Implementing Holistic Dimensions for a Facial Composite System

Charlie D. Frowd
Psychology Department, Stirling University, Stirling, UK
Email: cdf1@stir.ac.uk

Vicki Bruce
College of Humanities and Social Science, Edinburgh University, Edinburgh, UK
Email: Head.CHSS@ed.ac.uk

Alex H. McIntyre, David Ross, Stephen Fields, Yvonne Plenderleith, and Peter J.B. Hancock
Psychology Department, Stirling University, Stirling, UK
Email: ahm5@stir.ac.uk, david-a-ross@hotmail.co.uk, maverick_saf@yahoo.co.uk, yvonneplenderleith@hotmail.com, pjbh1@stir.ac.uk

***Abstract*—Facial composites are pictures of human faces. These are normally constructed by victims and witnesses of crime who describe a suspect's face and then select individual facial features. Unfortunately, research has shown that composites constructed in this way are not often recognised. In contrast, we are quite good at recognizing complete faces, even if the face is unfamiliar and only seen briefly. This more natural way of processing faces is at the heart of a new composite system called EvoFIT. With this computer program, witnesses are presented with sets of complete faces for selection and a composite is 'evolved' over time. The current work augments EvoFIT by developing a set of psychologically useful scales – such as facial weight, masculinity, and age – that allow EvoFIT faces to be manipulated. These holistic dimensions were implemented by increasing the size and variability of the underlying face model and by obtaining perceptual ratings so that the space could be suitably vectorised. The result of three evaluations suggested that the new dimensions were operating appropriately.**

***Index Terms*—facial composite, holistic, witness, crime, EvoFIT**

I. Introduction

Witnesses and victims of serious crime have an important role in the apprehension of criminals. It is normal for such eyewitnesses to describe the events surrounding a crime and the individuals involved. While the description of a face is of little value on its own for identifying a person, for example [1][2], witnesses normally construct a visual likeness of a face, known as a facial composite, and this representation is a more successful probe to identification [3]. These composite pictures are traditionally constructed using specialised computer software, specifically E-FIT and PRO-fit in the UK, or with the assistance of a sketch artist. Both of these methods have been the subject of laboratory research [4][5][6][7][8][9][10][11][12][13][14][15].

It has been known for a long time that recalling an unknown person's face can be difficult, although it is important for locating facial features – eyes, nose, mouth, etc. – within a composite system. The selection of facial features is not an easy task either [16]; it is perhaps not very surprising then that modern composites are poorly identified, especially after a person has waited a couple of days prior to composite construction [10][12][14][15][17], a timescale typically found in policework.

The general difficulty with recalling information is in contrast with our relatively good ability to recognise a face seen previously, or one similar, even if the face was only observed for a short time [18]. Our superior ability to process faces 'holistically' is at the heart of a new composite system called EvoFIT under development at the University of Stirling [10][12][14][19][20][21]. This computer program presents witnesses with a range of faces initially containing 'random' characteristics. Witnesses then identify the most similar faces to an assailant and the software 'breeds' these choices together to produce a new set. This process of selection and breeding continues until an acceptable likeness is achieved: a composite is thus created by 'evolution'. We know of two other similar systems under development, EigenFIT in the UK [22] and ID in South Africa [23].

EvoFIT therefore does not rely on recall (describing a face), but recognition (selecting similar-looking faces). The system already works better than other UK composite systems [10][12] when tested with 'mock' witnesses working from the memory of a person seen several days previously. However, the holistic nature of the model, which is built from faces in their entirety (except hair), has yet to be fully exploited. Using such a model, it is possible to make global changes to a face, for example by making it appear older, more masculine, or more threatening. Holistic operations of this kind are requested by witnesses, but are very difficult to achieve with the current systems.

The present paper describes the work we have carried out in order to implement and evaluate a range of holistic operations. Some of this work was originally published in Frowd et al. [24], but is presented here in expanded form, and also includes the development and evaluation of an ergonomic holistic tool suitable for use with EvoFIT witnesses. Further work is planned which uses these operations as part of a formal evaluation, that is, one where participant-witnesses construct EvoFITs from a two day memory of a target face and where comparison is made with another composite system. In addition, while the current work employed celebrity faces to (conveniently) evaluate the holistic dimensions, further research would use non-celebrity faces.

II. Current EvoFIT

The underlying mechanism that generates the face images in EvoFIT is currently built from 72 photographs of young male faces using Principal Components Analysis (PCA). PCA is a statistical technique that extracts the main axes of variation, the eigenvectors, or in this case, the eigenfaces, in a set of data. This technique works well for faces, for example [25][26], is often used for image compression [25], but is of particular value here as it provides a set of references faces (eigenfaces) that can be combined in variable amounts to produce a novel face, a key component of a face evolution system.

To build this novel face generator, approximately 250 coordinate landmarks were first located on the edges of the facial features (eyes, nose, mouth, etc) in each of the 72 faces, and these were then morphed to an average face shape [27]. PCA was then conducted separately on the resultant ‘shape free’ faces (pixels) and on the shape coordinates, to give facial texture and shape models respectively. The resultant models are essentially holistic in nature [19][28]. For example, one eigenvector may encode the sex of the face, making coordinated changes across the whole image. A novel face can be generated by adding random amounts of each set of eigen-components to the average image.

