Manipulating Facial Appearance through Shape and Color

Duncan A. Rowland
St. Andrews University

David I. Perrett
St. Andrews University

A technique for defining facial prototypes supports transformations along quantifiable dimensions in "face space." Examples illustrate the use of shape and color information to perform predictive gender and age transformations.

Applications of facial imaging are widespread. They include videophones, automated face recognition systems, stimuli for psychological studies,[1,2] and even visualization of multivariate data. Most work thus far has addressed solely the manipulation of facial shapes, either in two dimensions[3,4] or in three dimensions.[1] In this article, we examine both shape and color information, and document transformational techniques in these domains by applying them to perceived "dimensions" of the human face.

The examples considered here are restricted to manipulations of the facial dimensions of age and gender, but the transformational processes apply to instances of any homogeneous object class. For example, Knuth's idea of a Metafont[5] can be thought of as a template of feature points mapped onto letter shapes in order to parameterize specific base typefaces. This enables the creation of new fonts by interpolating between or extrapolating away from the base fonts.

We consider objects to form a visually homogeneous class if and only if a set of template points can be defined which map onto all instances of that class. By "template points" we mean a set defining salient component features or parts. In the processes we describe here for faces, the eyes, lips, nose, cheek bones, and so on are used as features for template points. The template points delineate major feature "landmarks," following previous conventions (notably Brennan[3]).

Automating the selection of template points would be a boon to the process, since we could then generate a transform automatically for any two groups within a homogeneous object class (for example, airplanes or letters). However, the present level in understanding of qualitative object features does not support this automation. Creating a class template

requires understanding what about an object is "essential" (in the Platonic sense of that word). As several authors have discussed, humans have this ability, but so far computers do not.[6]

After selecting the feature template set, we manually record (delineate) the point positions on an exemplar, though this process may soon be automated.[7,8]

The processes we describe begin with the creation of a facial prototype. Generally, a prototype can be defined as a representation containing the consistent attributes across a class of objects.[9,10] Once we obtain a class prototype, we can take an exemplar that has some information missing and augment it with the prototypical information. In effect this "adds in" the average values for the missing information. We use this notion to transform gray-scale images into full color by including the color information from a relevant prototype.

It is also possible to deduce the difference between two groups within a class (for example, male and female faces). Separate prototypes can be formed for each group. These can be used subsequently to define a transformation that will map instances from one group onto the domain of the other.

Methods

The following sections detail the procedure we use to transform facial images and show how it can be used to alter perceived facial attributes.

Feature encoding and normalization

In our derivations of prototypes we have used facial images from a variety of sources. However, we constrained the image selection to frontal views with the mouth closed. (Separate prototypes could be made for profile views, mouths open with teeth visible, and so forth.) We also omitted images that showed adornments such as jewelry and other items obscuring facial features, such as glasses and beards.

The faces used here came from two distinct collections:

1. Faces photographed under the same lighting, identical frontal views, neutral facial expression, and no makeup.
2. Faces taken from magazines (various lighting, slight differences in facial orientation and expression, and various amounts of makeup).

Originally, we thought it would be important to tightly control the factors mentioned for collection 1. We found, however, that if we use enough faces to form the prototype ($n > 30$), the inconsistencies disappear in the averaging process and the resulting prototype still typifies the group.

The faces forming collection 1 consisted of 300 male and female Caucasians (ages 18 to

65). The images (for example, Figure 1a) were frame grabbed in 24-bit color at an average interpupilary distance of 142 pixels (horizontal) and full-image resolution of 531 (horizontal) by 704 (vertical) pixels. Impartial subjects rated the images as to the strength of traits such as attractiveness and distinctiveness. We also recorded objective values such as age and gender. Then we used this information to divide the population into classes from which prototypes could be derived. The faces forming collection 2 (average interpupilary distance of 115 pixels) were also rated for similar traits, although objective values (except gender) were unavailable.
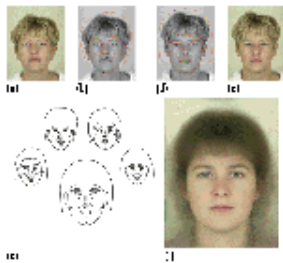

