EFIT-V - Interactive Evolutionary Strategy for the Construction of Photo-Realistic Facial Composites

Ben George
School of Physical Sciences
University of Kent
Canterbury, United Kingdom
bg37@kent.ac.uk

Dr. Stuart J. Gibson
School of Physical Sciences
University of Kent
Canterbury, United Kingdom
s.j.gibson@kent.ac.uk

Dr. Matthew I.S. Maylin
School of Physical Sciences
University of Kent
Canterbury, United Kingdom
m.maylin@kent.ac.uk

Dr. Christopher J. Solomon
School of Physical Sciences
University of Kent
Canterbury, United Kingdom
c.j.solomon@kent.ac.uk

ABSTRACT

Facial composite systems are used to create a likeness to a suspect in criminal investigations. Traditional, feature-based facial composite systems rely on the witness' ability to recall individual features, provide verbal descriptions and then select them from stored libraries of labelled features - a task which witnesses often find difficult. The EFIT-V facial composite system is based on different principles, employing a holistic (whole face) approach to construction.

The witness is shown a number of randomly generated faces and is asked to select the one that best resembles the target. A genetic algorithm is then used to breed a new generation of faces based upon the selected individual. This process is repeated until the user is satisfied with the composite generated.

This paper describes the main components and methodology of EFIT-V and showcases the strengths of the system.

Categories and Subject Descriptors

I.4.9 [**Image Processing and Computer Vision**]: Applications; I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*Industrial Automation*

General Terms

Human Factors

Keywords

Interactive Evolutionary Algorithm, Appearance Models, Facial Composites, Facial Synthesis

1. INTRODUCTION

Facial composite systems are used in criminal investigations as a means for a witness to generate a likeness to a suspect. By generating a composite image of sufficient likeness it is hoped that subsequent display to members of the public will result in recognition and capture. Current systems use a process which relies on the witness' ability to recall and describe individual facial features, select the best choice from a large library and arrange them in the correct spacial configuration. Although earlier feature based systems such as IdentiKit I and PhotoFIT have undergone substantial improvements the underlying process remains the same. There are two main drawbacks to the feature-based approach. Firstly, previous research has demonstrated the shortcomings of recall as a means of identification and it has been suggested that the weakest link in the generation of facial composites is the need for a witness to recall and verbally describe the suspect's face to the operator [9]. Secondly, a considerable body of evidence now suggests that the task of recognition and synthesis does not lend itself to simple decomposition into features but is a global process relying on the inherent spatial/textural relations between all features in the face [1].

We reason that the design of an effective facial composite system should be based upon the witness' ability to perform different tasks in facial processing and should be able to produce realistic and accurate composite faces. Based on previous work by by Gibson et al. [4] and Pallares-Bejarano [8], the EFIT-V system models global facial characteristics and allows the witness to produce plausible, near photo-realistic facial composites. Relted work includes Caldwell and Johnston [2] and Hancock et al. [5].

EFIT-V combines a compact model of the human face with an interactive evolutionary algorithm to search for a target face. In Section 2, we will discuss the production of a generative model of facial appearance that provides the genotypes. In Section 3, an evolutionary algorithm is described that balances speed and usability. Section 4 describes significant refinements to the system which have proved to be important in practical applications. Our conclusions on the efficiency

Figure 1: Example composites generated using EFIT-V

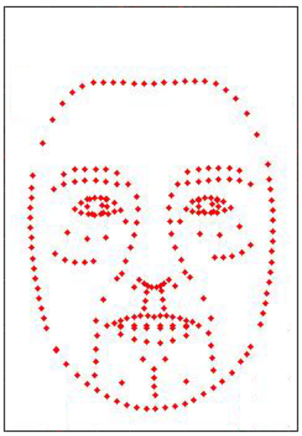


Figure 2: The landmark points used in EFIT-V

of the system, future developments and current usage are presented in the summary. Typical examples of composite images generated by EFIT-V are presented in Figure 1.

2. GENERATIVE MODEL OF FACIAL APPEARANCE

Appearance models are statistical learning methods which enable a compact representation of the human face to be generated [7].