The user is shown approximately 70 such faces and usually selects 6 that most resemble the target. An underlying Evolutionary Algorithm (EA) then generates new faces by randomly choosing pairs of selected faces and randomly combining, with a small amount of mutation, the underlying 72 shape and texture parameters (i.e. all available eigen-dimensions are used to produce the faces). Repeating the selection and breeding process allows the pool of faces to evolve towards an appearance selected by the witness.

It turns out that faces are represented well with PCA, but this is not the case for hair, since a blend of hairstyles is seldom meaningful. Therefore, this feature is normally taken from a current composite system, PRO-fit, and selected at the start of evolution. Since it was sometimes observed that a face was generated with an appropriate shape but a poor texture, and vice versa, the selection procedure was refined to first allow the selection of facial shape, then texture. Note that this method deliberately differs from the appearance model approach [29] where the shape and texture components are inherently combined. Witnesses subsequently select the optimum combination of shape and texture, to identify a ‘best-face’, which is given double the number of breeding opportunities in the EA.

Our experience is that witnesses are often able to suggest specific alterations to the best face, such as narrowing the face shape or moving the eyes apart, and such operations were permitted by a small software tool called the Feature Shift. This tool allows such shape changes to occur by simultaneously varying a number of underlying PCA parameters so that the faces remain within the model space and can therefore be evolved further.

EvoFIT has been evaluated using standardised procedures for evaluating composite systems, consistent as far as possible with current police procedures, for example [12]. Typically, witnesses are shown the picture of a target that is unknown to them. Two days later they undergo a Cognitive Interview, designed to help them recall as much as possible about the face [30], and then construct a composite. These composites are shown to other people who do know the targets to see if they are recognised. This procedure, unknown at construction and known at recognition, is important, as it mimics police use. In recent evaluations, EvoFITs were named significantly more often than composites from current commercial systems [10][12][14][21].

III. The Addition of Holistic Dimensions

To add holistic dimensions to EvoFIT, a new face model was first created, similar to the existing one but with more items and variability. The old model contained 72 faces, mostly in the age range from 20 to 40; the new set contains 200, ranging from mid teens to early seventies. The new faces were then rated along on a number of dimensions, such as masculinity and health. This allows computation of, for example, an average high masculinity face and an average low masculinity one. The difference between these averages defines a vector through the parameter space. By thus altering the relevant parameters, we can alter a given face to make it more or less masculine, healthy etc., as desired. We note that this approach, of rating model faces for the purpose of vectorising a face space, has been proposed for another type of holistic composite system [22].

The process of dimensionalising face space in this way is described further in the following sections. This is followed by details of an evaluation which tests the effectiveness of the new dimensions. Lastly, we describe and evaluate an ergonomic interface which manipulates faces holistically and can be used as part of composite construction with EvoFIT.

A. A new face model

A larger database than 72 faces was believed necessary to provide sufficient variability for a system with holistic dimensions. To this end, about 250 white male faces without glasses were carefully photographed in a front face pose and a neutral expression. As PCA is very sensitive to changes in ambient lighting, we used a pair of flashlights (positioned at approximately 30 degrees and 2m from the subject; the camera was the same distance away) and a small camera aperture (f-16). These data were collected at the Sensation Science Centre in Dundee and at the University of Stirling. Although sampling was opportunistic, we were able to collect a wide age range.

Two hundred and twenty of these faces were cropped and converted to 8 bit monochrome images at a resolution of 180 pixels (wide) x 240 pixels (high). The procedure described in II above was then used to build a new shape and texture face model. This was initiated by locating key facial landmarks in each face, except that an extra 48 coordinates were used, to allow a better representation of eyebags, jawline, brows, nose and nostrils. Then, as before, a new shape and texture PCA model was built using 200 of the faces (the remaining images were used elsewhere for testing purposes). The distribution of these 200 faces by age is shown in Table I.

TABLE I.
DISTRIBUTION BY AGE FOR THE NEW EVOFIT FACE MODEL

Age	15..20	20..29	30..39	40..49	50..75
Frequency	13	63	51	45	28

B. *Holistic dimensions*

Ratings were collected for each face along the following holistic dimensions, chosen to be those likely to be requested by witnesses: attractiveness, health, honesty, extroversion, threatening, and masculinity (a seventh dimension was included, facial distinctiveness, and data from this was saved for other projects). The raters were adult visitors to the Glasgow Science Centre and each person was tested individually, as featured elsewhere in this paper. The raters were presented sequentially with 44 faces and provided a rating that best described the face along the presented scale. The rating scale used was continuous, but anchored at the end-points with appropriate labels: unattractive … attractive, unhealthy … healthy, dishonest … honest, shy/introverted … outgoing/extroverted, friendly … threatening/hostile, feminine ... masculine, average-looking … unusual/distinctive. This exercise was carried out on a laptop with random sampling of both the faces presented and the associated rating scale. Three hundred and twenty visitors participated to provide a total of eight ratings of each face along each dimension (i.e. 220 faces x 7 scales / 44 ratings = 35 participants / repeat). These dimensions were supplemented by a facial weight scale, representing thin/narrow to wide faces, which were based on ratings of the final 200 faces from six staff and students at Stirling.

The 40 faces with the lowest rating and the 40 faces with the highest rating were averaged for each dimension, as illustrated in Fig. 1, and the corresponding averages were computed in the PCA face space (an average of 40 face coefficients) to provide the reference points for the various holistic vectors. To make a face appear more youthful, for example, the coefficients of a face would be progressed along the aging vector in the direction of the average young face.

