ac.uk (Christopher J. Solomon), s.j.gibson@kent.ac.uk (Stuart J. Gibson), jm441@kent.ac.uk (Joseph J. Mist)

developed by Brunelli and Mich named 'SpotIt!' [4]. This system relied on a Principal Component Analysis (PCA) model for each facial feature class (e.g. noses) achieving a reduction in the dimensionality of the problem and providing a basis from which novel features can be constructed by altering the contribution of each component. Its primary shortcoming was the mismatch between the calculated Principal Components (PCs) and the typical adjustments which real witnesses request, making it difficult to easily alter facial appearance in a satisfactory way.

Evolutionary techniques based on Darwinian theory that simulate complex structures and textures for use in computer graphics and animation were described as early as 1991 [19]. DiPaola [8], also working in the field of computer graphics, describes such an algorithm for facial appearance. In DiPaola's work, crude facial representations were encoded using 25 element parameter vectors and then evolved using an interactive genetic algorithm. DiPaola's work allows the representation of a person's face as a point in a multidimensional face-space. The concept of face-space also exists in the psychological literature and was first proposed by Valentine [25] as a model for human memory of faces. It has been shown that PCA models encapsulate some aspects of human face perception [13].

Motivated by the early work on evolutionary approaches to facial composite construction [14, 11], Kurt et al. [16] conducted a study which demonstrated that a number of Interactive Evolutionary Algorithms (IEAs) could in principle be used in an evolutionary facial composite system. The IEAs compared were two different interactive genetic algorithms, an Interactive Evolution Strategy (IES), an interactive particle swarm optimization algorithm and an interactive differential evolution algorithm. All of the algorithms tested produced composites that were recognizable in most cases.

Conceived originally as a research system operating under the name EigenFIT [11] and assuming its first commercial form [12, 20] in 2006, the EFIT-V system [2] described in this paper is the first system using evolutionary principles to receive wide commercial acceptance. EFIT-V combines a parametric model of facial appearance with an interactive evolutionary search within the parametric space. EFIT-V also represents the first viable composite system to combine full colour face imagery with a simple interactive mechanism for the witness to communicate the appearance of a suspect's face. The commercial EFIT-V system has since been refined and developed in many aspects of system operation to improve field performance. In this paper, we provide an account of recent developments and an overview that encapsulates the essence of the system from technical and operational perspectives. The first core element of the system is a statistical appearance model of human faces whose construction is described in Section 2. The second is an IEA which manipulates the parameter vectors defined by the statistical appearance model. Our work to develop and evaluate an effective search method is described in Section 3. The use and real-world performance of the system is described in Section 4. In Section 5 we outline recent, preliminary studies to improve the performance of the search algorithm. Finally, we summarize the effectiveness of our approach and outline future research directions in Section 6.

2. Mathematical model of facial appearance

'Active' appearance models have been used for computer vision applications [6, 26, 22, 17] in which images were synthesized and morphed to match a target object using an automated iterative fitting procedure. The method incorporates the frequently used data reduction technique of Principal Components Analysis (PCA). In our facial composite application the fitting procedure is guided by the witness' response to face stimuli. A training sample of face images is used to construct a statistical Appearance Model (AM) and thereby define a parametric search space which we refer to herein as face-space. Provided that the training sample is sufficiently comprehensive across age, gender, ethnicity and other variables, we posit that a satisfactory approximation to any desired face may be represented by a point, or localized contiguous region, within face-space. The model leads to a distribution of parameters that is very well approximated by an independent, multivariate, Gaussian probability density function. New instances of facial appearance, which constitute the candidate solutions to our problem, may be conveniently synthesized by randomly sampling from the learned distribution. The construction of an acceptable facial likeness therefore reduces to a search for the corresponding parameter vector within the learned space and conversion into the corresponding facial composite through the equations described in this section. The crucial advantage of the AM is the reduction in

the dimensionality of the search problem. We have established through empirical means that parameter vectors of 60 real variables are sufficient for representing facial appearance to quasi-photographic quality — a substantial reduction compared to the original image pixel space.

Our AM was constructed according to the method first described by Cootes et al. [6]. A training sample of 2729 images comprising males from nine distinct ethnic groups and females from five distinct ethnic groups was annotated using a point model that delineates the face shape and the internal facial features. In this process, 22 key landmark points were initially placed by hand on each face image. Polynomial spline sections, connecting the key landmarks, were positioned on the feature boundaries using control points. Additional landmark points were subsequently obtained by sampling equidistant points on the fitted spline sections to define the 190 points of the complete point model. For each face, landmark co-ordinates (x, y) were concatenated to form a shape-space vector, $\mathbf{x}$. The data was centred by subtracting the mean face shape, $\bar{\mathbf{x}}$, from each observation. The shape PCs were derived from the set of mean subtracted observations and placed in the columns of the matrix $\mathbf{P}_s$. A key property of the PCA is that the PCs are ordered by decreasing importance. For object classes that contain highly correlated data, such as the human face, most of the variance of the original data is represented by a relatively small number of PCs compared to the total number of observations. In our application a face shape is approximated by a linear combination of the first 60 PCs only. The synthesis of a face shape from the shape PCs is represented by the following equation

$$\widehat{\mathbf{x}} = \mathbf{P}_s \mathbf{b}_s + \bar{\mathbf{x}} \tag{1}$$

where $\widehat{\mathbf{x}}$ is an approximation to a specific individual's face shape and $\mathbf{b}_s$ is a vector in which the first 60 elements are normally distributed parameters that determine the linear combination of shape PCs and the remaining elements equal to zero.

