Gibson, Stuart J., Solomon, Christopher J. and Bejarano, Alvaro Pallares (2003) ***Synthesis of Photographic Quality Facial Composites using Evolutionary Algorithms.*** **In: BMVC.**

Synthesis of Photographic Quality Facial Composites using Evolutionary Algorithms

Stuart Gibson, Alvaro Pallares Bejarano, Christopher Solomon
School of Physical Sciences
University of Kent
Kent, CT2 7NR, UK
{sjg4,ap55,cjs5}@kent.ac.uk

Abstract

A facial composite system is described for use in criminal investigations which has distinct advantages over current methods. Unlike traditional feature based methods, our approach uses both local and global facial models, allowing a witness to evolve plausible, photo-realistic face images in an intuitive way. The basic method combines random sampling from a facial appearance model (AM) with an evolutionary algorithm (EA) to drive the search procedure to convergence. Three variants of the evolutionary algorithm have been explored and their performance measured using a computer simulation of a human witness (virtual witness). Further system functionality, provided by local appearance models and transformations of the appearance space which respectively allow both local features and semantic facial attributes to be manipulated, is presented. Preliminary examples of composites generated with our system are presented which demonstrate the potential superiority of the evolutionary approach to composite generation.

1 Introduction

Facial composites are widely used in criminal investigations as a means to generate a likeness to a suspected perpetrator of a crime. Current commercial systems for producing composites such as EFit [15] and ProFit [1] rely on the ability of a witness to recall individual facial features, select the best choice from a large sample of examples and then place them in the appropriate spatial configuration. There are two major drawbacks to this approach. Firstly, many authors working in the field of psychology as early as the late 70s demonstrated the shortcomings of recall as a means of identification [10, 13, 9, 3] and it has been suggested that the requirement for the witness to recall (as distinct from recognise) the face is the weakest link in the composite process [18]. Secondly, a considerable body of evidence now suggests that the task of face recognition and synthesis does not lend itself to simple decomposition into features and is partly, a global process [6, 11, 2] relying as much on the inherent spatial/textural relations between all the features in the face. In 1987, Sirovich and Kirby [16] demonstrated that a principal components analysis (PCA) on a suitably normalized set of faces produced a highly efficient representation of a human face as a linear superposition of global principal components or ”eigenfaces”.

This paper instigated a significant amount of research such that PCA is now a standard technique in face recognition and both 2-D and 3-D face modelling research [5]. Much less explored is the possibility of using PCA in facial synthesis/composite production. Brunelli and Mich [7] employed ”feature” principal components in their composite generator but the problems associated with eyewitness recall were inherently neglected. Hancock [12] describes a facial composite system based on recognition, which circumvents the recall problem. In this work, global facial characteristics of shape and texture are modelled independently using PCA. A more elegant solution is to perform a further PCA, which combines both shape and texture models into a single unified appearance model [8]. A single set of normally distributed appearance parameters then enables the generation of plausible faces. In this paper, we propose a novel approach to the synthesis of facial composites which combines two key components - I) an appearance model of the human face providing both a highly compact representation of any individual face and a tractable statistical model of human facial variation II) A steady-state, evolutionary algorithm which drives the process of composite production to convergence (i.e. best match to target face).

In the next section, we briefly describe the theoretical basis and construction of the facial appearance model. Evolutionary search strategies for achieving convergence on the target face are then discussed in section 3. In section 4, we briefly describe the basis of enhancements to our global evolutionary method in which local features and semantic attributes of a facial composite can be altered without undermining the global nature of the evolutionary search. Section 5 discusses the basis of a computer-generated ”ideal witness” designed to optimise the performance of the EA. Finally, we give a summary of the results achieved in preliminary human trials and discuss current developments.