The shape of a face is described by a set of coordinates placed around the main features (Figure 2). Each feature is thus represented by a vector of landmark coordinates, $\mathbf{x}$. By applying principal components analysis (PCA) to the data set $\{\mathbf{x}\}$ the dimensionality of the data is reduced [6]. Any example can then be approximated using:

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

where $\bar{\mathbf{x}}$ is the mean shape, $\mathbf{P_s}$ is the set of orthogonal modes of variation and $\mathbf{b_s}$ is a vector of shape parameters.

Each example face image is then warped to the mean face shape and the intensity value of each pixel is sampled. This is represented by a column vector $\mathbf{g}$ containing the red, green and blue values. PCA is applied to this data to obtain a linear model similar to that for the shape:

$$\mathbf{g} = \overline{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g \tag{2}$$

where $\bar{\mathbf{g}}$ is the mean texture, $\mathbf{P_g}$ is the set of orthogonal modes of variation and $\mathbf{b_g}$ are the texture parameters.

The shape and texture of any example can thus be summarized by the vectors $\mathbf{b_s}$ and $\mathbf{b_g}$. Correlations exist between the shape and texture of an example so it is possible to apply a further PCA to the data yielding a simultaneous representation of shape-texture:

$$\mathbf{b} = \begin{pmatrix} \mathbf{b}_s \\ \mathbf{b}_g \end{pmatrix} = \begin{pmatrix} \mathbf{P}_s^T(\mathbf{x} - \overline{\mathbf{x}}) \\ \mathbf{P}_g^T(\mathbf{g} - \overline{\mathbf{g}}) \end{pmatrix} \tag{3}$$

We apply a further PCA on these vectors, giving a reduced model:

$$\mathbf{b} = \mathbf{Q}\mathbf{c}, \ \mathbf{Q} = \begin{pmatrix} \mathbf{Q}_s \\ \mathbf{Q}_g \end{pmatrix} \tag{4}$$

Where the columns of $\mathbf{Q}$ are eigenvectors and $\mathbf{c}$ is a vector of appearance model parameters controlling both shape and texture. The linear nature of the model allows the expression of both shape and texture directly as a function of $\mathbf{c}$:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{P}_s \mathbf{Q}_s \mathbf{c} \tag{5}$$

$$\mathbf{g} = \overline{\mathbf{g}} + \mathbf{P}_g \mathbf{Q}_g \mathbf{c} \tag{6}$$

An example image can be synthesized for a given $\mathbf{c}$ by generating the 'shape-free' texture from the vector $\mathbf{g}$ and warping it using the control points described by $\mathbf{x}$. To generate plausible faces, the probability density function of appearance model parameters must be estimated. This can be modelled using a standard multivariate normal probability density function:

$$N(\mathbf{c}; 0, \Lambda) = (2\pi)^{-\frac{n}{2}} \, |\Lambda|^{-\frac{1}{2}} \, e^{-\frac{1}{2}(\mathbf{c}^T \Lambda^{-1} \mathbf{c})} \tag{7}$$

Where Λ is the covariance matrix of appearance model parameters. Truncating the appearance model vector, $\mathbf{c}$, allows for a very concise representation of a face, typically 50-75 parameters.

2.1 Appearance Model Used in EFIT-V

The model used in EFIT-V was built from 822 images comprising an assortment of male and female faces and a broad range of ethnicities and ages. Each image was labelled with over 150 landmark points to describe face shape, Figure 2. Polynomial curves were fitted to the facial features using manually placed control points. Landmarks were defined by automatic sampling of the polynomial curves at equidistant intervals. To allow prior demographic knowledge to be incorporated in the search procedure, faces that exhibit characteristics associated with the chosen ethnicity and age can be generated using an appropriate sub-sample model. Accordingly, probability density functions for the sub-sample model were defined as:

$$N(\mathbf{c}; \mu_c, \Sigma_c) = (2\pi)^{-\frac{n}{2}} |\Sigma_c|^{-\frac{1}{2}} e^{-\frac{1}{2}((\mathbf{c}-\mu_0)^T \Sigma_c^{-1} (\mathbf{c}-\mu_0))} \quad (8)$$

Where μ_c is the mean of the sub-sample and Σ_c is the covariance matrix of appearance model parameters comprising of a sub-sample of the set, $\{\mathbf{c}\}$. Figure 3 shows some random faces created from different demographic groups.