A useful preprocessing step in the treatment of the pixel data is to warp each face image to the sample mean face shape. Shape normalization is achieved by calculating a Delaunay triangulation of the landmarks followed by a piecewise affine shape transformation [21]. This non-rigid shape alignment procedure greatly improves feature correspondence between images and consequently reduces

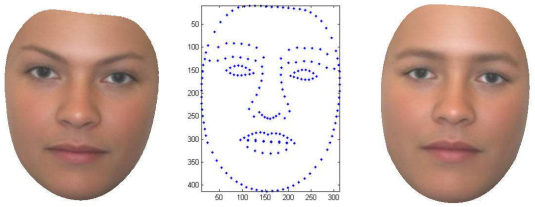


Figure 1: From left to right: segmented face, point model and shape normalized face (texture map).

blurring in the images generated by the model. The basic procedure is summarized in Figure 1. The next step in the construction of the appearance model is to derive a set of texture PCs from the sample of training images. For each face in the training sample, texture vector, $\mathbf{g}$, comprising the RGB values of every pixel in the shape normalized image is extracted. The 60 retained PCs are then arranged in order to form the columns of the matrix $\mathbf{P}_g$. The equation for expressing a texture vector in terms of the texture PCs follows the same form as in the case of shape:

$$\widehat{\mathbf{g}} = \mathbf{P}_g \mathbf{b}_g + \bar{\mathbf{g}}. \tag{2}$$

A weak correlation exists between the face shape and face texture. This provides the motivation to perform a third and final PCA on the concatenated shape and texture parameter vectors, thereby providing a highly compact representation of the human face. Using the AM we construct a compact vector representation, $\mathbf{c} = [c_i, c_2, \ldots, c_n]^T$ of the appearance of an out-of-sample face. The relationship between these appearance parameters and the corresponding shape, $\mathbf{x}$, and texture, $\mathbf{g}$, is expressed as follows,

$$\mathbf{c} = \mathbf{Q}^T \begin{bmatrix} w\mathbf{b}_s \\ \mathbf{b}_g \end{bmatrix} \equiv \mathbf{Q}^T \begin{bmatrix} w\mathbf{P}_s^T (\widehat{\mathbf{x}} - \bar{\mathbf{x}}) \\ \mathbf{P}_g^T (\widehat{\mathbf{g}} - \bar{\mathbf{g}}) \end{bmatrix} \tag{3}$$

where the columns of $\mathbf{Q}$, $\mathbf{P}_s$ and $\mathbf{P}_g$ are the appearance, shape and texture PCs of the sample data respectively. $(\widehat{\mathbf{g}} - \bar{\mathbf{g}})$ is a face texture vector in mean subtracted form and $(\widehat{\mathbf{x}} - \bar{\mathbf{x}})$ is the face shape vector in mean subtracted form. $\mathbf{b}_s$ and $\mathbf{b}_g$ are vectors of shape and texture parameters respectively and w is a scalar that determines the relative significance of shape and texture [21]. In the context of this work, the appearance PCs are the dimensions of a mathematical representation of face-space. In Figure 2 we show the first few PCs of face shape and texture which result from the model building stage.