Figure 1. Deriving a facial prototype. Original shape of an individual face depicted (a) in color and (b) in black and white with superimposed feature delineation points; (c) four face shapes delineated, surrounding a prototype shape made by averaging the feature positions across 60 female faces (ages 20 to 30). Original face warped into prototypical shape, (d) with and (e) without corresponding feature points marked; (f) color prototype made by blending the 60 original faces after they had been warped into the prototype shape.

The choice of feature points was described previously.[3,4,10] We allocate 195 points to a face such that point 1 refers to the center of the left eye, points 2 to 10 refer to the outer circle around the left iris, and so on as shown in Figure 1b. Using a mouse, an operator delineates these points manually for each face. (Operators practiced a standard delineation procedure to maintain high interoperator reliability.) If a feature is occluded (for example, an ear or eyebrow hidden by hair), then the operator places the points so that the feature is either an average size and shape or symmetrical with visible features. An additional 13 feature points are placed manually to form a border outside the hairline, and a further 22 points are defined automatically at equally spaced intervals along the external borders of the image. These extra points provide additional anchors for tessellating the image during warping.[4] Hairstyles vary considerably and are not the subject of current manipulations, although the color of the "hair region" is affected by the transformation.

Creating prototypes

To derive the prototypical shape for a group of faces, the delineation data must be "normalized" to make the faces nominally of the same size and orientation. The left and right pupil centers provide convenient landmark points for this process. The first step is to calculate the average left and right eye positions for the whole population. Then, we apply a uniform translation, scaling, and rotation to the (x, y) positions of all the feature points, thus normalizing each face to map the left eye to the average left eye position and the right eye to the average right eye position. This process maintains all the spatial relationships between the features within each face but standardizes face size and alignment.

The four outer faces in Figure 1c indicate the shape variations between four group members. When we calculate the average positions of each remaining template point (after alignment), the resulting data constitutes the mean shape for the given population (Figure 1c center).

We obtained the prototypical color information p by warping each face in the set to this population mean shape (see Figure 1d and Figure 1e). A detailed discussion of the face warping technique is given elsewhere.[4] The resultant images were then blended with equal weighting. Every pixel position in each image has a red, green, and blue (RGB) brightness value describing its combined color and intensity. For an individual face i, the triple (R, G, B) can describe the value of a pixel position k. Let the function $F(i, k)$ yield this value. The prototype color-value triple $F(p, k)$ for an individual pixel is defined as the mean of the RGB values for the same pixel position in each constituent normalized face. In other words, since we have aligned each face for all features to coincide across the normalized faces, this resultant blend contains the population mean coloration and the population mean face shape (Figure 1f).

$$F(p, k) = \left[\sum_{i=0}^{n-1} F(i, k) \right] / n$$

where n is the number of faces in the group.

Color addition

By creating a prototype from just the females in the face set, we obtain mean (or prototypical) shape and color information relating to Caucasian females. In this example we augment a black-and-white image of a female face with this color information to produce a full-color image. RGB values are useful as a color coordinate system, since they define a 3D color space with each axis describing the brightness of the red, green, and blue phosphors on a color monitor. Within this color space, we can calculate the mean RGB value for a set of colors (and also interpolate between two colors) since a color is specified by a vector (R, G, B).

Because RGB space is defined by red, green, and blue axes, it does not have any axis relating to luminance. One system that does separate color and luminance information is HLS space (hue, lightness, saturation). These three values correspond conceptually to the controls on some color television sets. Hue refers to the tint (or color bias), saturation corresponds to the amount of hue present, and lightness is comparable to the luminance (or the black-and-white information). Saturation can vary from a maximum, which would display on a TV image as vivid colors, to a minimum, which is equivalent to black-and-white film. Although HLS color space is useful for splitting a color into meaningful conceptual categories, it is less useful than RGB color space for calculating means or for interpolation purposes (mainly due to the circular nature of the hue).