3. INTERACTIVE EVOLUTIONARY FACE SEARCH

The appearance model described in Section 2 provides a powerful method of generating plausible facial images by selecting the appropriate vector $\mathbf{c}$. In EFIT-V, $\mathbf{c}$ is a 75 element vector of real numbers which comprises the genotype in our problem. This section describes a method that allows witnesses to determine the optimal vector of appearance model parameters that can be used to construct a suitable likeness to the target face.

We define 'face space' as the vector space spanned by the columns of $\mathbf{Q}$. The method of searching the 'face space' for the fittest vector of appearance model parameters has to satisfy both practical and technical criteria. The process needs to be both easy for a witness to use, and also allow for a fast convergence to the target.

The genetic algorithm designed for the task has been termed a select, multiply, mutate algorithm, SMM [3]. A small population is presented to the user who is asked to select the fittest individual from the selection - the individual that most closely resembles the target. This individual (named the stallion) is cloned a number of times, each being mutated slightly. The mutated clones, as well as the stallion from the previous generation, are then placed in the new population. The process is repeated until the user is satisfied they cannot generate a better likeness to the suspect. This process is detailed in Figure 4. The population size chosen for the algorithm is important. A large population size allows the algorithm to search a larger area of 'face space' at each generation and therefore achieve faster convergence, however this may be overwhelming for a witness. A smaller population is more accessible, but leads to slower convergence. A population size of 9 was chosen as it allows a neat 3×3 grid layout and was found to be a good compromise between convergence speed and usability.

In the SMM algorithm, the only genetic operator used is mutation. The mutation rate determines the amount of variation in each new population. At the start of the search process it is desirable to have a larger mutation rate, to allow the witness to explore as much of the search space as possible. As the witness converges on a target, the mutation rate should be lower to allow more fine grained control. The method of mutation rate control in the algorithm follows the exponential decay:

$$p(t) = 0.100 + 0.417t^{-0.558} \quad (9)$$

1. An initial population of faces is presented to the witness via a graphical user interface
2. The witness *selects* the phenotype that exhibits the best likeness to the suspect - the *fittest* face, labelled the *stallion*
3. The genotype of the stallion is cloned (*multiplied*) eight times
4. Each of the eight genotypes is *mutated* by making random changes to the genetic material (appearance model parameters)
5. The eight mutated genotypes and the stallion are then placed into an array
6. Phenotypes are generated and displayed to the user for rating
7. Steps 1 to 6 are repeated until the witness is satisfied they could not generate a better likeness to the suspect

Figure 4: Structure of the Evolutioanry Algorithm used in EFIT-V

Where p is the probability that a parameter will mutate and t is the generation number. As the number of generations increases, the amount of variation exhibited in the population decreases as the chance of mutation decreases. The mutation rate is the probability that a locus is mutated. When a mutation occurs the paramter value is resampled from the distribution in Equation (8) between ± 3 standard deviations.

The rate of decay was determined by testing the system with simulated ideal and non-ideal witnesses. An ideal witness always selects the individual closest to the target, the fittest at each generation. A non-ideal witness tests the robustness of the system by attempting to model the behaviour of a human user who does not always select the fittest individual at each generation.

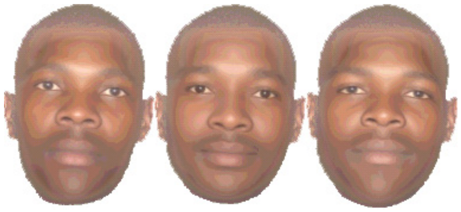
Figure 5 shows the steps taken to generate a new population.

3.1 Evolutionary Rejection Strategy

To improve the rate at which a user can converge on the target without making the process too cognitively demanding it is also possible for the user to reject individuals at each generation. By creating 'no-go' zones in the search space around the rejected individuals, the search will not return to these areas.