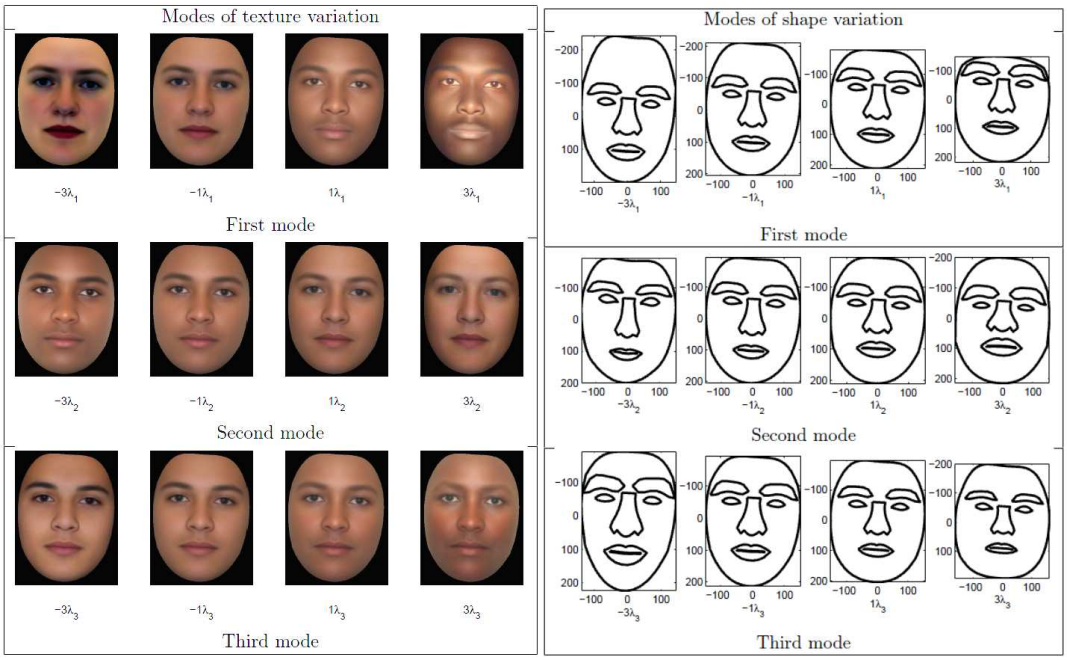


Figure 2: Illustration of the shape and texture principal modes of variation extracted from the training sample of faces. Here negative and positive multiples of each mode have been added to the mean, indicating their effect on the synthesized face. Note how the modes affect multiple aspects of the face simultaneously — e.g. the third shape mode significantly affects head shape ('long and pointed' to 'square'), eyebrows ('thick' to 'thin') and mouth shape ('full and curved' to 'thin and straight')

New examples of faces are synthesized by sampling the parameters $\{c_i\}$ from a multivariate distribution and then manipulating Equations 1–3 to obtain the corresponding face shape and face texture. The final step is to arrange the elements of $(\widehat{\mathbf{g}})$ into the columns of an image and warp to the correct face shape defined by $\widehat{\mathbf{x}}$. To generate plausible faces, the Probability Density Function (PDF) of appearance parameter values must be estimated. This can be approximated by a standard multivariate normal,

$$N(\mathbf{c}; \mathbf{0}, \Lambda) = (2\pi)^{-\frac{n}{2}} |\Lambda|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left(\mathbf{c}^T \Lambda^{-1} \mathbf{c}\right)\right\}$$

where Λ is a diagonal variance matrix due to the statistical independence of the PCs. Figure 3 shows some examples of randomly generated faces using EFIT-V (Hispanic male, White male, Arab male and Black female). As is evident, the faces show a near-photographic realism, unlike previous systems.



Figure 3: Random generation of faces in EFIT-V through sampling a statistical model of facial appearance. Examples are shown for Hispanic male, White male, Arab male and Black female.

3. Interactive search algorithm

The AM outlined in the previous section provides the means for synthesizing plausible face im-

ages. To be suitable for application in the field of operation, the method for composite construction employed must achieve a good representation of the face in the witness' memory without imposing a cognitive overload on the witness or the operator. It is important to recognize that any optimum search procedure in this sense must be constrained by these very practical considerations. To reduce the cognitive burden, rather than have the witness and operator generate a composite from a description, groups of faces are generated which the witness then evaluates subjectively for their similarity to the target face. Some practical means of searching the face-space for the appropriate values of the parameter vector is therefore required. We address this problem using an IEA and the details of our algorithm are presented in the remainder of this section.

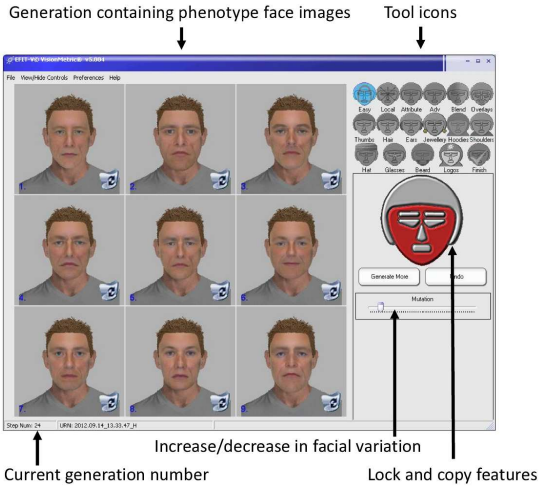


Figure 5: Screen shot of EFIT-V's user interface.

3.1. Algorithm development

The need for cognitive simplicity, time constraints and the effects of witness fatigue mean that any interactive search algorithm that is suitable for use by witnesses to crimes is constrained in several ways. The number of candidate solutions (i.e. faces) presented in each generation must be modest and the fitness evaluation applied by the eyewitness must be simple. The speed of convergence should be related entirely to the time required for the witness to achieve an acceptable likeness to the target face and not to computational speed.

Three basic approaches were explored in the development of EFIT-V. A detailed description of the relative performance of these methods can be found in [18]. In the first of these, each member of the population was rated on a scale of 0–10 for its similarity to the target face. Standard crossover methods were then applied in proportion to the assigned scores to breed subsequent generations. This algorithm was time-consuming and inconsistencies in scoring led to erratic performance. A second very simple algorithm required the witness to repeatedly select the best of two faces at each step of the process. The selected face was copied and randomly mutated. The copy was then displayed alongside its parent. The mutation pressure was left as an adjustable parameter. This algorithm proved unsatisfactory due to the long time required to attain a satisfactory likeness. The third approach, which we termed Select-Multiply-Mutate (SMM), requires the witness to select a single, elite member from a population of nine to seed the following generation. This algorithm offered the best compromise between cognitive simplicity and the volume of information transferred through the human computer interface at each iteration. Details of our SMM evolutionary algorithm, used in EFIT-V, are given in Table 1.