A transformation that maps triplets of RGB values to a corresponding HLS triplet has been defined and reviewed elsewhere.[11] These alternative coordinates of color space are useful for separating and combining luminance and color information. Since it is possible to split an image into HLS components, we can also take the color information (hue and saturation) from one image and add it to a second image containing only lightness information. Thus, we can take a black-and-white photograph of a Caucasian face and

transform it into a color image by adding artificial color from a second appropriately "shape matched" prototype derivative.

Figure 2 illustrates this process. Delineating the source black-and-white photograph defines the example face's shape information (see Figure 1b). Warping the young female prototype into this shape, as shown in Figure 2a, generates the relevant (or at least plausible) color information for the source (for example, the average red pigmentation of the lips from the prototype is aligned with the gray lip region of the source black-and-white photograph. Taking the lightness component for each pixel from the original photograph (Figure 2b) and adding the hue and saturation components from the corresponding pixels in the shape-matched prototype (Figure 2c), we can generate an HLS color reconstruction of the source face (Figure 2d).

Figure 2. Adding color to a black-and white face image. (a) Female face prototype warped into shape of the black-and-white example face in Figure 1b; (b) lightness information from the example face; (c) color information (hue, saturation with uniform lightness) from the prototype; (d) HLS color image resulting from adding lightness information in (b) to hue and saturation color information in (c). The reconstructed color closely matches the original in Figure 1a.

The faces included in the female prototype were matched to the source face with respect to gender, age, and ethnic background. In this example, the original color information was available (see Figure 1a), which allows a comparative assessment of the success of the color addition technique in Figure 2d. The technique could, however, be applied to face images for which there was no color information, such as film stars from old monochrome films.

Difference analysis between face prototypes

By splitting face collection 2 into two groups, males and females (age rated 20 to 30), we can derive separate prototypes p_m and p_f respectively (see Figure 3). The shape and color differences between the male and female prototypes should be entirely due to the difference in gender (since the samples are matched for age and so on). If there were no consistent differences between male and female faces, both blends would be identical.
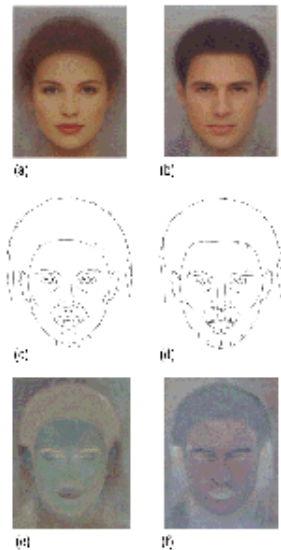
Figure 3. Defining shape and color differences between prototypes. Prototypes of (a) 61 female and (b) 43 male Caucasian fashion models. The shapes of the prototype model (c) female faces and (d) male faces displayed in (c) and (d), along with superimposed vector differences between the position of each feature point. Color differences between prototypes expressed as (e) the way the female differs from the male and (f) vice versa.

We can use any shape and color differences existing between prototypes to define a transformation relating to the gender of a face. The transform itself will contain no information relating to identity because, as we have said, the only differences between one prototype and its counterpart are the dissimilarities due to the trait used to group the constituent faces.

We must be wary though that the chosen trait is "real" (that is, physically apparent). Francis Galton[9] created the notion of a prototypical photographic image by applying the techniques of composite portraiture to various groups of faces. In one example, he tried to create a prototype from "criminal types," perhaps hoping to save the police a lot of time in their inquiries! To his surprise, the resulting composite image was not fundamentally different from composites created by "blending" other segments of the community.

The visual differences between face prototypes may have biological origins (for example, chin shape) or cultural origins (as in hair length).