Figure 1. Example averages of the holistic dimensions for age (top row), facial weight (middle row), and threatening (bottom row).

IV. EVALUATION

To explore the effectiveness of the new dimensions, three main evaluations were conducted. The first involved systematically manipulating a set of faces in the model along each dimension and verifying the transforms by further ratings. The second, involved constructing a set of composites using the new model, manipulating them to improve the likeness with the holistic tools, and then comparing the quality of the veridical and transformed images. In the third evaluation, an ergonomic interface was designed and evaluated which contained the holistic tools. Note that for these investigations, eight dimensions were considered. These included age, plus seven of the others mentioned above: attractiveness, health, honesty, extroversion, threatening, masculinity and facial weight.

A. Evaluation 1: Perceptual tests

Twelve of the faces used for the model were selected at random and manipulated by a fixed amount in both the positive and negative direction along each dimension. The amount of change was taken as twice the vector length for each dimension, as this produced a sizeable change that was not too extreme: very large changes tended to produce unacceptable distortions. Example transforms can be seen in Fig. 2.

Figure 2. Example holistic transforms. From left to right, top to bottom: reduced age, increased age; reduced health, increased health; and reduced weight, increased weight.

The same rating procedure and scale as II-B was used, this time using volunteer staff and students at Stirling University. Each person provided ratings (1 = low / 10 = high) from four target faces plus manipulations (negative and positive) thereof along the scale that matched the dimension being manipulated – for example, masculinity ratings were collected from faces manipulated along the masculinity dimension. Twenty-four participants each provided 96 ratings (4 faces x 3 combinations x 8 scales) to give a total of eight ratings for each target face at each level of manipulation. The order of image presentation was randomised for each person.

The ratings obtained from the four different faces were combined to give, for each participant, an average rating for each of the three levels of manipulation (negative / veridical / positive) along each dimension (age / attractiveness / health / honesty / extroversion / threatening / masculinity / weight). These were subjected to a repeated-measures Analysis of Variance (ANOVA), which was significant for dimension, $F(7, 161) = 18.9$, $p < 0.001$, and level, $F(2, 46) = 27.7$, $p < 0.001$. However, these factors also interacted, $F(14, 322) = 11.0$, $p < 0.001$, as all positive and negative manipulations gave rise to a significant change in ratings except for positive attractiveness.

The range of average ratings for each scale was generally quite large, spanning for example 3.0 to 5.3 for attractiveness, and 3.0 to 8.2 for facial weight (SD by-items ranged from 6.0 to 8.2). Appropriately, average rating scores indicated that a positive manipulation along a scale always led to an increase in rating for that scale, and similarly that a negative manipulation consistently led to a decrease. Overall, average ratings increased by 34% for positive manipulations and decreased by 26% in the opposite direction.

The data thus suggest that all manipulations except positive attractiveness operated appropriately. We believe that the degree of positive manipulation for attractiveness was too great and only served to go beyond the region of increasing attractiveness. If this is indeed true, then perhaps a smaller positive change might be viewed as being more attractive. We tested this notion by reworking the positive manipulation, to a level half that of the previous setting, and repeating the rating task. This time, eight staff and students at Stirling provided purely attractiveness judgments, for all 36 faces (12 faces x 3 levels of attractiveness). The mean rating was 3.3 for the negative manipulation, 4.8 for veridical faces, and 5.3 for the positive manipulation. The ANOVA was significant, $F(2, 14) = 22.5$, $p < 0.001$, as were both the positive and negative manipulations, $t(7) > 3.3$, $p < 0.02$. Therefore, faces manipulated along the holistic dimensions, including attractiveness, appear to be perceptually sensible.

B. Evaluation 2: Identification

When employed normally with witnesses, we anticipate that the new dimensions would be used either to enhance the best face at the end of each generation, thus accelerating the evolution of a face, or as a final stage in the construction process. For the purposes of the current evaluation, we focused on the latter of these possibilities, and opted to construct a set of composites using the new model and then to manipulate them using a simple prototype interface. The success of the holistic dimensions was measured in this way by comparing composite quality before and after manipulation with the interface.

A set of famous face composites were first constructed using the new EvoFIT face model. To do this, an EvoFIT operator looked at a famous face for 1 minute and evolved a composite. The normal EvoFIT procedure was used, for example [10][21], and involved the operator repeatedly selecting six facial shapes and six facial textures from a set of about 70 faces and running two breeding cycles. For simplicity, the hair, ears and neck for each composite were taken from the 200 references faces used to construct the texture model (rather than from the PRO-fit system). This construction procedure was repeated twice for each of eight well-known celebrities in the UK, with a randomized construction order, to produce a 16 item composite set. The celebrity targets were David Beckham, Stephen Hendry, Tim Henman, Ronan Keating, Ant McPartlin, Michael Owen, Robbie Williams, and Will Young.

The resulting composites were then modified using the holistic dimensions to make them appear visually better. To do this, the operator had available a prototype software tool containing eight Windows sliders, one for each dimension. For each composite, he selected a promising-sounding scale, for example facial weight, set it to give the most identifiable likeness and then moved on to another scale. The tool was designed such that a change made on one dimension was taken as a starting point for the next. Examples are shown in Fig. 3.