2 Statistical Model of Facial Appearance (Search Space)

An appearance model was constructed that captures both shape and texture properties of the human face. Here we give a brief overview of the procedure. For a full mathematical account the reader is referred to [8].The model was formed from a training set of example face images. The set contained equal numbers of male and female faces in the age range 20-70. Each training example was labelled with 150 landmark points to describe the face shape. 79 key landmarks were placed by hand and the remaining 71 points were estimated using a polynomial curve fitting algorithm. After labelling, the shapes were aligned using procrustes analysis. Subsequently, principal component analysis (PCA) was used to generate a model that defines the modes of shape variation. Modes of texture variation were also modelled using PCA. To obtain a true correspondence between pixels, each face image was first warped to the mean face shape prior to PCA using a piece-wise affine linear transformation. The shape model and texture model were combined into a single, unified shape-texture model (or appearance model) using further PCA. The key result of this procedure is that each training example can be expressed in terms of the appearance model using an N element vector of parameters, $\mathbf{c} = [c_1\, c_2\, c_3 \ldots c_N]^T$ each element of which can influence the global facial appearance. The relationship between the appearance parameters, the face shape, $\mathbf{x} = [x_1\, y_1\, x_2\, y_2 \ldots x_M\, y_M]^T$, and face texture, $\mathbf{g}$ is given by,

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \tag{1}$$

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \tag{2}$$

Where $\bar{\mathbf{g}}$ is the mean face texture, $\bar{\mathbf{x}}$ the mean face shape and $\mathbf{Q}_s, \mathbf{Q}_g$ are matrices that describe modes of variation in the training set.

The parameter vectors for all training examples in N dimensional space generated in this way obey a multivariate normal distribution. The generation of plausible faces is ensured by allowing the value of the appearance parameter, c_j to vary within the range $-3\sigma_j < \mu_j < +3\sigma_j$ (where σ_j is the standard deviation of c_j and μ_j the mean of c_j over the training set.) For practical applications the number of dimensions of this space can be reduced by truncating the parameter vectors to t elements such that $t < N$ and $\mathbf{c} = [c_1\, c_2\, c_3\, \ldots c_t]^T$. This gives a compact search space within which, in principle, any given face can be represented as a unique vector of parameters. For the generation of facial composites the dimensionality of the search space should be low to minimize the duration of the search whilst retaining enough variance to construct a good likeness to any given face. New examples of faces can be generated by randomly sampling for the vector of appearance parameters from the multivariate, normal distribution. As can be seen from Figure 1, these faces are completely plausible in appearance.

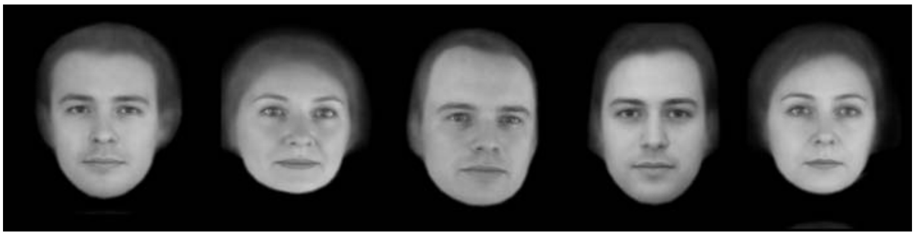

Figure 1: New examples of faces generated from the appearance model

3 Evolutionary Search Methods

Producing a composite face can be regarded as a search for a target face within a multidimensional 'face-space' where the Cartesian axes of the space correspond to the appearance parameters. It is crucial to recognize that the optimum search strategy for this task (in its practical implementation) must be an algorithm that is *a suitable compromise between human usability and speed of convergence* (i.e. the required number of faces seen and rated by the user before a satisfactory composite is achieved). To some degree, these are conflicting goals. Thus, a particular interactive search strategy, which can be demonstrated to converge in fewer iterations based on computer simulated studies, will be unusable if the input/decision required from the human user at each iteration is too difficult or time-consuming (e.g. the numerical rating of a large number of faces for their similarity to the target face). Alternatively, if the information input/decision from the human user at each iteration is restricted to a very simple task (selection of the best face from only two examples), we may expect that convergence will be correspondingly longer. Accordingly, we have explored and successfully developed three evolutionary approaches to conducting the search, each of which requires a different input from the user. The key elements in each approach are described below

Scale Rating (SR) algorithm - In this approach we employ an elitist genetic algorithm, applying the operations of selection, crossover and mutation between an elite member of the population (*the stallion*) and another member. At each iteration the offspring produced are rated on a simple numerical scale of 0-10 by the user for their perceived similarity to the target face. To encourage consistency and avoid fitness scaling problems which might induce premature convergence to an incorrect solution, the current stallion (i.e. best likeness generated so far) and its assigned score is made visible at all times to the user. The process of assigning numerical ratings to the faces displayed and replacing the stallion as appropriate is continued until the user considers the best possible reconstruction has been achieved.