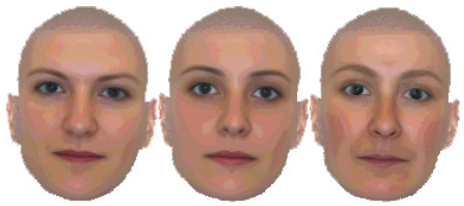
Alongside the sample space, a decision space $D(\mathbf{x})$, initialised to $D(\mathbf{x}) = 1$, is formed that acts as the probability of a new individual being added to the population. As individuals are rejected by the witness the decision space is modified through multiplication of a weight function, $W(\mathbf{x})$. For the $k + 1^{th}$ rejected individual we have:

$$D_{k+1}(\mathbf{x}) = W_k(\mathbf{x}) D_k(\mathbf{x}) \quad (10)$$



(a) Random sample of faces synthesised using the sub-sample of black males aged 40-45.

(b) Random sample of faces synthesised using the sub-sample of white males aged 50-55.



(c) Random sample of faces synthesised using the sub-sample of white females aged 15-20.

Figure 3: Random faces synthesised from different demographics

Where:

$$W_k(\mathbf{x}) = 1 - \mathrm{e}^{(-\alpha(\mathbf{x}-\mathbf{x}_k)^2)} \quad (11)$$

And $\mathbf{x}_k$ is the k^{th} rejected individual.

This sets $D(\mathbf{x}) = 0$ when $\mathbf{x} = \mathbf{x}_k$ and makes the function small in the region where $\mathbf{x}$ approaches $\mathbf{x}_k$. α is an adjustable parameter that effectively controls the zone of influence of a rejected individual.

For any value of $\mathbf{x}$ sampled we then randomly sample from the uniform distribution $P(y) = 1(0 \leq y < 1)$. If $y \leq D(\mathbf{x})$ the individual is inserted into the new population, otherwise we disallow the individual and we resample for $\mathbf{x}$.

Experience with the system has shown that a process consisting of just selection and rejection can achieve satisfactory results. However, in many cases users feel that deterministic changes to the facial appearance are also advantageous when creating a facial likeness. This functionality has been included in EFIT-V and is outlined in Section 4.

4. DETERMINISTIC CHANGES TO FACIAL APPEARANCE

The stochastic method described in Section 3 allows a user to search 'face space' for a suitable likeness to the target. To allow the user to have greater control over the composite generation process, functionality has been provided for making deterministic changes to the image, should the user feel they are beneficial. This functionality includes:

- Scaling, translation and rotation of features
- Age transformation
- Blending of two or more faces within a population
- Adding high spatial frequency overlays

(a) A composite generated without overlays

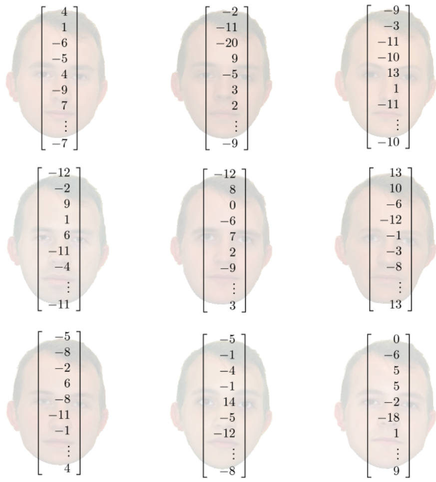
(b) Overlays allow high-frequency details to be added to to composite image

Figure 6: A composite image with and without overlays added

EFIT-V also allows the user to add overlays to a composite to display details that are not well modelled by the facial appearance model. These include high-frequency details such as wrinkles and other lines associated with ageing (see Figure 6). The overlays are currently selected deterministically and independently of the evolutionary process but will be determined by the genetic algorithm at a later date, extending the genotype to include support for adding overlays. The user can use these deterministic tools during the composite generation; to accelerate the convergence to a target face.

