Creating Face Models from Vague Mental Images

V. Blanz[1], I. Albrecht[2], J. Haber[3], and H.-P. Seidel[2]

[1]Universität Siegen, Germany
[2]MPI Informatik, Saarbrücken, Germany
[3]SIZ / BLKA, München, Germany

Abstract

We present a novel approach to create plausible 3D face models from vague recollections or incomplete descriptions. This task plays an important role in police work, where composite facial images of suspects need to be created from vague descriptions given by the eyewitnesses of an incident.

Our approach is based on a morphable model of 3D faces and takes into account correlations among facial features based on human anatomy and ethnicity. Using these correlations, unspecified parts of the target face are automatically completed to yield a coherent face model. The system uses a novel paradigm for navigating face space and provides high-level control of facial attributes as well as the possibility to import facial features from a database. In addition, the user can specify a set of attribute constraints that are used to restrict the target face to a residual subspace. These constraints can also be enforced on the example faces in the database, bringing their appearance closer to the mental image of the user, and thus avoiding confusing exposure to entirely different faces. We also propose a novel approach for adapting the system to local populations based on additional image databases that are converted into our 3D representation by automated shape reconstruction.

We demonstrate the applicability of our system in a simulated forensic scenario and compare our results with those obtained by a professional forensic artist using state-of-the-art software for creating composite images in police work.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques I.4.10 [Image Processing and Computer Vision]: Image Representation—Hierarchical,Multidimensional,Statistical J.m [Computer Applications]: Miscellaneous—Forensic Sciences

1. Introduction

Creating pictures of faces from mental images or vague recollections has important applications in both art/design and forensic sciences: new virtual characters for movies, advertisement or computer games are created from the artist's imagination, and illustrations of criminals need to be generated from the eyewitnesses' descriptions. In both cases, the artist/witness (denoted as *source* in the following) has a mental image of the *target face* to be created, and it is important not to influence this mental image during the process of reconstructing the face. In general, the sources cannot give a complete verbal description of the target face, and they tend to recollect only some striking details, for instance bushy eyebrows or a hooked nose, while they are unaware of many other characteristics.

In this paper, we present an approach for generating models of human faces from vague mental images or incomplete descriptions. To demonstrate our approach, we have developed a system that could substitute commercial systems used by police for creating composite or photofit pictures.

Figure 1: *The face of a "cab robber" created from the cab driver's recollection is rendered into a background image. The inset shows a photograph of the actual person who was described by the cab driver.*

In contrast to most of these commercial systems, our approach has the following advantages:

- Our GUI offers intuitive ways to modify features of the target face in arbitrary order.
- Unspecified parts of the reconstructed face are automatically completed according to statistical properties.
- Anatomical/ethnical correlations within a face are taken into account automatically during reconstruction.
- A 3D face model is created that can be viewed from an arbitrary view point with arbitrary lighting conditions.
- The resulting face model can be rendered automatically into background images in appropriate pose and illumination.

While these features represent the contributions on the application side, the technical contributions of our system are as follows:

- We propose a new algorithm for navigating face space, which uses a set of attribute constraints that restrict the face to a residual subspace.
- The prediction of unspecified facial features is based on correlation between different face regions and features learned from the database. Our system makes the most plausible prediction, given the information provided by the user.
- Photographs of new individuals are used to augment the databases available for the 3D modeling process.
- Using attribute constraints and face exchange, the database of example faces is automatically adjusted to the user's specifications.

2. Related Work

A large body of literature on modeling faces has been published in the last decades. A good overview can be found in the textbook by Parke and Waters [PW96] and in the survey by Noh and Neumann [NN99]. In the context of this paper, we focus on publications that address the creation of models or images from incomplete descriptions.

Remarkably, one of the oldest publications in this context is the only one to our knowledge that takes into account statistical correlations between facial features [GC75]. The system supports only line drawings. Starting from the average face, it applies affine transformations and intensity changes of a fixed set of 17 features, and supports selection from a feature data base. The user is prompted to deal with features in a fixed order. Laughery and Fowler [LF80] found that artist sketches are superior to identi-kit images due to the limited amount of facial features and lack of shading in identi-kit images. This statement, however, may be outdated by recent commercial systems that are equipped with huge libraries of facial features. An interface for gradient-based face space navigation has been proposed in [CF04]. Principal components are used as navigation axes in face space, but facial features cannot be controlled individually. The authors focus on a comparison of different user interfaces.

The use of *genetic algorithms* for tracking the target face in a large face space is proposed in several publications, initiated by Caldwell and Johnston [CJ91]. A system for exploring face space is described by DiPaola [DiP02], where a descriptive language is used to specify features and state of a face in terms of numerical parameters. To navigate through face space, a genetic algorithm using approximately 25 parameters is employed. Gibson *et al.* [GBS03] describe a 2D facial composite system that uses an evolutionary algorithm. The EvoFIT facial composite system [FHC04] is a recent approach based on an evolutionary algorithm working on shape and texture PCA coefficients extracted from a database of 72 monochrome photographs.

Approaches to search for and retrieve an image from a mug-shot database are typically based on composite images. The database used in [WAL*94] is indexed by attributes of individual features obtained by PCA and landmarks. The best matches to a photofit target image are retrieved based on normalized correlation and weighted distance of the feature vectors. The system also supports fuzzy search queries. Both Brunelli and Mich [BM96] and Baker and Seltzer [BS98] proposed systems that use facial composites for database retrieval and PCA for computation of similarity. In addition, the user can modify the input image by feature selection from database images or by changing the PCA coefficients of individual features [BM96], or random composites can be created from the current database selections [BS98].

Our own approach is based on a morphable model as proposed by Blanz and Vetter [BV99]. In addition, we use the method presented in [BSVS04] to render our reconstructed faces into background images containing hair styles and/or scenery.