To check whether other people also thought the manipulations were visually better, the first composite constructed from each target was presented along with its manipulated counterpart and the target face, and 18 students at Stirling University selected the image they thought best. It was found that 75% of the time, the manipulated image was preferred, and this was significantly more often than the original composites, $X^2 = 16$, $p < 0.001$.

The 16 original and the 16 manipulated composites were given to another group of 34 student volunteers, told that they were of famous faces and asked to provide a name for each one where possible. Each image was presented sequentially in a random order, and the participants provided a name for each where possible. Original composites were correctly named 4.8% of the time, compared with 9.6% for those given a holistic manipulation, a significant increase, $t(33) = 3.3$, $p = 0.002$. Therefore, both perceptually and by naming, the manipulated composites were better, thus suggesting that the holistic tools were operating appropriately.

In the above work, a computer operator enhanced the composites along the holistic dimensions in a prototype software tool. In the next part, we describe a more ergonomic interface that was developed for use by witnesses, and outline a study to evaluate it.

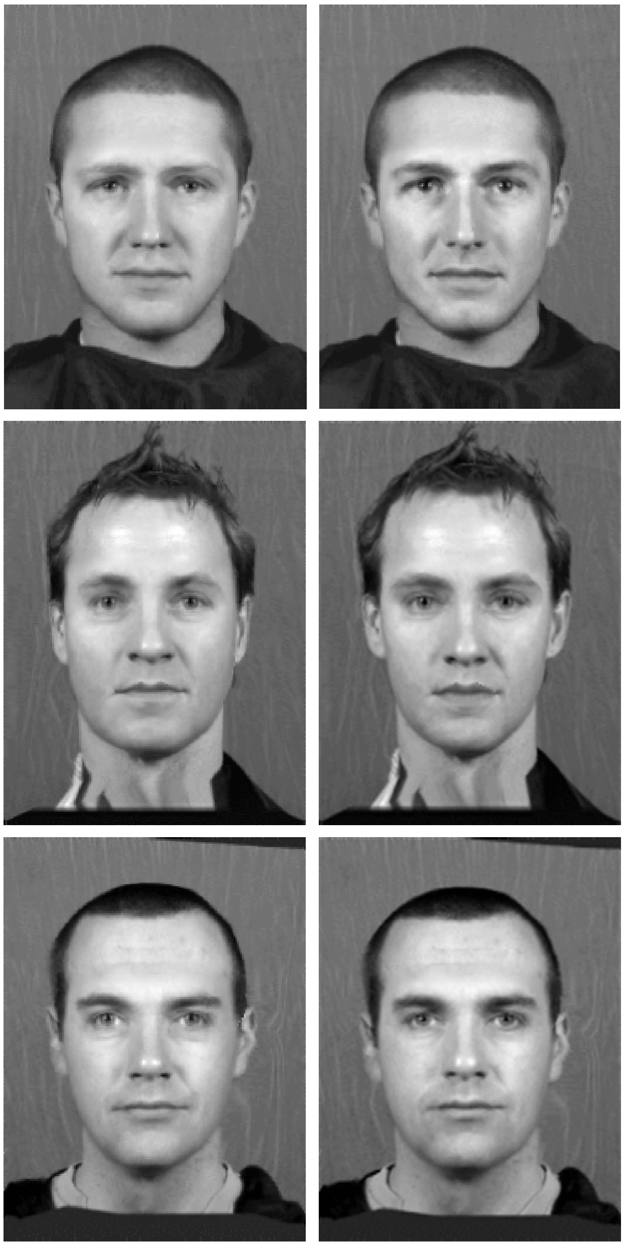

Figure 3. Famous face composites before (left) and after (right) holistic enhancements: top row, the British footballer David Beckham, and after increasing health and attractiveness; middle, singer Will Young, and after decreasing health, extroversion and age; and bottom, pop singer Robbie Williams and after decreasing honesty.

C. *Evaluation 3: A forensically-valuable tool*

The operator in section IV-B above commented that it was generally quite difficult to decide (a) which dimension should be used first, although all dimensions were valuable some of the time, and (b) the value required to produce the best likeness. This feedback was used to design a new interface, called simply the Holistic Tool, that could be used at the end of composite construction with witnesses.

The new interface presented the eight holistic dimensions sequentially and in a fixed order. Age and facial weight dimensions were presented first, arguably two of the most important dimensions for perceiving faces [31], followed by the others: attractiveness, outgoing / extroversion, health, honesty, masculinity and threatening. Each scale could be adjusted by a user using a Windows slider control that spanned a +/-100% change for all dimensions except age and weight, where +/-50% change appeared to be sufficient. Dimensions had a resolution of +/-5 steps. After each slider had been used, the best representation was taken as the starting point for the next dimension, as before. At the end of scale use, the tool presented the original and manipulated image at the same time, to allow changes to be discarded by the user if desired. An option was also made available to use the scales again to further refine the face.

The Holistic Tool was subjected to a formal evaluation. Eight of the 16 composites produced above were used as stimuli, one per celebrity target. A total of 40 participants randomly selected a composite of a familiar target, with equal sampling, and attempted to make the face as identifiable as possible using the presented sliders. The procedure was carried out entirely from the participant's memory of the chosen celebrity. Participants were sampled more widely than before, and were (a) staff and students at Stirling University, (b) staff at a local company (HSBC), and (c) students attending a Psychology Open Day at High Wycombe school.