It is useful to be able to quantify or at least visualize the differences between prototypes. Figure 3 illustrates an analysis of shape and color differences for female and male prototypes (from face collection 2). Shape differences are displayed in the middle row and color differences in the bottom row. The female prototype shape is shown (middle left) as a facial outline. The vector difference from female to male face shapes for each template point has been superimposed (after aligning the pupil centers for the male and female prototypes). The vector differences (lines leading away from the outline shape) show that male and female faces consistently differ in shape. Similarly the shape of the male prototype (middle right) is shown with the vector difference in shape from male to female.

We derived color differences between prototypes by first warping the male and female prototypes to the same intermediate "androgynous" shape (that is, half way between the average male and average female shape) and then calculating the difference between RGB values for corresponding pixels. For each pixel, we subtracted the female pixel value from the male pixel value. Then we added the result to (or in the later case, subtracted it from) a

uniform gray image $g = (R_{max}/2, G_{max}/2, B_{max}/2)$. For example, for 24-bit images with 8 bits for each of the red, green and blue pixel values, $g = (127, 127, 127)$. Addition defines the way the female face prototype p_f differs in color from the male prototype p_m (Figure 3e). Subtraction defines the way the male prototype differs from the female prototype (Figure 3f).

More generally, for two prototypes (p_1 and p_2) we can visualize the difference between the two prototypes by adding the difference to a uniform gray box. Thus, the RGB value in the resultant image r at pixel position k can be defined as $F(r, k)$, so

$$F(r, k) = g + [\ F(p_1, k) - F(p_2, k) \ ]$$

The methods we have described are useful for visualizing the differences in shape and color between any two face prototypes. For example, the procedures for defining shape vectors were used to illustrate the differences between attractive and average face shapes.[2] When the difference between prototypes is small, visualization can be aided by amplifying differences in shape and color with the methods described below.

Transformation based on prototypes

By taking an individual facial image and applying a color and shape transform, we can selectively manipulate the image along a chosen facial dimension. To define the transform, we need the color and shape information for two prototypes that define the chosen dimension (for example, male-female, young-old, happy-sad).

Manipulation of apparent gender. We can use the shape and color differences between prototypes illustrated in the last section to define a transform manipulating the apparent gender of an individual face. The transformations in shape and color are applied separately, but each follows equivalent principles.

Figure 4 (a and b) illustrates the shape transform associated with gender. To perform the transform, we first normalize the shape information for male and female prototypes with respect to the eye positions of the source face. For each feature point, we calculate a difference in position between the prototypes. This corresponds to a scaled version of the vector difference illustrated in Figure 3c and Figure 3d. We can then add a percentage (1) of this vector difference to each corresponding feature point from the source face.
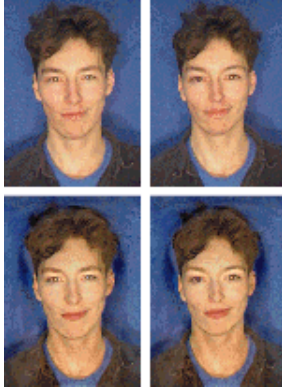
Figure 4. Applying shape and color transformations for gender differences. (a) Original face, (b) shape transformed: original face + 100 percent shape difference between female and male prototypes, (c) color transformed: original face + 100 percent color difference between female and male prototypes, (d) shape and color transformed: original face + 100 percent shape and color difference between female and male prototypes.

The shape transform can be expressed by simple parametric equations: For the x and y horizontal and vertical pixel positions of each feature point,

$$x_r = x_s + 1(x_f - x_m)$$

$$y_r = y_s + 1(y_f - y_m)$$

where r denotes the resultant face, f denotes the first prototype (female), m denotes the second prototype (male), s denotes the source face, and 1 is the percentage of transformation to be applied.

Figure 4b illustrates the effect of the shape transform with 100 percent of the vector difference from male to female prototype added to the shape of the original source male face.