Several commercial systems for generating composite facial images are available. The PROfit™ system [ABM05] is a standard b/w photofit system based on a library of over 200.000 facial features, which can be composited in both frontal and 3/4 views. In addition to some caricature and aging simulation features, composite images from different witnesses can be morphed to generate average or in-between results. The E-FIT™ program [Asp05] is another photofit system for compositing b/w images from individual features. Default features can be used where no description is available. The system supports aging simulation and—together with the 3D E-FIT extension—fitting of 3D faces to the 2D composite images. In the PHANTOM PROFESSIONALxp© system [UNI05], a color *basic image* is selected from a large database based on key attributes such as gender, age, skin color, etc. The features of this basic image are then modified using affine transformations and/or import functionality according to the description of the witness. Aging within a limited time frame of approximately ten years is also supported. The Identi-Kit.NET™ system [Smi05] is a b/w implementation of the classical, hands-on identi-kit system on the computer.

To our knowledge, no commercial system available today supports anatomical/ethnical correlations among facial features and automatic completion of unspecified features according to statistical properties. Arbitrary 3D views are supported by only one system through an extension.

In recent computer games, such as *The Sims 2* or *FIFA Soccer 2006*, the user can model avatars according to his ideas. Although the character's faces can be adapted in quite detailed ways, the techniques employed are targeted towards generating rather coarse, textured face models suitable for hardware rendering. The user cannot select features from a database, and due to missing anatomical correlations it is possible to create non-plausible "alien" faces.

3. System Overview

In this section, we give a brief overview of the functionality and the features of our system. Upon start, an *average face* is shown in the display area of the program. The average face is computed as the average of 100 male and 100 female faces as described by Blanz and Vetter [BV99]. All subsequent user operations such as transformations or attribute editing result in an immediate update of the displayed face. To facilitate user interaction, we have implemented an operation history with multiple levels of "undo".

3.1. General Settings

As a first step, the user (i.e., either the source or an operator) may specify age, gender, and ethnicity of the target face. This is an optional step, which does not directly affect the target face. Instead, setting these parameters restricts the selection of example faces shown in the import dialog box to match the criteria specified for age, gender, and ethnicity (see Section 3.5).

3.2. Segments

Various editing operations on individual facial features (such as the nose, eyes, eyebrows, mouth) may affect the whole face due to correlations among the features and the overall face shape. For instance, bushy eyebrows are usually correlated to a male face and to a darker skin, whereas a wide mouth is typically correlated to a broader face and also to a broader nose. It is one of the main advantages of our approach that these correlations are taken into account automatically. In cases where only little is known or remembered of the target face, exploiting these correlations may add significantly to the faithfulness of the reconstruction. On the other hand, if the user knows exactly what detail should be changed without influencing any other facial attributes, it is desirable to constrain the effect of editing operations to a local area.

To this end, we have implemented the possibility to select an arbitrary combination of *segments*, which define the area on the face where editing operations may have an effect. Figure 2(a) shows the GUI of our program in the selection mode, where the different segments are color-coded. The user clicks on one or several different segments, which are then added to the selection mask shown in the small pixmap on the right. Some facial features are represented by a hierarchy of segments, which are color-coded in different shades of the same base-color. The nose, for instance, consists of individual segments for the nose bridge, the alar wings, the tip, and the overall nose area. In the selection mode, the user may add segments from any hierarchy level of any facial feature to the selection mask. To ensure a smooth transition at the border of segments, we apply a multiresolution decomposition and blending technique [BA83].

3.3. Affine Transformations

Any time during the reconstruction of the target face, affine transformations (rotation, translation, non-uniform scaling) may be applied to the entire face or a currently selected segment. These transformations are useful, for instance, to adjust the position and orientation of the face to fit into a background image or to account for an asymmetric layout of facial features.

3.4. Facial Attributes

The majority of editing operations in our system are used to change the shape and/or color of facial features. The editable characteristics of all facial features are denoted as *facial attributes* in the following. These facial attributes are pooled in groups according to the facial feature they affect. For example, all facial attributes that affect the shape or color of the nose are grouped together. Facial attributes can be modified on a continuous scale from $[-1 \ldots 1]$ (where 0 denotes the default value of the average face) using sliders. Figure 2(b) shows a dialog of our GUI with sliders for a variety of facial attributes of the nose. As a first step, we have included the following facial attributes in our system:

|	
overall	: masculine–feminine, slender–obese, dark–light skin, younger–older, intensity of freckles, intensity of beard shadow, attractiveness, caricature level
face shape	: round–angular, narrow–broad, pear–heart shape, hollow–puffy cheeks, prominence of cheek bones, pointed–broad chin, retruded–protruded chin, distance between lips and chin, intensity of double chin, intensity of nasolabial fold
eyes	: slitted–round, upwards–downwards inclination, horizontal distance of eyeballs, dark–light iris, color of iris, dark–light eyebrows, thin–bushy eyebrows, straight–curved eyebrows (separate for left and right side), horizontal distance of eyebrows, distance between eyebrows and eyes