All participants reported that the tool improved the likeness of their composite face and most of them (67.5%) used it more than once (M = 1.7 cycles).

The overall mean scores by dimension are presented in Table II. While we did not necessarily expect these overall scores to be substantially different from zero, since settings may be positive for one composite, negative for another and thus cancel out over the set, this did turn out to be the case for age (-26.3%), attractiveness (+26.5%) and honesty (+22.0%). These data do suggest that in spite of quite large individual differences in settings, as indicated by the sizeable standard deviations in Table 2, there are certain quite strong preferences in general for enhancing the composites.

The participant data were compared against a rating of zero in a repeated-measures ANOVA. This analysis was found to be significant for holistic dimension, $F(7, 273) = 4.5$, $p < 0.001$, and for comparisons against zero, $F(1, 39) = 4.8$, $p < 0.05$. These factors also interacted with each other, $F(7, 273) = 4.5$, $p < 0.001$, as age, attractiveness and honesty were significantly different from zero, $p < 0.02$, and facial weight and masculinity similarly approached a significant difference, $p < 0.09$. Overall, the data suggest that the EvoFITs evolved in section IV-B could be rendered more identifiable when made to look younger, more attractive and more honest.

TABLE II.
OVERALL PARTICIPANT MEAN SCORES FOR EACH HOLISTIC DIMENSION

Age	Weight	Attractiveness	Outgoing
-26.3*	-8.8†	26.5*	-3.5
(38.4)	(28.1)	(55.4)	(48.3)

Health	Honesty	Masculinity	Threatening
10.0	22.0*	17.0†	0.0
(54.7)	(55.2)	(61.5)	(56.2)

Values are percent change along each dimension and those in brackets are standard deviations. * indicates significance from zero at $p < 0.05$, and † similarly at $p < 0.1$.

As a final stage in this evaluation, we manipulated each composite by the mean setting for that face along each dimension. Examples are presented in Fig. 4. Note that by comparing Fig. 3 and 4, some manipulations appear markedly different to those produced by the operator in IV-B, as illustrated by the David Beckham composites, while others are quite similar, as can be seen for those of Robbie Williams and Will Young.

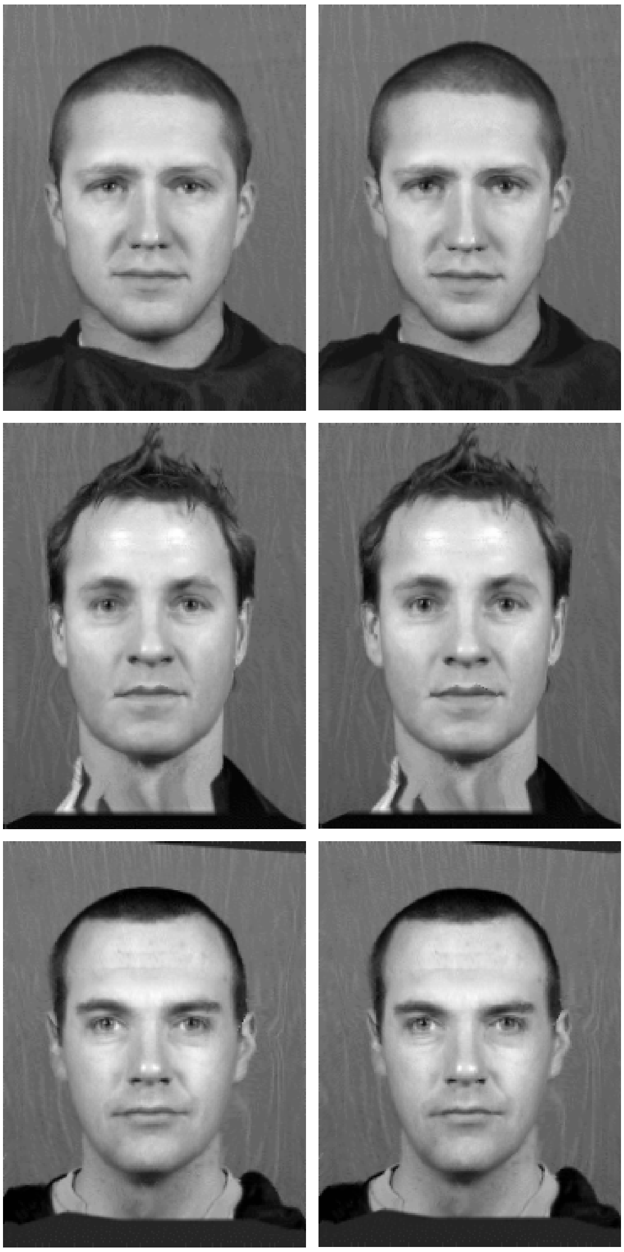

Figure 4. Famous face composites before (left) and after (right) holistic enhancements made using average participant settings. Top, David Beckham; middle, Will Young; and bottom, Robbie Williams.

We next tested whether the manipulated images were considered better representations of the celebrities than the veridical images. In section IV-B, participants were asked to select the best likeness of the original and manipulated image, and this was carried out in the presence of a photograph of a celebrity face. However, the presence of the target photograph may have encouraged participants to compare the individual features of a face, which may not be the most appropriate method for evaluating a set of essentially holistic operations. For this test, participants worked from their memory of the celebrities, and, while a potentially more challenging exercise, it may encourage them to evaluate the faces in a more holistic way. Thus, a further 24 participants, who were drawn mainly from students at Stirling, were given the correct name for each pair of composites and selected the one they thought to be most identifiable. The position of the manipulated image for each pair was randomized, as was the order of presentation of the pairs for each person.