3.2. Operation of EFIT-V

A summary of the procedure for creating a facial composite using EFIT-V version 5.004 is detailed in Figure 4. The system is initialized using a pseudo-random number generator to obtain nine genotypes each comprising 60 real, random numbers drawn from a standard normal distribution. A face image (phenotype) is synthesised from each of the genotypes as described in Section 2. A screen shot of EFIT-V's user interface is given in Figure 5. The witness then views the array of nine faces and is required to select the single face, which we call the 'stallion', that most closely resembles the suspect. Optionally, the user may first hide one or more faces deemed to be poor likenesses. This appears to have real psychological benefits for many witnesses. The genotype corresponding to the stallion is duplicated eight times, thereby copying the genetic code of the selected face into a new generation of nine individuals. The cloned genotypes are mutated but the stallion is left unaltered.

Extensive studies [18] using a simulated witness established an optimal starting value of the mutation probability, $p = 0.05$. The mutation probability decays by 0.001 per generation, although the

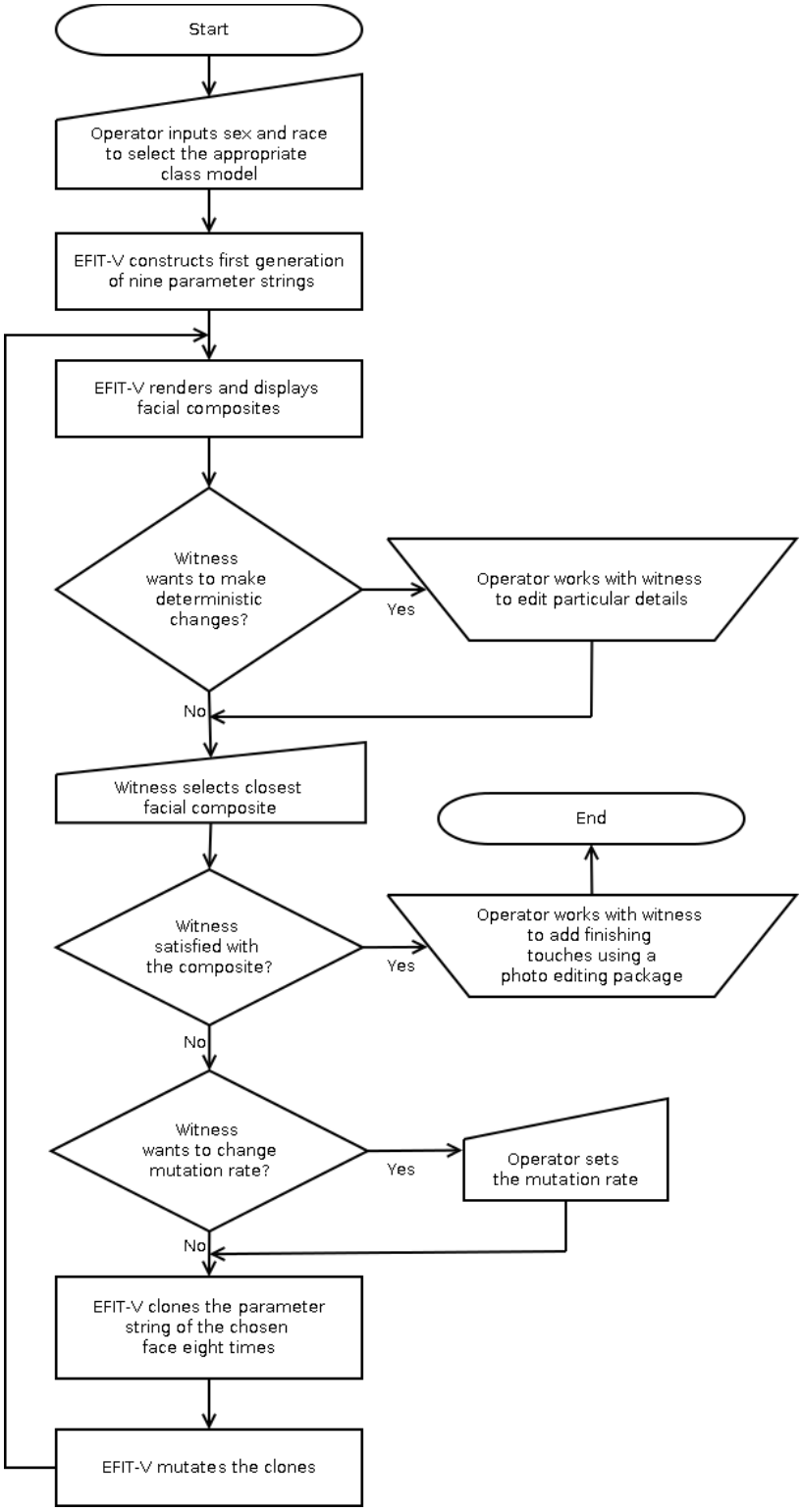


Figure 4: Flowchart summary of the process of creating a composite with EFIT-V

Term	Definition
Phenotype	The phenotypes are the face images synthesised from the appearance model.
Genotype	The genotype is a parameter vector of 60 real numbers that is used to encode the phenotype face image.
Population	The population is the nine faces that are displayed to the witness each time they choose the face with the closest match to their remembered face.
Fitness function	EFIT-V has no fitness function; the fitness of the faces is determined by the opinion of the witness.
Parent selection	In EFIT-V a particularly elitist strategy is employed in which only one individual is selected per generation to be parent to the next generation. This is partly driven by the strong psychological need for witnesses to progress to increasingly good likenesses without any intermediate retrograde steps.
Mutation	In EFIT-V this is achieved by adding a random number to some of the values in the genotype. More details are given in Section 3.2.
Generation	In EFIT-V each set of nine faces displayed to the witness represents a new generation.