(a)	(b)	(c)	(d)
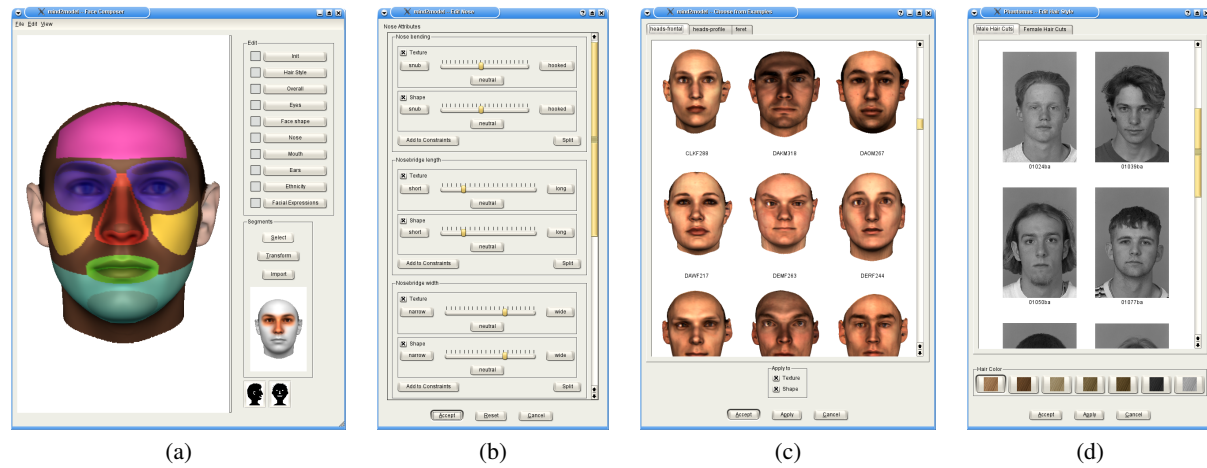
Figure 2: *Snapshots of our system's GUI. Left to right: (a) Main window in selection mode. The user clicks on one or several segments (color-coded in the main drawing area), which are then added to the selection mask shown in the small pixmap on the right side (see also Section 3.2). In this example, the user has selected both eyes and the nose for further editing operations. (b) Dialog for setting individual facial attributes of the nose. See Section 3.4 for details. (c) Database dialog for selecting features from a large database of example faces (see Section 3.5). (d) In the hair style dialog, the user selects a hair style and color (see Section 3.6).*

nose	: short–long nose bridge, narrow–wide nose bridge, narrow–wide alar wings, flat–round alar wings, snub–hooked nose, distance between nose and mouth
mouth	: narrow–wide, thin–full lips, dark–light lip color, convex–concave lip line
ears	: small–large, flat–jug ears
ethnicity	: Caucasian, Asian, African
expressions	: smiling, angry, surprised, scared, deranged, disgusted

The *caricature level* attribute is useful to overstate all characteristic attributes of a reconstructed face. Psychological findings [DJVO00] have shown that making faces more typical by caricaturing techniques increases subjects' ability to recognize faces. Caricaturing is achieved by morphing a face away from the average. Additional facial attributes can be easily integrated through parameter files. Section 4.1 explains how attributes are learned from a database of example faces.

During editing operations, the effect of each facial attribute on the target face (or on the current selection) can be restricted to the shape and/or texture of the target face. In addition, different values can be set for the shape and texture parameter of each attribute.

As explained in Section 3.2, some facial attributes are correlated to other facial features. Although this is a desired feature of our system, it might happen that the user wants a specific value for a facial attribute *A* to remain untouched when performing further editing operations on other facial attributes. This can be achieved by adding the current setting of attribute *A* to a list of constraints. Section 4.2 explains

how these constraints are enforced throughout further editing operations.

3.5. Importing Features from a Database

In addition to modifying the face or a feature of the face using the attributes described in the previous section, the user may also select the desired feature(s) from a database of example faces as shown in Figure 2(c). If a selection mask has been defined previously, only the corresponding features are imported from the selected face in the database. Thus it is possibly, for instance, to select the nose of the target face and then look for the desired nose among the example faces. Upon selecting a face from the example database, the nose of that face will be imported seamlessly into the target face and can be further edited if needed. As with attribute editing, the import of features from a database can be restricted to the shape and/or texture of the feature.

3.6. Hair Style

The face models from the database that have been used to learn facial attributes (see Section 4.1) have been captured without hair. Thus, we cannot handle hair in the same way as other facial attributes. Despite impressive progress in recent years, 3D hair modeling and rendering is still an open field of research. Even with photorealistic rendering techniques, collecting a database of hair styles would require to re-model existing styles. Therefore, we follow the approach of [BSVS04] and render the 3D face model into photographs of hair styles taken from different viewpoints.

For our hair style database, we selected a frontal and a profile image of people with several different hair styles from the FERET database [PWHR98]. After manually selecting 5–15 feature points in each hair style image once, the algorithm automatically estimates pose, illumination, and contrast for inserting arbitrary target faces. This estimation works for both color and monochrome photographs. We have chosen the monochrome images from the FERET database, since it was easy to color the hair in these images in seven different colors (different shades of blond, brown, red; black; grey) using an image manipulation program. Figure 2(d) depicts the hair style dialog of our system.

3.7. Exporting the Target Face

The target face can be saved as a 3D model for further operations. In particular, it is possible to change individual facial attributes later on or to display the face with different rendering parameters. In addition, the target face can be exported into an image using any desired set of rendering parameters. Rendering a face into an image containing scenery may help the source to recall details.

4. Learning-Based Face Modeling

The core element of the face modeling system is a morphable model of 3D faces [BV99]. The morphable model is a vector space of shape vectors $\mathbf{S}$ and texture vectors $\mathbf{T}$,

$$
\left.
\begin{aligned}
\mathbf{S} &= (x_1, y_1, z_1, \ldots, x_n, y_n, z_n)^T \\
\mathbf{T} &= (r_1, g_1, b_1, \ldots, r_n, g_n, b_n)^T
\end{aligned}
\right\} \quad (1)
$$

that contain the (x, y, z) positions and (r, g, b) color values of the vertices of a dense polygon mesh. Due to point-by-point correspondence between the $n = 75972$ vertices of our model, linear combinations of shape and texture vectors of the $m = 200$ example heads from our database produce realistic new 3D faces.

We perform a Principal Component Analysis (PCA) on shape and texture separately, ignoring the correlation between shapes and textures in the database. Still, we will be able to learn correlations between shape and texture in facial attributes as described in Section 4.1.