The shape transformation is limited in impact because a lot of gender information is missing, namely the color differences between male and female faces. We can perform a similar transformation in color space. Figure 4c shows the impact of a color transformation while maintaining the original shape of the source face. This transformation begins by warping each prototype into the shape of a source face. The RGB color difference between prototypes at each pixel is then calculated and subsequently added to the corresponding RGB value from the original image.

Again, we describe each pixel position k as a triple by representing the RGB color values as $F(p_m, k)$ for the males, $F(p_f, k)$ for the females (both warped to the shape of the source face), and $F(s, k)$ for the source face. By adding a percentage (1) of this difference between the male and female prototypes (exactly as we did with shape), the resultant color information $F(r, k)$ becomes more female.

$$F(r, k) = F(s, k) + 1 [F(p_f, k) - F(p_m, k)]$$

Combining these two transformations (see Figure 4d) is more effective than either shape or color manipulations alone (Figure 4b and Figure 4c). To combine the two transformations, we warp the prototypes to the shape of the image resultant from the shape transformation (for example, Figure 4b) and apply the color difference.

Figure 5 shows the effects of applying the "gender" transformation in different degrees and directions. By applying the transformation with 1 = -50, -100, we can move in slight steps from the image of a source female face (upper row, column 2) to an image with a more masculine appearance (columns 3 and 4). Similarly, by applying the inverse of this transformation (with 1 = +50, +100), a source image of a male face (lower row, column 2) can be transformed into a more female appearance (columns 3 and 4).

Figure 5. Varying the degree of transformation for gender. The figure shows two faces, one female (original image, upper row second column) and one male (original image, lower row second column) with different degrees (expressed as a percentage) of gender transformation applied. The color and shape transformations are applied in a positive direction to add same-gender characteristics (to the left) and in a negative direction to add opposite-gender characteristics (to the right). From left to right the images are (1) 50 percent (own gender attributes enhanced), (2) original images, (3) -50 percent (near "androgyny"), (4) -100 percent (gender switched), and (5) -150 percent (opposite gender attributes enhanced).

We can accentuate the masculine or feminine characteristics past those typical for an average male or female. To do this, we apply a gender transform greater than 100 percent in the direction towards the opposite gender (that is, 1 < -100 for females and 1 > +100 for males, as in column 5 top and bottom, respectively) or greater than 0 percent to enhance the same gender characteristics (that is, 1 > 0 for females and 1 < 0 for males, column 1 top and bottom, respectively). The effective amount of transform depends on the salience of the gender traits present in the source face images.

Manipulation of apparent age . To illustrate the generality of the processes described, we have used the same techniques to define prototypes for male faces of different ages and derived an aging transform from these. Two prototypes were formed, the first from the faces of 40 males aged 25 to 29 years and a second from the faces of 20 males aged 50 to 54 years.

Figure 6 shows the effects of the aging transformation applied to one face. We used the color and shape differences between young and old male prototypes to simulate shape changes (top right), color changes (bottom left), and combined shape and color changes (bottom right) associated with aging. Several points are worth noting. First, the shape and color transformations applied separately both induce an increase in apparent age. Therefore, the increased age of the face subjected to the combined transform (bottom right) is not due solely to one of these factors. Second, the transform is global in that it affects the hair and all facial features simultaneously. Third, the identity of the face is maintained across the transformation. Fourth, the perceived change in age between the original and the "shape and color" transformed face is less than the 25-year difference between the prototypes. This is probably because the blending process used when creating prototypes reduces some of the textual cues to age such as wrinkles. Such textual changes may not be so important when altering other attributes, such as gender.

Figure 6. Applying shape and color transformations for age. (a) Original face, (b) shape transformed: original face + 100 percent shape difference between young male (40 individuals from collection 1, 25 to 29 years old) and older male prototypes (20 individuals from collection 1, 50 to 54 years old), (c) color transformed: original face + 100 percent color difference between young and old prototypes, (d) shape and color transformed: original face + 100 percent shape and color difference between young and old prototypes.