Table 1: Summary of the IEA for the SMM algorithm

probability can be adjusted manually by the operator if the witness thinks that the faces shown are too similar or too different from each other. Each one of the 60 elements $\{c_i\}$ constituting the genotype $\mathbf{c}$ is mutated with a probability p. If $\mathbf{r}$ is a vector of random numbers of the same length as $\mathbf{c}$, where $r_i \sim U(0,1)$, the elements $\{c'_i\}$ of the new genotype $\mathbf{c'}$ are given by

$$c'_i = \begin{cases} c_i & \text{if } r_i \geq p \\ c_i + N(0,1) & \text{if } r_i < p. \end{cases}$$

From these new genotypes, the phenotypes are constructed and displayed to the witness as a new generation. The stallion from the previous generation is placed randomly amongst the mutated clones in order to avoid any positional bias whereby the witness' eye gaze becomes fixated on a single face location such that the remaining faces in the generation are not given due consideration.

3.3. Deterministic modification of facial appearance

Our SMM algorithm allows the construction of a facial composite even when the witness is unable to provide a detailed description. In practice, the witness may recall a good description of one or more features — especially if those features are distinctive. In these circumstances a direct deterministic method is more effective than evolving the desired effect over a number of generations. Our approach is to effect deterministic alterations to facial appearance in one of three ways: intra face-space transforms, extra face-space transforms and image layers.

We use the term intra face-space transform to describe the process of 'pushing' a face in a predefined direction through the face-space [12, 10] to achieve a holistic (whole face) increase or decrease of a facial attribute such as a subject's age or their weight. A direction through the face-space, that maximizes the perceived increase in a particular facial attribute, is learnt using a linear regression model in which a sample of training faces are assigned attribute scores by human scorers. Attributes modelled this way can be perceptual traits as well as physical facial characteristics. Examples of perceptual traits modelled using this intra face-space method include hostility, health [5], angularity, hardness, happiness, kindness, weight and friendliness. These alterations to facial appearance are achieved in software using slider controls.

An extra face-space transform is any modification in appearance that results in a face which lies outside the span of the appearance model's PCs. In general, alterations to the position and aspect ratio of individual facial features can be categorized as extra face-space transforms. For example, the witness may request that the nose is made wider and this can be achieved using EFIT-V's Local Feature Tool in which the feature is adjusted using push button controls. Other operations in this category include copying a feature from one face into the

other faces of the current generation and fixing the appearance of one or more features over consecutive generations. These tools use the witness' inherent general knowledge of facial appearance to ensure that alterations result in plausible facial appearances.

Other important aspects of appearance which cannot be modelled effectively using the statistical approach of Section 2 include additions such as hairstyle, facial hair and effects bearing on the complexion of the skin including wrinkles. These additions are included as database items and can be superimposed onto the composite as an image layer using an alpha channel to control transparency.

4. Use and performance of the system

Although the only true test of the performance of the system is through use in real cases, it is possible to estimate the efficiency of the IEA described in Section 3 using the simulated witness model. Simulating the witness' response to the displayed images allows us to estimate the optimal parameters of the evolutionary algorithm without the need for extensive human trials. The simulated witness works in such a way that in each generation the computer selects the face which lies closest to a known target face — simulating the behaviour of an ideal witness. The distance measure employed in our system was the Mahalanobis distance and the algorithm terminates when the distance between the 'stallion' image and the target falls below a recognition threshold. In a study involving 1000 randomly generated target images, using only the SMM algorithm and no deterministic modifications, an acceptable likeness to the target face was achieved on average within just over 42 generations and a standard deviation of 17 generations.

In practice, because EFIT-V has a variety of tools which allow direct manipulation of the faces, an acceptable composite can generally be achieved in considerably fewer generations. The time taken in real cases depends on many highly variable factors such as the witness' memory, emotional state and willingness as well as operator skill. The extensive human input and experimentation required to estimate a meaningful average for the number of iterations required to achieve a good facial composite are prohibitive. However, extensive field use and feedback over a period of six years from police forces using the system routinely, suggests that by combining the SMM algorithm with the system's other feature manipulation capabilities, a satisfactory composite can normally be achieved in between 2 and 25 generations.

West Yorkshire police reported that suspects were correctly identified from their EFIT-V composite images in 55% of cases. This naming rate was determined over a twelve month period spanning 2010 to 2011 and encompassed more than 400 interviews. In contrast, some academic studies have suggested that success rates using traditional feature-based systems may be as low as 5% on average [9]. Functional developments of the EFIT-V system have also been, and will continue to be, driven by the real-life requirements of operators and witnesses but also by advances in our understanding of facial processing and image processing.