The following algebra is described only for shape, and applies to texture in the same way. Given the dataset of shape vectors $\mathbf{S}_i$, let $\overline{\mathbf{S}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{S}_i$ be the mean shape vector, and $\mathbf{X}$ the data matrix formed by columns $\mathbf{x}_i = \mathbf{S}_i - \overline{\mathbf{S}}$. The covariance matrix of the data is defined by $\mathbf{C} = \frac{1}{m} \mathbf{X}\mathbf{X}^T$, and PCA performs a diagonalization $\mathbf{C} = \mathbf{U} \operatorname{diag}(\sigma_i^2) \mathbf{U}^T$, where the columns $\mathbf{u}_i$ of the orthogonal matrix $\mathbf{U}$ are the principal components, and the eigenvalues are the squared standard deviations σ_i along the principal components. This diagonalization can be computed by a Singular Value Decomposition of $\mathbf{X}$:

$$
\mathbf{X} = \mathbf{U}\mathbf{W}\mathbf{V}^T, \quad (2)
$$

with a diagonal matrix $\mathbf{W} = \operatorname{diag}(w_i)$ and an orthogonal matrix $\mathbf{V}$. Substituting $\mathbf{X}$ in the definition of $\mathbf{C}$, we obtain the standard deviations $\sigma_i = \frac{1}{\sqrt{m}} w_i$.

Transformed into the orthonormal basis of principal components, shape vectors can be written as

$$
\mathbf{S} = \sum_i c_i \mathbf{u}_i + \overline{\mathbf{S}}. \quad (3)
$$

Even though most of the operations in the face modeling system work on the shape and texture vectors $\mathbf{S}$, $\mathbf{T}$ directly, the basis of principal components is essential for the following algorithm.

4.1. Learning Attributes from Examples

This section describes how to learn a vector $\mathbf{a}$ from examples, such that adding multiples of $\mathbf{a}$ to a given face changes a specific attribute while leaving all other individual characteristics unchanged. On 2D data, an intuitive method for finding $\mathbf{a}$ is to compute the averages (prototypes) of classes, such as male and female faces, and to use the difference vector between these prototypes [RP95]. For gradually varying attributes, such as body weight, a generalized method has been proposed [BV99], based on a weighted sum of examples. In this section, we derive this technique from a general statistical approach and discuss it in more detail, as it is the basis for the constrained modification paradigm presented in the following section.

Let $\mathbf{x}_i$ be a set of sample vectors (shape or texture), and $b_i \in \mathbb{R}$ be the ratings of a given attribute, such as body weight, for these 3D faces. b_i can be either given as ground truth, e.g. for age, or based on subjective ratings by the user. We have found that the learning approach presented in this section picks up trends in ratings reliably, so the ratings may be spontaneous and do not require too much care. Without loss of generality, we assume that the means of $\mathbf{x}_i$ and b_i are zero. Neither the vectors $\mathbf{x}_i$ nor the ratings b_i have to be scaled.

Our approach is to estimate the function f that assigns attribute values to faces, and then follow the gradient of this function in order to achieve a given change in attribute at a minimal, most plausible change in appearance. Given the limited set of data, we choose a linear regression for f, and minimize the least squares error

$$
E = \sum_{i=1}^{m} (f(\mathbf{x}_i) - b_i)^2. \quad (4)
$$

According to Riesz representation theorem, any linear functional f can be written in terms of a dot product with some vector $\mathbf{a}$. As we demonstrate later in this section, the gradient of f critically depends on the choice of the dot product. Intuitively speaking, the dot product defines the criterion for the direction of steepest ascent or descent in f. The canonical dot product of two vectors $\mathbf{x}$, $\mathbf{y}$ is the well known function $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_i y_i$ related to the L_2 norm. In contrast,
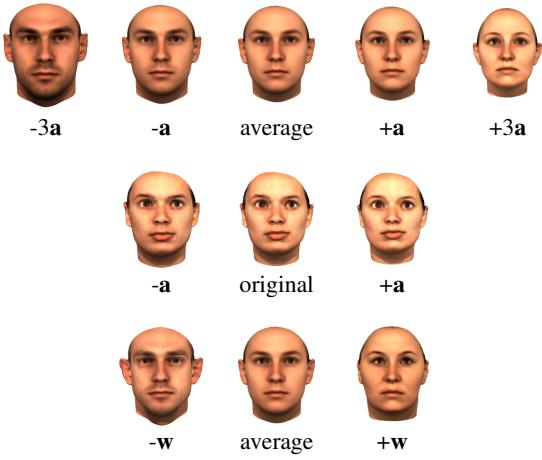
-3a   -a   average   +a   +3a

-a   original   +a

-w   average   +w

Figure 3: *Top row: attributes such as gender can be changed by adding or subtracting multiples of the attribute vector **a** to the average face. Second row: if applied to an individual face, characteristics such as the shape of the mouth are retained. Bottom row: simply adding the gradient $\nabla f = \mathbf{w}$, which defines the steepest slope of the attribute in L_2-norm, would affect individual features.*

we define a dot product

$$\langle \mathbf{x}, \mathbf{y} \rangle_M = \langle \mathbf{x}, \mathbf{C}^{-1}\mathbf{y} \rangle \qquad (5)$$

which is related to the Mahalanobis distance and the probability density estimated by PCA. In Mahalanobis distance $\|\mathbf{x}\|_M = \langle \mathbf{x}, \mathbf{x} \rangle_M$, distances are measured relative to the standard deviation observed in the dataset of examples. Let $f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{a} \rangle_M$, so

$$E = \sum_{i=1}^{m} (\langle \mathbf{x}_i, \mathbf{a} \rangle_M - b_i)^2 = \|\mathbf{X}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{a} - \mathbf{b}\|^2 = \min \qquad (6)$$

with the data matrix $\mathbf{X}$ and $\mathbf{b} = (b_1,...,b_n)^T$. Using the Singular Value Decomposition (2), we obtain

$$E = \|(\mathbf{UWV}^{\mathrm{T}})^{\mathrm{T}}(\frac{1}{m}\mathbf{UW}^2\mathbf{U}^{\mathrm{T}})^{-1}\mathbf{a} - \mathbf{b}\|^2, \qquad (7)$$

which simplifies to $E = \|m \cdot \mathbf{VW}^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{a} - \mathbf{b}\|^2$. The optimal solution **a** can be found using the Pseudo-Inverse, which is easy to compute here since the problem is already decomposed into orthogonal and diagonal matrices:

$$(m \cdot \mathbf{VW}^{-1}\mathbf{U}^{\mathrm{T}})^{+} = \frac{1}{m}\mathbf{UWV}^{\mathrm{T}} = \frac{1}{m}\mathbf{X}, \qquad (8)$$

so

$$\mathbf{a} = \frac{1}{m}\mathbf{Xb} = \frac{1}{m}\sum_{i=1}^{m} b_i \mathbf{x}_i, \qquad (9)$$

which is simply a weighted sum of the input vectors. It can be shown that $f(\mathbf{x} + \frac{m}{\|\mathbf{b}\|^2} \cdot \mathbf{a}) = f(\mathbf{x}) + 1$ for any vector $\mathbf{x}$, and that the direction **a** defines the steepest ascent or descent in terms of Mahalanobis distance. The result of adding multiples of **a** to faces is shown in the top two rows of Figure 3.
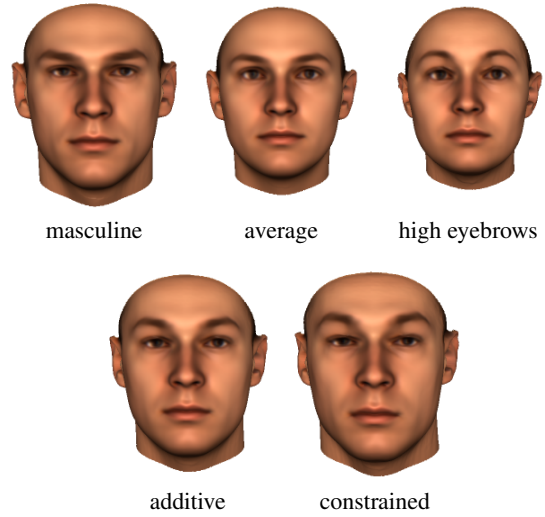

masculine   average   high eyebrows

additive   constrained

Figure 4: *Effect of constrained combination. Top row: the shape attributes "masculine" and "high eyebrows" have partly conflicting effects on the average face. The masculine face has low eyebrows, while the face with high eyebrows appears feminine. Unlike the additive solution, where the attributes partly cancel out (bottom left), the constrained combination retains the selected masculine appearance and eyebrow distance (bottom right).*

In order to motivate the use of the dot product $\langle \cdot \rangle_M$, consider the same problem with the standard dot product, which will lead to a vector **w** different from the previous solution:

$$f(\mathbf{x}_i) = \langle \mathbf{x}_i, \mathbf{w} \rangle, \quad E = \|\mathbf{X}^{\mathrm{T}}\mathbf{w} - \mathbf{b}\|^2 = \min. \qquad (10)$$

This can be solved using the Pseudo-Inverse $\mathbf{X}^{T+}$ of $\mathbf{X}^T$: $\mathbf{w} = \mathbf{X}^{T+}\mathbf{b}$. For manipulating attributes of faces, we would then add multiples of the gradient $\nabla f = \mathbf{w}$. This vector achieves the desired change in $f(\mathbf{x})$ even with a very small effect on shape and texture in terms of L_2-norm, which is hardly visible (see Figure 3 bottom row). However, if larger multiples of **w** are added for more perceptible changes, individual characteristics are no longer retained.

For each attribute, we estimate vectors **a** both for shape and texture, and the user may apply both of them at the same time. Our method, therefore, captures correlations between shape and texture, such as the fact that masculine versus feminine features can be found both in shape and in texture, or that faces with bright skin complexion might differ in shape from darker faces.

4.2. Enforcing Constraints on Faces

Previous systems for face modeling have relied either on morphing between existing example faces, or additive changes that add multiples $r \in \mathbb{R}$ of vectors **a**: $\mathbf{x} \mapsto \mathbf{x} + r\mathbf{a}$. Dealing with correlated attributes, however, these methods are suboptimal, as demonstrated by the following example:
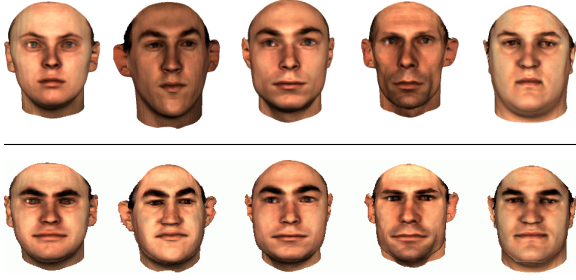
Figure 5: *Applying constraints to example faces. Top row: original faces. Bottom row: transformed faces with common settings for skin color, obesity, mouth width, and eyebrow bushiness.*

gender is correlated with the distance between eyes and eyebrows, and if the user first selects a value for masculine appearance and then lifts the eyebrows, the result will look less masculine than desired, see Figure 4 (bottom left). As a result, the user would have to re-adapt attribute values in an iterative procedure. In contrast, our method retains the masculine appearance by restricting subsequent modifications to the residual subspace of shapes and textures. Higher eyebrows are compensated automatically by a more male overall shape as shown in Figure 4 (bottom right).

In our system, the user may decide to fix attribute values as constraints for subsequent operations after he has selected a face from the database or applied modifications already. After each subsequent operation, our system will then map the face back to the residual subspace defined by the constraints. Moreover, our system transforms all faces in the database to meet these constraints (see Figure 5), so the user may select subsequent faces or facial features from a choice that is automatically adapted to his settings.

Let $\mathbf{x}_0$ be the shape or texture value of the current face (after subtracting the mean, without loss of generality). For each attribute j, $j = 1 \ldots N$, we compute the current value

$$b_j = \langle \mathbf{x}_0, \mathbf{a}_j \rangle_M \qquad (11)$$

These values are stored as current settings, and enforced on other faces. In the following, we describe the algorithm for mapping a new face $\mathbf{x}$ to the closest element $\mathbf{x}'$ that meets the constraints $\langle \mathbf{x}', \mathbf{a}_j \rangle_M = b_j$, $j = 1 \ldots N$.