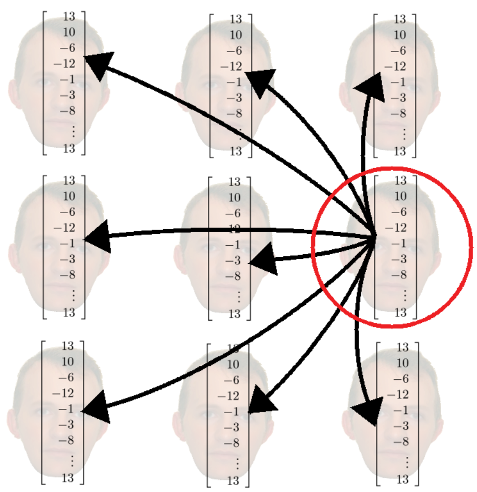
5. SUMMARY

An evolutionary approach to facial composite generation was described and implemented. The use of a global face model depends upon the witness' ability to recognise rather than recall in the composite generation process, a task which is more suited to the cognitive processes involved in human facial processing. A genetic algorithm allows the user to search the 'face space' for a likeness. The system has been

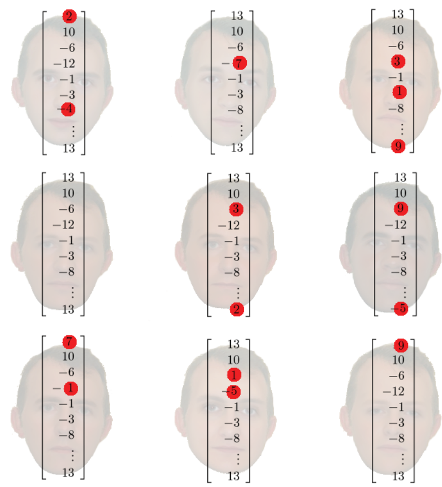


(a) **Initial population:** Random genotypes generated.

(b) **Initial population:** Phenotypes synthesised from the genotypes - fittest phenotypes (as selected by the witness) circled in red.



(c) **Cloning (Multiply):** Genotype corresponding to fittest phenotype cloned nine times.



(d) **Mutation:** Random mutations on eight of the nine clones. The stallion remains unaltered, although its position in the new generation is randomised.

(e) **New Generation:** Based on mutated genetic material from select face.

Figure 5: **Schematic representation of the process which results in generation of phenotype facial images. The user is shown a number of randomly generated faces and selects the one they recognise as looking most like the suspect. A genetic algorithm is then used to breed a new generation of facial composites.**

designed to be sympathetic to the cognitive processes involved in human face recognition. Deterministic changes complement the evolutionary procedure and can be performed during the generation process if the user feels it is necessary.

A system based upon this approach called EFIT-V has been implemented. EFIT-V is currently the subject of a UK Home Office funded trial taking place in two UK police forces. In these trials, EFIT-V successfully provided intelligence, arrest or detection in 30% of cases in which it was used. Moreover, these were cases in which the witness was judged to be incapable of producing a composite using a standard, feature-based, commercial system.

6. REFERENCES

[1] A.W.Young, D. Hellawell, and D. Hay. Configurational information in face perception. *Perception*, 16:747–749, 1987.

[2] C. Caldwell and V. Johnston. Tracking a criminal suspect through face-space with a genetic algorithm. *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 416–421, 1991.

[3] S. Gibson, A. Pallares-Bejarano, M. Maylin, and C. Solomon. The generation of facial composites using and evolutionary algorithm. In K. Sirlantzis, editor, *6th International Conference on Recent Advances in Soft Computing*, pages 249–254. RASC, July 2006.

[4] S. J. Gibson, C. J. Solomon, and A. P. Bejarano. Synthesis of photographic quality facial composites using evolutionary algorithms. *Proceedings of the British Machine Vision Conference 2003*, Vol 1:221–230, 2003.

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

[6] I. Jolliffe. *Principal component analysis.* Springer-Verlag, 1986.

[7] A. Lanitis, C. Taylor, and T. Cootes. Automatic interpretation and coding of face images using flexible models. *Human-computer interaction*, 2:18.

[8] A. Pallares-Bejarano, C. Solomon, and S.J.Gibson. Eigenfit: Building photographic quality facial composites using evolutionary algorithms. *Proceedings of the 3rd IASTED international conference Visualization, imagingand image processing*, page 1, Sept 2003.

[9] 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.