At the time of writing, EFIT-V is in routine use by over 60% of the UK's police forces and in 7 other countries including the USA and Canada. EFIT-V facial composites and corresponding custody suite images for two suspects are shown in Figure 6. These images relate to robberies that took place in 2012 in the Merseyside area of the UK. In both cases the composite image was constructed by a trained operator, under the guidance of the witness to the crime, in accordance with the procedure outlined in Association of Chief Police Officer's Facial Identification Guidance document [1].

5. Towards shaping the search space

In Section 3, we described the SMM algorithm used in the commercial version of the system. The SMM algorithm has been found empirically to establish a good compromise between simplicity of operation, in which a human witness can easily make the required decisions, and fast convergence properties. However, we now draw explicit attention to the fact that our implementation of the SMM algorithm provides the opportunity for a witness to explicitly reject one or more faces in each generation. The ability to reject weak members in each generation appears to have some psychological benefit to witnesses as it is often easier for the witness to choose the worst examples and thereby simplify the task of selecting the preferred face. It is apparent that the rejection of certain faces provides definite information (in effect, "don't show that face again or one that it close to it") which may, in principle, be harnessed to accelerate convergence to a satisfactory solution. The essence of the approach we have explored is as follows.



Figure 6: Examples of EFIT-V composite images produced in the field of operation. Both suspects were recognized by members of the public immediately after the composites were released.

Candidate solution vectors are real-valued, N-dimensional genotypes $\mathbf{c} = [c_1, c_2, \cdots, c_N]$ and there exists a target vector corresponding to the 'ideal' solution denoted by $\mathbf{c}^t = [c_1^t, c_2^t, \cdots, c_N^t]$. In each generation, the user (witness, in the case of EFIT-V) selects a preferred candidate solution with corresponding genotype $\mathbf{c}^s = [c_1^s, c_2^s, \cdots, c_N^s]$. The witness may optionally reject one or more candidate solutions with the i-th rejected potential solution having corresponding genotype $\mathbf{c}^{r_i} = [c_1^{r_i}, c_2^{r_i}, \cdots, c_N^{r_i}]$.

Consider a point P lying on the line $\mathbf{w} = \mathbf{c}^s - \mathbf{c}^{r_i}$ and distance $\alpha |\mathbf{w}|$ from point $\mathbf{c}^{r_i}$. We construct an N-dimensional hyperplane which passes through point P and which is orthogonal to the line $\mathbf{w}$. The hyperplane defines a discriminant function $f(\mathbf{c})$ which has the form

$$f(\mathbf{c}) = \mathbf{w} \cdot \mathbf{c} + \omega_0$$

where

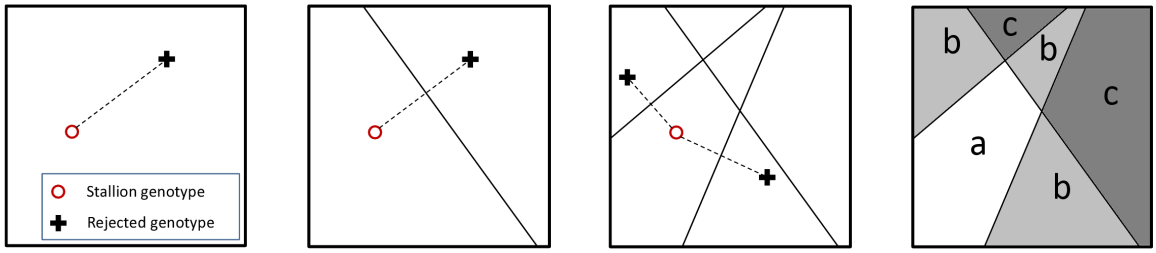
$$\omega_0 = -\left(\mathbf{w} \cdot \mathbf{x}^{r_i} + \alpha |\mathbf{w}|^2\right).$$

The discriminant function divides the space into two mutually exclusive regions: $\mathbb{R}_s^N$ the region in which $\mathbf{c}^s$ is located and $\mathbb{R}_{r_i}^N$ the region in which $\mathbf{c}^{r_i}$ is located. In general, for an arbitrary genotype $\mathbf{c}$, we have

$$\begin{aligned} f(\mathbf{c}) > 0 \quad &\Leftrightarrow \quad \mathbf{c} \in \mathbb{R}_s^N \\ f(\mathbf{c}) \leq 0 \quad &\Leftrightarrow \quad \mathbf{c} \in \mathbb{R}_r^N. \end{aligned}$$

After we have constructed the discriminant function, any subsequent genotype produced by the SMM procedure which satisfies $f(\mathbf{c}) \leq 0$ has its phenotype conversion probability correspondingly reduced through multiplication by factor p_r where $0 \leq p_r \leq 1$. In this way, the probability landscape is successively modified to favour the generation of genotypes from within those regions of the search space lying closer to $\mathbf{c}^s$. The essentials of the algorithm are given in Figure 7.