SMM (Select-Multiply-Mutate) algorithm - In a similar fashion to the SR algorithm described above, this algorithm also employs an elitist strategy. However, in this case, the user is required at each iteration to select the best likeness from a small group (5 is typical) which are simultaneously displayed (*select*), this likeness is then cloned a number of times (*multiply*) and all but one is randomly mutated to produce a new generation (*mutate*). The best likeness is then selected from this new generation and so on. The basic *SMM* algorithm is depicted in the flow chart in Figure 2

Follow-The-Leader (FTL) algorithm - This strategy is the easiest algorithm for the human operator. At each step of the iterative process, a new face is displayed alongside the current best likeness (*the stallion*) and the user is simply asked to make a choice between the two faces. The new face presented at each iteration is the result of breeding the stallion with a new member. In this variation of the EA the recent evolutionary history is a factor in determining future offspring. For instance the recent evolutionary history may suggest that the process is following a well defined direction (as opposed to a random totally random path) through the search space. If so a preference will be made for this direction at subsequent iterations, accelerating the search process along a more efficient path through the appearance space and reducing the number of iterations required for convergence.

The performance of these algorithms in extensive trials employing a simulated, virtual witness is discussed in section 5.

4 Enhancing Functionality - Interactive Face Modification Techniques

4.1 Local Feature Manipulation

There is a substantial body of psychological literature to suggest that the process of human face recognition involves both local (isolated facial features), and global (facial features in a configurational context) information [17, 4]. Accordingly, a truly effective facial

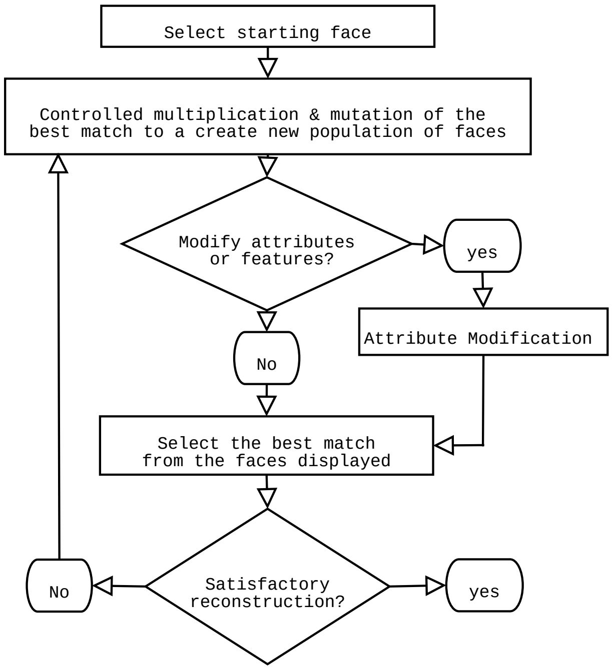

Figure 2: *SMM* Algorithm flowchart

composite system should be sufficiently flexible to operate in both the global and local modes. Global facial change is naturally effected by our core, evolutionary approach in which the modification of any of the appearance model parameters produces a change in the entire facial appearance. To achieve a similar effect on isolated facial features, we have built analogous feature appearance models in which the shape and texture of the features can be represented and manipulated from a compact parametric form. In this way, we make provision for the user to manipulate isolated facial features (e.g. change the width of the nose, or fullness of the lips) but then seamlessly resume the global evolutionary procedure which can be accelerated by such interventions. Figure 3 shows the effect of using a local feature appearance model of the nose in conjunction with a global model to ensure globally plausible face shapes.