Constrained attributes j are re-adjusted to values b_j by adding multiples r_j to the current head vector:

$$\mathbf{x}' = \mathbf{x} + \sum_{j=1}^{N} r_j \mathbf{a}_j. \qquad (12)$$

Then, the set of constraints is

$$b_j = \langle \mathbf{x} + \sum_{k=1}^{N} r_k \mathbf{a}_k, \ \mathbf{a}_j \rangle_M \qquad (13)$$

$$= \langle \mathbf{x}, \mathbf{a}_j \rangle_M + \sum_{k=1}^{N} r_k \cdot \langle \mathbf{a}_k, \mathbf{a}_j \rangle_M. \qquad (14)$$

eyes — nose
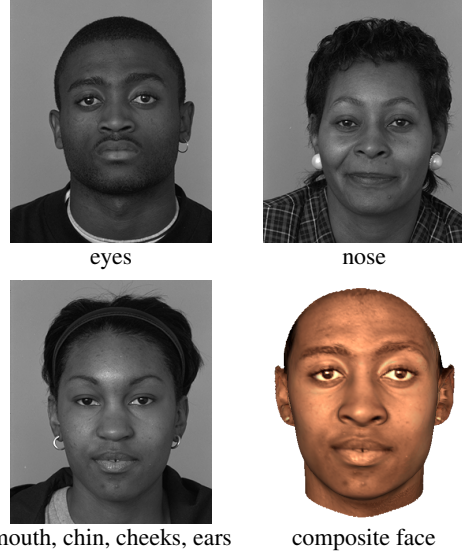
mouth, chin, cheeks, ears — composite face

Figure 6: *Adapting our system to local populations. From single photographs of three Afro-American individuals (top row and bottom left), we reconstructed 3D models using the 3D dataset of 199 Caucasian faces and one Asian face. The composite face (bottom right) is built from facial features of these three individuals.*

Our algorithm solves this system of equations for r_k, and computes the output face $\mathbf{x}'$ using Equation (12).

Note that for computing dot products such as $\langle \mathbf{x}, \mathbf{a} \rangle_M$, an explicit computation of the high-dimensional matrix $\mathbf{C}^{-1}$ is not required, due to the decomposition of $\mathbf{C}$ into lower dimensional matrices obtained from PCA: $\langle \mathbf{x}, \mathbf{a} \rangle_M = \langle \mathbf{x}, \mathbf{C}^{-1} \mathbf{a} \rangle = \langle \mathbf{x}, \mathbf{U} \mathbf{W}^{-2} \mathbf{U}^T \mathbf{a} \rangle = \langle \mathbf{W}^{-1} \mathbf{U}^T \mathbf{x}, \mathbf{W}^{-1} \mathbf{U}^T \mathbf{a} \rangle$.

4.3. Manipulation of Faces in Images

An important part of the proposed system is the replacement of faces in given photographs. This is achieved by fitting the morphable model to an image, which gives an estimate not only of the original head, but also of all relevant rendering parameters such as pose and illumination [BSVS04]. The preprocessing step of the background images requires manual clicking of 5–15 feature points, such as the tip of the nose or the corners of the mouth. Using the estimated rendering parameters, the new face is rendered automatically into the image. The method includes a background continuation step that removes the silhouette of the original face from the image, and some semi-automated processing of the foreground hair.

4.4. Adapting the Model to Local Populations

The current database of 3D scans contains 199 Caucasian faces and one Asian face, with an equal number of male and female individuals. Depending on where the system is used,
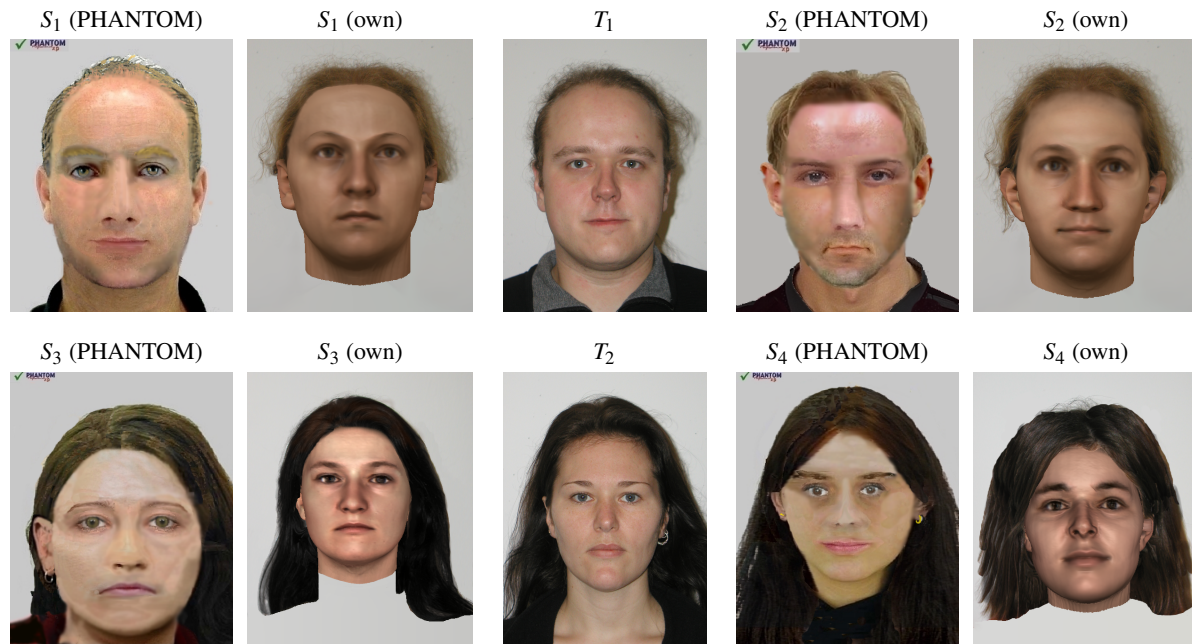
Figure 7: *Results of user study. Two target persons (T_1 and T_2) have been shown face-to-face to four source persons (S_1–S_4) for a duration of 60 seconds. The sources described the target faces after a delay of half-a-day up to one day. A professional forensic artist used the PHANTOM PROFESSIONALxp$^©$ system [UNI05] and Adobe Photoshop$^®$ for post-processing to create the composite images indicated by the keyword PHANTOM in about 2–3 hours each. The images created with our own approach took about 1.5–2 hours each.*

it will be desirable to adjust the choice of faces to the local population.