The fundamental assumption in our approach is that the preferred face in the generation, which has genotype $\mathbf{c}^s$, will lie closer to the target vector $\mathbf{c}^t$ than the rejected vector $\mathbf{c}^{r_i}$. If this assumption were strictly valid, it would guarantee that the target face $\mathbf{c}^t$ always lies within $\mathbb{R}_s^N$. Under these circumstances, we would be justified in setting $\alpha = 1/2$ and $p_r = 0$ thereby maximally reducing the volume of the search space at each step. However, this assumption does not always hold. There are two reasons for this. The first is that the search space



(a) (b) (c) (d)

Figure 7: The essentials of the hyperplane algorithm in a simple 2D space. (a) The current stallion and a genotype explicitly rejected by the user. (b) A hyperplane is placed perpendicular to the vector between the stallion and the genotype. (c) Two more genotypes (for example) are rejected by the user, hence two more hyperplanes are added to the search space. (d) A genotype generated in region 'a' has no chance of being rejected by the algorithm. A genotype generated in a region 'b' is accepted with a probability p_r. A genotype generated in a region 'c' is accepted with probability p_r^2.

is likely to be non-linear so that two phenotypes may seem to be equally similar to the target phenotype when their genotypes are not equidistant from the target genotype in the search space. The second reason is that due to the human threshold of perception, there is uncertainty in the perceptual distance between phenotypes. As well as accounting for these factors by setting p_r to be non-zero, the hyperspace approach can be made more 'forgiving' by setting a lifetime, l, for each hyperplane so that after l generations the hyperplane is removed.

Intuitively, it is reasonable to assume that progressively small (or even negative) values for α, which move the hyperplane closer to $\mathbf{c}^{r_i}$, will progressively increase the likelihood that $f(\mathbf{c}^t) > 0$ and thus that $\mathbf{c}^t \in \mathbb{R}_s^N$. However, increasing α will generally reduce the volume of the search space which is partially suppressed.

In an attempt to establish the viability of this approach on a lower dimensional problem, we implemented this algorithm on the interactive colour matching task described by Breukelaer et al.[3]. In this task the participant was required to repeatedly select the colour swatch which best matched their subjective perception of a target colour from each generation of nine colours. The selected colours were used to synthesise the colours for the following generation. The process would continue until the participant was satisfied with their colour match. This task was chosen for two reasons. The first is that it is cognitively simple, which means that effects due to fatigue are less likely. The second reason is that because the search was conducted in the CIELab colour space the Euclidean distance between two colours corresponded to their perceptual distance.

To find optimal values of α, p_r and l for the colour matching task a realistic, non-ideal, simulated participant was constructed. The behaviour of the simulated participant was based on human responses collected from 500 generations of evaluation of the hyperplane SMM algorithm detailed above. Based on simulations run with the simulated participant, the values of the three parameters were set at $\alpha = 0.55$, $p_r = 0.6$ and $l = 4$.

In the experiment there were three treatments: the basic SMM algorithm described in Section 3, the hyperplane SMM algorithm and a 'dummy' algorithm which was indistinguishable from the hyperplane SMM algorithm from the participants' perspective but in fact made no use of the participants' rejections.

A total of twenty-four people participated in the experiment. Each participant completed six runs, one for each of the three treatments trying to match a shade of orange and one for each of the three treatments trying to match a shade of green. After completing the three runs on the first colour, and then after completing the three runs on the second colour, the participant was asked to rank the algorithms. The algorithms were ranked on: the perceived time taken to achieve a colour match, their ease of use and how much control the participant felt they had over the algorithms. As well as the subjective data from the participants, objective data were also gathered: the number of generations required to complete each run, the time taken and the proximity of the final estimated colour to the target colour.

The results of the experiment are summarized in Table 2. The most important measure is the number of generations required to produce a satisfac-

	Basic SMM	Hyperplane SMM	Dummy SMM
Number of generations	20.8	17.9	22.8
Time taken (s)	149	169	181
Final distance (CIELab colour space)	3.19	3.29	3.38
Perceived time taken (ranked)	1.90	1.94	2.17
Perceived ease of use (ranked)	1.88	1.90	2.23
Perceived control (ranked)	2.42	1.60	1.98

Table 2: Means of the measured variables. For the ranked variables, a low mean rank score is better than a high mean rank score.

tory match since this relates directly to the cognitive load placed on the observer. In this we observe a modest increase in efficiency using the hyperplane SMM algorithm. It is also interesting to note that participants generally felt more in control of the search procedure using the hyperplane and dummy SMM algorithms than they did the basic SMM algorithm. The experiment described in this section is preliminary but appears promising as a basic approach to improving the convergence of the search as defined by our specific problem. Future work will seek to extend this and understand its behaviour in higher dimensional search spaces such as face-space models.

6. Summary and future development

The elements of the EFIT-V system have been presented. In this system, a statistical appearance model of human facial appearance was constructed and an interactive evolutionary algorithm was employed to enable witnesses to crimes to create the facial appearance of a suspect. The conceptual simplicity of this method and its inherent appeal to facial recognition capacity is its major strength. In essence, the witness is only required to make simple decisions in response to the facial stimuli. Field use of the system by European police forces has indicated significant improvements in identification rates and ease of use.

In respect of ongoing improvements to the performance of EFIT-V, there appears to be scope for continued research into more effective and efficient search methods. In particular, preliminary work suggests that there may be considerable potential for faster attainment of the target identity by using explicitly rejected faces to weight the search space. Composites generated using the system can exhibit near photo-realistic quality and it would be interesting to compare their performance with feature based systems when used as input to a face recognition system.

References

[1] Facial identification guidance document. Produced on behalf of the Association of Chief Police Officers by the National Policing Improvement Agency, Specialist Operations Centre, Wyboston Lakes, Great North Road, Wyboston, Bedfordshire, UK, 2009.