The results indicated that participants considered the manipulated composites of TV presenter Ant McPartlin and snooker player Stephen Hendry were the best representations (37% original, 63% manipulated for both). Across all composites, the manipulated images were selected as best likenesses 9% more often than the original images, a difference that approached significance in a Chi-Square test, $p < 0.1$ (one-tailed).

Finally, we investigated whether the manipulated composites would be recognised better than the veridical images. Two testing books were constructed of 8 composites each, half of which were veridical and half were manipulated composites (order counterbalanced). A final set of 24 volunteers attempted to name a book of composites (with random and equal sampling to the books). We also attempted to elevate naming levels, by asking participants to name the composites a second time after they had been shown the celebrity targets – a so called ‘cued’ naming procedure [9].

It was found that the ‘uncued’ naming level was a little higher for the manipulated composites (M = 7.3% vs. 5.2%), but this increase did not approach significance, $t(23) = 0.6$, $p > 0.1$. However, for the ‘cued’ procedure, correct naming was higher by 15.6% for the manipulated images (M = 54.2% vs. 38.5%), and this difference was significant, $t(23) = 2.7$, $p < 0.02$. Therefore, the holistic dimensions, implemented as part of an ergonomic tool, again appear to be operating appropriately.

V. General Discussion

The purpose of this paper is to present current developments to the EvoFIT facial composite system. Previous data suggested that EvoFIT produces better composites than computerised systems in current use, but could be improved by adding transforms such as ageing, masculinity or weight. To realize this, 250 faces were photographed to allow a new face model to be constructed. Two hundred of these faces (along with 20 other ‘test’ faces) were then rated along 8 forensically-useful dimensions to allow the top and bottom 40 faces to be identified, one pair for each dimension. Averaging of these faces provided reference vectors in the PCA face space and allowed a face to be manipulated holistically. The results of three evaluations using perceptual and identification tasks suggested that the holistic transforms were operating appropriately.

The work presented here is the first stage in the production of a composite system with holistic dimensions. Clearly, the augmented system now requires testing with witnesses in the laboratory, as we have done elsewhere during its development [10][12][13][14]. This would typically involve participant-witnesses viewing a target, then constructing a composite a couple of days later using procedures normally used with real witnesses, such as a Cognitive Interview. Arguably this evaluation would initially use the Holistic Tool at the end of construction, as done here. Follow-up work could explore its use throughout the composite session, rather than at the end, as this may guide the search more effectively. In both cases, comparison would be made against another composite system, for example PRO-fit.

We were surprised that the level of naming from the un-manipulated (veridical) composites from the new model was so low, at 5%, since previous work using these same targets and procedures with the existing model produced an average naming level of about 25% [21]. It is possible that the larger model is more difficult to search, due to the increase in the number of faces therein, and so the likenesses produced are correspondingly less accurate than those produced previously.

It is also possible that composites evolved in section IV-B above using the new model contained key dimensions that were relatively more average than those produced from the old model [20], as suggested in IV-C above. It turns out that the new model has an average age of 36 years, which is older than the celebrity faces featured here (M = 29 years); it is also evident that celebrity faces are quite attractive, while the average attractiveness of the new model is lower. Therefore, making the composites appear younger and more attractive, as suggested in IV-C, may improve the match with the target faces (a similar story may also apply to honesty). The holistic enhancements of IV-C may therefore overcome in part limitations of a more complex face model. It is perhaps worth mentioning that criminal faces tend not to be as attractive as famous faces and therefore sizeable positive attractiveness manipulations are less likely for composites constructed with real witnesses.

There are a number of other potential solutions to counter an increase in face space complexity. One possibility is to generate the initial faces more carefully so as to provide a better set of initial solutions than random points in face space. For example, the initial faces could be produced with characteristics that match a witness’s memory, by presenting only thin, middle-aged faces, if that is what is required. This could be achieved by estimating the position of each initial (random) face along a dimension of interest, such as weight, then manipulating the faces to have the same attribute, such as a thin face shape. While potentially providing a better set of starting faces, the approach would also allow the use of the holistic tools during evolution.

Clearly, this method might improve performance, but it may be the case that the space to be searched is still overly large and that poor performance may persist. An alternative could be to build smaller models that contain just faces which match a person's memory of a suspect, such as a model with thin faces. The data collect here also indicate that it may be quite important to get the average age of the model correct. While this general approach would preclude the use of the holistic tools during construction, the tools could be used in a post-hoc way, the same as we have done here, and with seemingly good performance.

A more optimal solution however might be to build a small face model that allowed the holistic tools to be used throughout construction. This is possible to implement. Firstly, the influence of each holistic dimension could be removed from the starting images, with reference to the subjective ratings in III-B, and thus produce a 'dimension-free' set. This could be achieved by manipulating each reference image to the average setting along each holistic dimension. For example, the influence of facial weight could be removed by transforming each reference face to have average weight. Secondly, the holistic dimensions themselves could be taken as the initial Principal Components in a PCA model, thus allowing their use during evolution. Finally, a small face model, perhaps containing 36-48 faces that broadly matched a witness's description of a suspect, could be built from the 'dimension-free' set and used to evolve a face. It is even conceivable that a witness might use the holistic tools at the start of construction, to get the holistic settings roughly correct, and then to continue the process as normal to evolve the remaining 'dimension-free' parameters. We are currently exploring this possibility.