Figure 3: Image sequence showing a change in nasal width

4.2 Facial Attribute Manipulation

There are many semantic attributes that are commonly ascribed to a human face. Simple examples are *attractive*, *kind*, *masculine*, *honest* and so on. Some attributes have a more objective nature (e.g. a masculine/feminine face) in the sense that a sample population will exhibit much closer agreement in their perception whereas others are more subjective and generate more disagreement (e.g. an honest face). There is evidence to suggest that the human ability to define/recognise subjective facial attributes may be based on hormonal and other underlying biological factors. It would therefore seem plausible that a powerful, complementary way to navigate our appearance search space is along directions that are known to be associated with attributes, allowing facial synthesis to proceed along perceptually more salient directions. These directions can be defined by weighting the appearance model parameters of a sample of faces according to the degree to which they are considered to possess the given attribute. Our preliminary implementation of this method is now briefly described.

A survey was first conducted in which 40 subjects were asked to assign a score, $\mathbf{s}_k(i)$, between one and five to each of the training images based on the degree to which a given face exhibited a particular attribute (thus a highly masculine face would be given a score of $\mathbf{s}_k(i) = 5$). For each subject participant a new vector $\mathbf{d}_k$ can be defined as a weighted combination of the appearance model parameters.

$$\mathbf{d}_k = \mathbf{A}\mathbf{s}_k \tag{3}$$

Where $\mathbf{A}$ is matrix in which each column is a vector of appearance model parameters corresponding to one of N training examples

$$\mathbf{A} = [\mathbf{c}_1\, \mathbf{c}_2\, \mathbf{c}_3 \ldots \mathbf{c}_N] \tag{4}$$

The vector $\mathbf{d}_k$ defines a direction in appearance space that the k^{th} subject perceives to be most closely associated with the attribute of interest. Averaging over all such vectors for all 40 scorers gives the consensus as to the direction, $\bar{\mathbf{d}}$, which best defines the attribute. Changing an attribute of a given face is simply achieved by adding a small quantity of the vector $\bar{\mathbf{d}}$ to the vector of appearance parameters $\mathbf{c}$ according to Equation 5. Changes in perceived gender characteristics and race using this method are shown in Figure 4 and Figure 5

Figure 4: Moving in appearance space to modify gender

$$\acute{\mathbf{c}} = \mathbf{c} + \Delta\bar{\mathbf{d}} \tag{5}$$

Figure 5: Moving in appearance space to modify racial appearance

5 System Optimization and Testing

5.1 Simulated Witness and Witness Trials

The behaviour of evolutionary algorithms can be strongly affected by a number of parameters such as probability of crossover and mutation, selection method and genotype length. Ultimately, it is anticipated that the system described in this paper will be subject to psychological evaluation and testing with human subjects. As such evaluation is time-consuming and costly, we developed a virtual witness program in which the computer simulated the role of the human witness, responding to the stimulus according to a distance metric between it and the target face. The aim of this was to study and optimize the convergence properties of our algorithms in order that human trials can then be conducted effectively using a properly tuned system. The *ideal* virtual witness assigns a fitness score to a facial phenotype according to Equation 6 for the SMM and FTL algorithms and to Equation 7 for the SR algorithm, where ρ is the correlation coefficient and ε is the euclidean distance between the vector of appearance parameters of the given face and the vector of the actual (known) target.

$$f = \varepsilon \times \frac{\left(101 - \rho^2\right)}{100} \tag{6}$$

$$f = (1+\rho)^2 \; ; \; f_N = \frac{(1+\rho)^2}{4} \tag{7}$$

A non-ideal virtual witness has also been developed that aims to mimic the imperfect scoring of a typical human witness. This has been achieved by introducing a random error to the fitness function as shown in Equation 8, where α is sampled from a zero mean normal distribution with standard deviation σ (where σ lies in the range 0.2 to 0.5). The non-ideal virtual witness si intended to asses the robustness of the evolutionary approach in human trials and the number of iterations that will be required to obtain a good likeness to a target face.