[2] EFIT-V. `http://www.visionmetric.com`, Sept 2012. Accessed: 21/09/2012.

[3] R. Breukelaer, M. Emmerich, and T. Bäck. On interactive evolution strategies. In Franz Rothlauf et al, editor, *Applications of Evolutionary Computing, EvoWorkshop2006: EvoINTERACT*, volume 3907, pages 530–541. Springer-Verlag GmbH LNCS, 2006.

[4] R. Brunelli and O. Mich. SpotIt! an interactive identikit system. *Graphical models and image processing: GMIP*, 58(5):399–404, 1996.

[5] David Clarke, Marianne Riggs, and Stuart Gibson. Prototyping perceptions of health for inclusion in facial composite systems. In *International Conference on Emerging Security Technologies*, 2010.

[6] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. In H.Burkhardt and B. Neumann, editors, *Proceedings of the European Conference on Computer Vision*, volume 2, pages 484–498. Springer, 1998.

[7] G.M. Davies and T. Valentine. *Handbook of eyewitness psychology.Volume 2: Memory for people.*, chapter Facial composites: forensic utility and psychological research., pages 59–83. Mahwah: LEA, 2007.

[8] S. DiPaola. FaceSpace: a facial spatial-domain toolkit. In *Proceedings of the Sixth International Conference on Information Visualisation*, 2002.

[9] C.D. Frowd, P.J.B. Hancock, V. Bruce, A. McIntyre, M. Pitchford, R. Atkins, W. Webster, J. Pollard, B. Hunt, E. Price, S. Morgan, A. Stoika, R. Dughila, S. Maftei, and G. Sendrea. Giving crime the 'evo': catching criminals using EvoFIT facial composites. In G. Howells, K. Sirlantzis, A. Stoica, T. Huntsberger, and A.T. Arslan, editors, *IEEE International Conference on Emerging Security*, 2010.

[10] S.J. Gibson, C.M. Scandrett, C.J. Solomon, C. Wilkinson, and M.I.S. Maylin. Computer assisted age pro-

gression. *Journal of Forensic Science, Medicine and Pathology*, 5(3):174 181, 2009.

[11] S.J. Gibson, C.J. Solomon, and A. Pallares Bejarano. Synthesis of photographic quality facial composites using evolutionary algorithms. In Richard Harvey and J.Andrew Bangham, editors, *British Machine Vision Conference 2003*, volume 1, pages 221–230, 2003.

[12] S.J. Gibson, C.J. Solomon, M.I.S. Maylin, and C. Clark. New methodology in facial composite construction: from theory to practice. *International Journal of Electronic Security and Digital Forensics*, 2:156–168, 2009.

[13] P. Hancock, A. Burton, and V. Bruce. Face processing: human perception and principal components analysis. *Memory and Cognition*, 24:26–40, 1996.

[14] P.J.B Hancock. Evolving faces from principal components. *Behavioral Research Methods, Instruments and Computers.*, 32(2):327–333, 2000.

[15] W. Konen. Comparing facial line-drawings with gray-level images: A case study on phantomas. In C.V.D. Malsburg and W.V. Seelen, editors, *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*, Bochum, North-Holland, 1996.

[16] B. Kurt, A. S. Etaner-Uyar, T. Akbal, N. Demir, A. E. Kanlikilicer, M. C. Kus, and F. H. Ulu. Active appearance model-based facial composite generation with interactive nature inspired heuristics. In B. Gunsel et al., editor, *International Workshop on Multimedia Content Representation, Classification and Security*, volume 4105 of *LNCS*, pages 183–190. Springer-Verlag, 2006.

[17] X. Li, S.Maybank, S.Yan, D. Tao, and D.Xu. Gait components and their application to gender recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(2):145–155, Mar. 2008.

[18] A. Pallares-Bejarano. *Evolutionary Algorithms for Facial Composite Synthesis*. PhD thesis, University of Kent, 2006.

[19] K. Sims. Artificial evolution for computer graphics. *Computer Graphics*, 25(4):319–328, 1991.

[20] C.J. Solomon, S.J. Gibson, and M.I.S. Maylin. *A New Computational Methodology for the Construction of Forensic, Facial Composites*, volume 5718/2009, pages 67–77. Springer-Verlag LNCS, August 2009.

[21] M. B. Stegmann. Active appearance models: Theory, extensions and cases. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, aug 2000.

[22] Mikkel B. Stegmann. Object tracking using active appearance models. In Sren I. Olsen, editor, *Proceedings of the 10-th Danish Conference on Pattern Recognition and Image Analysis*, volume 1, pages 54–60, Copenhagen, Denmark, July 2001.

[23] J.W Tanaka and M.J. Farah. Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, 46A:225–245, 1993.

[24] J. Turner, G. Pike, N. Towell, R. Kemp, and P. Bennett. Making faces: Comparing E-FIT construction techniques. *Proceedings of The British Psychological Society*, 7(1):78, 1999.

[25] T. Valentine. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43A:161–204, 1991.

[26] J. Wan, X. Renm, and G. Hu. Automatic red-eyes detection based on AAM. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, volume 7, pages 6337 – 6341, 2004.