The system proposed in this paper can be extended easily to cover a wider range of faces without collecting more 3D scans. Given a set of single photographs of additional faces, we reconstruct 3D models by fitting the morphable face model [BV99] and include these 3D faces for selecting segments. Even though the reconstructed shapes are in the linear span of the original database, the method produces faithful reconstructions that capture the details of the new faces. In terms of texture, mapping the color values from the images to the model using illumination-corrected texture extraction as proposed in [BV99] adds new dimensions to the vector space. Figure 6 demonstrates that typical features of faces can be transferred to the composite image given only single images of new individuals.

5. Results

To evaluate our system, we conducted a user study involving four source persons (S_1 and S_3 female, S_2 and S_4 male, age range 25–55 years) and two target persons (T_1 male, T_2 female, age range 25–30 years) unknown to each other. Each target person had been shown face-to-face to two source persons for a duration of 60 seconds. After a delay of half-a-day up to one day, all source persons participated in reconstructing the target faces using both our system and the PHAN-

TOM PROFESSIONALxp$^©$ system [UNI05] according to the following schedule:

target	source	after half-a-day	after one day
T_1	S_1	own	PHANTOM
T_1	S_2	PHANTOM	own
T_2	S_3	PHANTOM	own
T_2	S_4	own	PHANTOM

The sources described the target faces to the operators of the systems and provided feedback on the intermediate stages during the process. The operators did not know the target persons.

A forensic artist with a professional background of having created more than 2000 composite facial images created images of the target faces using the PHANTOM system and Adobe Photoshop$^®$ for post-processing. In the process, a face is crudely assembled from parts of faces in the database of the photofit system and then imported into the image processing system for the final touches. Each composite image has been created in about 2–3 hours.

In contrast, our system was operated by computer scientists who naturally lack the forensic professional's psychological background and finesse in conjuring up a detailed image of the target face before the source's mental eye. Since the variety of hair styles in our system was not sufficient for the sources to find a satisfying hair style, we interactively created hair style images according to the sources descrip-

tions using an image editing program. It took on average 1.5–2 hours to create each target face model.

A comparison of the results created by both our and the commercial PHANTOM system is shown in Figure 7. The overall quality of the results shows that creating recognizable composites from source descriptions is a hard problem. In their experiments Laughery and Fowler [LF80] compared the quality of sketches to identi-kit composites and concluded that the identi-kit technique itself was the main limiting factor. The extensive use of Adobe Photoshop® by the forensic artist excludes this reason. Both for the commercial product operated by the expert and our system operated by laymen, the main constraint was the sources' limited ability for accurate recall rather than the flexibility of the system.

For both target faces, the forensic artist considered a reconstruction with our system the best (S_1 for T_1 and S_3 for T_2, except for the missing ear rings) and stated that he anticipated them to be helpful in a real investigation. In particular, he judged our results as superior in terms of overall face shape, and shape and position of facial features (eyes, nose, mouth). He also appreciated the intuitiveness and continuous variability of the attribute sliders.

One source person favored the PHANTOM system, while the other three sources felt more comfortable with our system. One reason for their preference was the fact that the professional composite showed a static front view only. Although in the PHANTOM system it is theoretically possible to also create an additional $90°$ view of the face, this requires the database to be present in a second side-face version. Even if this had been the case in our forensic artist's version, it would have been very time-consuming, since all post-processing steps for the front view must be executed analogously for the side view. The sources also missed the possibility to replace facial features during the post-processing without loosing all intermediate steps. Some subjects had problems with the database approach in general. One source complained that "*having seen so many other faces* [in the database]*, I can't recall what's right or wrong anymore*", while another said: "*Despite the huge amount of faces in the database, I wasn't able to find the right eyes: I couldn't imagine what all those eyes would look like in the face I was going to describe.*"

We found indeed that most of the statements of sources were of the type "*make the nose wider*", so the slider-interface turned out to be more appropriate in most of the reconstruction process than the standard data base selection of the professional systems that our software also offers.

Rendering the composite face into an image of a simulated crime scene matching the experience of the witness (as in Figures 1 and 8) will set the face into proportion to its surroundings and may thus help the source to remember valuable detail.

Figure 8: *The perpetrator's face has been replaced by a face model created with our system. The underlying mental image is shown as an inset in Figure 1.*

6. Conclusions and Future Work

Creating recognizable facial composites is a difficult problem due to the inadequacy of the human brain with respect to face recall. The main advantage of our system is that, due to the underlying statistical model, it supports the user in bringing his mental image to the screen without restraining him. For overall editing operations, the program presents the most probable solution while leaving the user the freedom to override this result. The system is intuitive to control and fine-grained. The user always works with an entire face and hence need not consider isolated facial features. Viewing the composite face model from several directions is helpful when defining the overall face shape, when editing silhouette features, or to simulate a different perspective, for instance, of a small person describing a tall man. By definition of the morphable model, face models of a single target created from several sources can be morphed easily to obtain weighted combinations.

Psychological studies indicate that human memory for faces is affected through retroactive interference if subjects are exposed to images of different faces between the study and test phase [DCL81]. Therefore, we reduce the variety of faces our sources are exposed to by adjusting the example faces to meet the user-specified constraints. Moreover, wrong clues about features of the target face interfere with the sources' recollection significantly [JD85]. Our technique accounts for this by showing at any time the most plausible face according to the correlation estimated from the database.