We have also been exploring other approaches that might improve performance. One of the limitations of the current system is that only a relatively small number of points (about 70) are used to search a very large face space. There are other user interfaces that may do this better. For example, one approach might be to use a car-type steering wheel to 'drive' through face space, 'stopping' at faces with a preferable likeness. A moving interface of this type has already been successful in elevating the recognition of composites in general, through a moving caricature animation [9], but if applied here could allow a better set of faces to be presented to the Evolutionary Algorithm for breeding, thus facilitating the evolution.

VI. Conclusion

Composite systems are typically based on the selection of individual facial features. These systems are clearly not a good interface with human memory, which is more holistic in nature. The EvoFIT system was designed to capitalize on holistic face processing and has demonstrated good performance relative to feature systems. We have shown here that it is possible to create a set of psychologically useful dimensions and use them to successfully enhance composites produced from EvoFIT. Further work is planned that employs the tools with laboratory witnesses in a more formal evaluation. The work has also indicated that larger, more complex face models may not be optimal when evolving faces, and that versions more tailored to the description of a suspect may be better.

Acknowledgement

This work was funded by a grant from the Engineering and Physical Sciences Research Council, EP/C522893/1. The authors would like to thank both the Sensation Science Centre and the Glasgow Science Centre for granting permission to collect data, as well as Emily Nicholls, Shona Tait and Fiona McNeill for their assistance with composite production and data collection. Thanks are also given to Cindy Frowd for her insightful comments on an early draft of the paper.

References

[1] A.G. Goldstein, and J.C. Chance, "Recognition of faces and verbal labels", *Bulletin of the Psychonomic Soc.*, 7, 1976, pp. 384-386.

[2] D. Christie and H. Ellis, "Photofit constructions versus verbal descriptions", *Journal of Applied Psychology*, 66, 1981, pp. 358-363.

[3] H. Ness, "Improving facial composites produced by eyewitnesses", *Ph.D. thesis*, University of Stirling, unpublished.

[4] N. Brace, G. Pike, and R. Kemp, "Investigating E-FIT using famous faces", In A. Czerederecka, T. Jaskiewicz-Obydzinska and J. Wojcikiewicz (Eds.). *Forensic Psychology and Law*, Krakow: Institute of Forensic Research Publishers, 2000, pp. 272-276.

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

[6] G.M. Davies and M. Little, "Drawing on memory: Exploring the expertise of a police artist", *Medical Science and the Law*, 30, 1990, pp. 345-354.

[7] G.M. Davies, P. van der Willik, and L.J. Morrison, "Facial composite production: a comparison of mechanical and computer-driven systems", *Journal Applied Psychology*, 85, 2000, pp. 119-124.

[8] C.D. Frowd, V. Bruce, A. McIntyre, and P.J.B. Hancock, "The relative importance of external and internal features of facial composites", *British Journal of Psychology*, in press.

[9] C.D. Frowd, V. Bruce, A. McIntyre, D. Ross, and P.J.B. Hancock, "An application of caricature: how to improve the recognition of facial composites", *Visual Cognition*, under revision.

[10] C.D. Frowd, V. Bruce, H. Ness, C. Thomson-Bogner, J. Peterson, A. McIntyre, and P.J.B. Hancock, "Parallel approaches to composite production", *Ergonomics*, submitted.

[11] C.D. Frowd, V. Bruce, K. Storås, P. Spick, and P.J.B. Hancock, "An evaluation of morphed composites constructed in a criminal investigation", *Proceedings of The 16th Conference of the European Association of Psychology and Law*, 2006, in press.

[12] C.D. Frowd, D. Carson, H. Ness, D. McQuiston, J. Richardson, H. Baldwin, and P.J.B. Hancock, "Contemporary composite techniques: the impact of a forensically-relevant target delay", *Legal and Criminological Psychology*, 10, 2005, pp. 63-81.