$$f_N = \frac{(1+\rho)^2}{4} \times (1+\alpha) \tag{8}$$

Extensive simulations were carried out to optimize the free variables of the algorithms. The number of iterations to obtain a quasi-perfect composite was reduced dramatically, from typical values of several thousand to average values of 150 iterations for the SMM algorithm and 350 iterations for the FTL algorithm. An example of typical composite generated in 160 iterations using the SMM algorithm is shown in Figure 6

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

Figure 6: SMM ideal witness search process: a) Starting face, b)and c) are intermediate points in the evolutionary process, d) final generated composite, e) actual target face

6 Human Trials and Discussion

We have demonstrated that it is possible to produce a good likeness to a target face using an evolutionary search algorithm to navigate through an appearance space. The concept of virtual witness' was introduced as a tool for testing and optimizing evolutionary algorithms. Preliminary human trials suggest that convergence is possible in a viable number of iterations. Two such examples generated by a human operator are given in Figure 7 and Figure 8.

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

Figure 7: SMM human trial: a) Starting face, b)and c) are intermediate points in the evolutionary process, d) final generated composite after 27 iterations (162 faces viewed), e) actual target face

In Figure 7 the target face remained visible to the human operator while generating the composite. An acceptable convergence was obtained in approximately 20 minutes which is faster than current systems. Figure 8 shows a likeness generated from memory of the current British Prime Minister, Tony Blair. In the final facial composite shown in Figure 8e, hair has been added from a database of hair styles since hair is not modelled well by the appearance model. Current developments include the incorporation of a database of facial appendages such as facial hair, scars and glasses as well as a comprehensive set of images for building appearance models based on gender and ethnic origin.

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

Figure 8: SMM human trial for famous target face: a) Starting face, b)and c) are intermediate points in the evolutionary process, d) final generated composite after 23 iterations (138 faces viewed), e) addition of hair to facial composite

The approach that we have taken to facial composite generation has potential applications outside criminal investigations in fields such as plastic surgery [14] and entertainment.

Acknowledgment

We gratefully acknowledge the support provided for this work from the EPSRC award GR/S06738.

References

[1] ABM. Profit. http://www.abm-uk.com/.

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

[3] A.D. Baddeley. *The psychology of memory*. Harper and Row, 1976.

[4] J.C. Bartlett and J. Searcy. Inversion and configuration of faces. *Cognitive Psychology*, 25:281–316, 1993.

[5] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH'99 Conference Proceedings*, pages 187–194, 1999.

[6] V. Bruce, P. Hancock, and A. M. Burton. *Human Face Perception and Identification, in Face Recognition - From Theory to Applications*, pages 51–72. Springer Verlag, 1998. ISBN3-540-64410-5.

[7] R. Brunelli and O. Mich. Spotit! an interactive identikit system. Technical report, IRST Technical Report, 1994.

[8] T. Cootes and C. Taylor. Statistical models of appearance for medical image analysis and computer vision. In *SPIE Medical Imaging*, pages 236–248, 2001.

[9] G.M. Davies and D. Christie. Face recall: An examination of some factors limiting composite production accuracy. *Journal of Applied Psychology*, 67:103–109, 1982.

[10] H.D. Ellis, G.M. Davies, and J.W. Shepherd. A critical examination of the photofit system for recalling faces. *Ergonomics*, 21:297–307, 1978.

[11] A.J. O'Toole et al. Structural aspects of face recognition and the other-race effect. *Memory and Cognition*, 22:208–224, 1994.

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

[13] D. Hines, L. Jordan-Brown, and K.R. Juzwin. Hemispheric visual processing in face recognition. brain and cognition. *Brain Cognition*, 6:91–100, 1987.

[14] V.S. Johnston, C.J. Solomon, S.J. Gibson, and A. Pallares-Bejarano. Human facial beauty: Current theories and methodologies. *AMA Archives of Plastic Surgery, Manuscript QSP20000*, 2003. in press.

[15] Aspley Ltd. E-fit. http://www.efit.co.uk/.

[16] L. Sirovich and M. Kirby. Low dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America*, 4(3):519–524, 1987.

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

[18] 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.