Our control experiment with an experienced professional who was operating commercial software shows that, apparently, the problem of reconstructing faces is intrinsically difficult, having seen these unknown faces for only 60 seconds and describing them half-a-day or one day later. The comparison of the results indicates that the limiting factor is the deficiency of human consciousness with respect to exact recollection of faces, not the accuracy of our system. The positive reactions of the forensic artist and the sources indicate that our system helps to obtain better facial composites in a less tedious process.

For adding details such as birth marks, scars, or jewelry, we currently propose to follow the forensic artist's approach and finish an image of the completed head in the desired view using an image editing program. Especially for computer games, however, it would be desirable to have data bases of such accessories and facial particulars as well as beards and hair styles in 3D, in order to be able to add such detail at any stage of the creation process, and to allow views of the final composite head from any direction.

Using a similar approach, one could generate photospreads for identification of suspects by witnesses. Typically, a photospread consists of a picture of the suspect and five to nine more or less similar "fillers". These fillers could be generated by defining constraints for the most prominent features of the suspect, such as nose curvature for a very hooked nose, and varying the rest of the face by constrained random permutation.

References

[ABM05] ABM UNITED KINGDOM LTD.: PROfit™. http://www.abm-uk.com/uk/products/profit.asp, 2005.

[Asp05] ASPLEY LTD.: E-FIT™. http://www.efit.co.uk, 2005.

[BA83] BURT P. J., ADELSON E. H.: A Multiresolution Spline with Application to Image Mosaics. *ACM Transactions on Graphics 2*, 4 (Oct. 1983), 217–236.

[BM96] BRUNELLI R., MICH O.: SpotIt! An Interactive Identikit System. *Graphical Models and Image Processing 58*, 5 (1996), 399–404.

[BS98] BAKER E., SELTZER M.: The Mug-Shot Search Problem. In *Proc. Vision Interface '98* (1998), pp. 65–72.

[BSVS04] BLANZ V., SCHERBAUM K., VETTER T., SEIDEL H.-P.: Exchanging Faces in Images. In *Computer Graphics Forum (Proceedings of Eurographics 2004)* (Aug. 2004), Cani M.-P., Slater M., (Eds.), vol. 23, pp. 669–676.

[BV99] BLANZ V., VETTER T.: A Morphable Model for the Synthesis of 3D Faces. In *Computer Graphics (SIGGRAPH '99 Conference Proceedings)* (Aug. 1999), Rockwood A., (Ed.), ACM SIGGRAPH, pp. 187–194.

[CF04] CHEN T.-P. G., FELS S.: Exploring Gradient-based Face Navigation Interfaces. In *Proc. Graphics Interface 2004* (2004), pp. 65–72.

[CJ91] CALDWELL C., JOHNSTON V. S.: Tracking a Criminal Suspect through "Face-Space" with a Genetic Algorithm. In *Proc. Fourth International Conference on Genetic Algorithms* (1991), pp. 416–421.

[DCL81] DEFFENBACHER K. A., CARR T. H., LEU J. R.: Memory for Words, Pictures, and Faces: Retroactive Interference, Forgetting, and Reminiscence. *J. Experimental Psychology 7*, 4 (1981), 299–305.

[DiP02] DIPAOLA S.: FaceSpace: A Facial Spatial-Domain Toolkit. In *Proc. Information Visualisation '02* (2002), pp. 105–109.

[DJVO00] DEFFENBACHER K. A., JOHANSON J., VETTER T., O'TOOLE A. J.: The face typicality-recognizability relationship: encoding or retrieval locus? *Memory and Cognition 28*, 7 (2000), 1173–1182.

[FHC04] FROWD C. E., HANCOCK P. J. B., CARSON D.: EvoFIT: A Holistic, Evolutionary Facial Imaging Technique for Creating Composites. *ACM Trans. Applied Perceptions 1*, 1 (July 2004), 19–39.

[GBS03] GIBSON S., BEJARANO A. P., SOLOMON C.: Synthesis of Photographic Quality Facial Composites using Evolutionary Algorithms. In *Proc. British Machine Vision Conference 2003* (2003), Harvey R., Bangham J. A., (Eds.), pp. 221–230.

[GC75] GILLENSON M. L., CHANDRASEKARAN B.: A Heuristic Strategy for Developing Human Facial Images on a CRT. *Pattern Recognition 7*, 4 (1975), 187–196.

[JD85] JENKINS F., DAVIES G.: Contamination of Facial Memory Through Exposure to Misleading Composite Pictures. *J. Applied Psychology 70*, 1 (1985), 164–176.

[LF80] LAUGHERY K. R., FOWLER R. H.: Sketch Artist and Identi-Kit Procedures for Recalling Faces. *J. Applied Psychology 65*, 3 (1980), 307–316.

[NN99] NOH J.-Y., NEUMANN U.: *A Survey of Facial Modeling and Animation Techniques.* USC Technical Report 99-705, USC, Los Angeles, CA, 1999.

[PW96] PARKE F. I., WATERS K.: *Computer Facial Animation.* A K Peters, Wellesley, MA, 1996.

[PWHR98] PHILLIPS P. J., WECHSLER H., HUANG J., RAUSS P.: The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing J. 16*, 5 (1998), 295–306.

[RP95] ROWLAND D., PERRETT D.: Manipulating facial appearance through shape and color. *IEEE Computer Graphics and Applications 15*, 5 (1995), 70–76.

[Smi05] SMITH & WESSON®: Identi-Kit.NET™. https://www.identikit.net/, 2005.

[UNI05] UNIDAS: PHANTOM PROFESSIONALxp©. http://www.unidas.com/html/phantome.html, 2005.

[WAL*94] WU J. K., ANG Y. H., LAM P., LOH H. H., NARASIMHALU A. D.: Inference and retrieval of facial images. *Multimedia Systems 2* (1994), 1–14.