[13] C.D. Frowd, D. Carson, H. Ness, J. Richardson., L. Morrison, S. McLanaghan, and P.J.B. Hancock, "A forensically valid comparison of facial composite systems", *Psychology, Crime and Law*, 11, 2005, pp. 33-52.
[14] C.D. Frowd, P.J.B. Hancock, and D. Carson, "EvoFIT: A Holistic, Evolutionary Facial Imaging Technique for Creating Composites", *ACM Transactions on Applied Perception (TAP)*, 1, 2004, pp. 1-21.
[15] C.D. Frowd, D. McQuiston-Surrett, S. Anandaciva, H. Ness, and P.J.B. Hancock, "An evaluation of US systems for facial composite production", *Ergonomics*, submitted.
[16] H.D. Ellis, H.D. "Practice aspects of face memory", In G. R. Wells and E. F. Loftus, editors, *Eyewitness Testimony. Psychological Perspectives*. England: Cambridge University Press, 1984. Ch. 2.
[17] C.E. Koehn, and R.P. Fisher, "Constructing facial composites with the Mac-a-Mug Pro system", *Psychology, Crime and Law*, 3, 1997, pp. 215-224.
[18] P.J.B. Hancock, V. Bruce, and A.M. Burton, "Recognition of unfamiliar faces", *TICS*, 4, 2000, pp. 330-337.
[19] P.J.B. Hancock, "Evolving faces from principal components", *Behav. Res. Meth., Instruments and Computers*, 32, 2000, pp. 327-333.
[20] C.D. Frowd, "EvoFIT: A Holistic, Evolutionary Facial Imaging System", PhD thesis, University of Stirling, 2002, unpublished.
[21] C.D. Frowd, V. Bruce, Y. Plenderleith, and P.J.B. Hancock, "Improving target identification using pairs of composite faces constructed by the same person", *IEE Crime and Security*, London, 2006, in press.
[22] S.J. Gibson, C.J. Solomon, and A. Pallares-Bejarano, "Synthesis of photographic quality facial composites using evolutionary algorithms", *Proceedings of the British Machine Vision Conference*, Editors R. Harvey and J.A. Bangham, 2003, pp. 221-230.
[23] C. Tredoux, and Y. Rosenthal, "Face reconstruction using a configural, eigenface-based composite system", Presented at SARMAC III, Boulder, Colorado, July 10, 1999.
[24] C.D. Frowd, V. Bruce, A. McIntyre, D. Ross, and P.J.B. Hancock, "Adding Holistic Dimensions to a Facial Composite System", *Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition*. Los Alamitos: Ca., 2006, pp. 183-188.
[25] L. Sirovich, and M. Kirby, "Low-dimensional procedure for the characterization of human faces", *J. Opt. Soc. Amer. A*, 4, 1987, pp. 519-524.
[26] C. Coin, and G. Tiberghien, "Encoding activity and face recognition", *Memory*, 5, 1997, pp. 545-568.
[27] I. Craw, and P. Cameron, "Parameterising images for recognition and reconstruction", *Proceedings of the British Machine Vision Conference*, Turing Institute Press and Springer Verlag, 1991.
[28] P.J.B. Hancock, V. Bruce, and A.M. Burton, "Testing principal component representations for faces", in J.A. Bullinaria, D.W.Glasspool and G. Houghton (Eds.), *Proceedings of 4th Neural Computation and Psychology Workshop*, London: Springer-Verlag, 1997, pp. 84-97.
[29] T.F. Cootes, K.N. Walker, and C.J. Taylor, "View-Based Active Appearance Models", *Proceedings of the International Conference on Face and Gesture Recognition*, 2000, pp. 227-232.
[30] R.E. Geiselman, R.P. Fisher, D.P. MacKinnon, and H.L. Holland, "Eyewitness memory enhancement with the cognitive interview", *American Journal of Psychology*, 99, 1986, pp. 385-401.
[31] H. D. Ellis, "Face recall: A psychological perspective", *Human Learning*, 5, 1986, pp. 1-8.

Charlie Frowd is based at the University of Stirling. He developed the first version of EvoFIT as part of his Ph.D. and has been improving it ever since. His research involves improving the quality of facial composites from composite systems in general, as well as the processes used with witnesses. More generally, he is interested in how we perceive familiar and unfamiliar faces. He is currently a Co-investigator on 2 EPSRC grants, one to develop EvoFIT, from which the current paper is produced; the other is a PPE-type project, and is developing a version of EvoFIT as an exhibit to the Sensation Science Centre, Dundee.

Vicki Bruce is Vice-Principal and Head of the College of Humanities and Social Science at the University of Edinburgh and holds a personal chair in Psychology there. As Head of College she is in charge of academic disciplines across the arts, humanities and social sciences including law, education and management. She is an experimental psychologist who has researched and published extensively in the areas of visual perception and cognition, particularly focussing on human face recognition and face perception. She continues her research in collaboration with colleagues at Stirling.

Alex McIntyre is a Research Assistant in the Department of Psychology at the University of Stirling, also undertaking a research degree, having recently graduated with first class honours in Psychology. She is interested in ways to improve the recognition of facial composites, by applying theoretical insights from elsewhere in the face perception literature. She is also researching how faces are learned and retrieved from memory.

David Ross has a first-class honours degree in Psychology from the University of Stirling. His research interests currently concern face-space and the caricature effect, and he has worked as a research assistant with Dr Charlie Frowd and Peter Hancock on the EvoFIT composite system. He is currently investigating some perceptual effects of caricature before going on to pursue a masters and PhD in the field of face recognition at Cardiff University.

Stephen Fields has an honours degree in Psychology from the University of Stirling. He has since been carrying out research at the University of Stirling on eyewitness testimony using the EvoFIT and PRO-fit facial composite systems.

Yvonne Plenderleith is a Clinical Psychology Assistant Practitioner for the Child and Adolescent Mental Health Service, NHS Forth Valley. As an undergraduate student of the University of Stirling, she developed an interest in the effect of Autism on perception and cognition. Since achieving her degree with First Class Honours, Yvonne has contributed to university research investigating face recognition and facial composite systems. Yvonne hopes to continue both experimental and clinical work.

Peter Hancock is a Senior Lecturer in Psychology at the University of Stirling. His PhD in Computing Science looked at the application of Genetic Algorithms to the design of Neural Nets. He moved on to study human face perception and the extent to which it might be explained by principal components analysis of face images. The combination of face PCA and Genetic Algorithms led to the idea behind EvoFIT. He has subsequently researched ways to further improve recall of face information by witnesses. He is also interested in evolutionary approaches to the understanding of human behaviour and in the role of temporal synchrony in the binding problem in early vision.