Discussion

The technique described here may seem similar to morphing, since both can change an object's appearance gradually. The similarity between techniques is superficial, however, because morphing does not preserve identity—it simply changes one object into another. Our transformation processes are designed to maintain identity and selectively modify perceived attributes. Furthermore, morphing requires knowledge of both endpoints prior to computation, whereas the transforms described here (once created) can be applied many times to a variety of faces, each time creating something novel. Moreover, our process is predictive and thus can be used to create an entirely new appearance of an object. For example, age transformations can start with a single image of a person and predict future or previous appearance.

We can think of the prototypes as each defining a point in a multidimensional "face space" and together defining an axis (of gender). The transformational techniques described here utilize such axes and shift specific faces along them (thus altering the apparent gender). Note that any two axes need not be orthogonal. In the example given in Figure 5, the prototypes were chosen to manipulate gender, but shifting a source face along this axis may also alter other perceptual attributions such as attractiveness.

Performing the transformations in 2D is fast and convenient, but the procedures can be easily extended to 3D. It may be possible to create prototypical texture maps and 3D meshes, and to carry out the transformations in a conceptually identical way. As noted in the introduction, the processes underlying the formation of prototypes and the construction of transformations can be performed for any homogeneous class of objects. ◆

Acknowledgments

References

1. H. Yamada, et al., "A Facial Image Processing System for Psychological Studies," *Proc. IEEE Int'l Workshop on Robot Human Communication,* IEEE, Piscataway, N.J., 1992, pp. 358-362.

2. D.I. Perrett, K.A. May, and S. Yoshikawa, "Facial Shape and Judgements of Female Attractiveness," *Nature,* Vol. 368, 1994, pp. 239-242.

3. S.E. Brennan, "The Caricature Generator," *Leonardo,* Vol. 18, 1985, pp. 170-178.

4. P.J. Benson, and D.I. Perrett, "Synthesizing Continuous-Tone Caricatures," *Image and Vision Computing,* Vol. 9, 1991, pp. 123-129.

5. D. Knuth, "The Concept of a Meta-Font," *Visible Language,* Vol. 16, 1982, pp. 3-27.

6. D.R. Hofstadter, *Metamagical Themas: Questing for the Essence of Mind and Pattern,* Penguin Books, London, 1986, pp. 232-296.

7. C. Kang, Y. Chen, and W. Hsu, "An Automatic Approach to Mapping a Lifelike 2.5D Human Face," *Image and Vision Computing,* Vol. 12, 1994, pp. 5-14.

8. S. Edelman, D. Reisfeld, and Y. Yeshurun, "Learning to Recognize Faces from Examples," *Lecture Notes in Computer Science,* Vol. 588, 1992, pp. 787-791.

9. F.J. Galton, "Composite Portraits," *Anthropological Institute of Great Britain and Ireland,* Vol. 8, 1878, pp. 132-142.

10. P.J. Benson, and D.I. Perrett, "Extracting Prototypical Facial Images from Exemplars," *Perception,* Vol. 22, 1993, pp. 257-262.

11. J.D. Foley, et al., *Computer Graphics Principles and Practice,* 2nd Edition, Addison-Wesley, Reading, Mass., 1990, pp. 592-295.

Duncan Rowland is a doctoral student in the School of Psychology at the University of St. Andrews, where he is employing computer graphics techniques in his study of visual perception. Rowland received a BSc in computer science from Hull University in 1991.

David Perrett lectures in visual perception at the University of St. Andrews. His research interests include brain mechanisms in object recognition. Perrett received a BSc in Psychology from St. Andrews University in 1976 and a D.Phil from Oxford University in 1981.

Readers may contact the authors at the University of St. Andrews, School of Psychology, St. Andrews, Fife, Scotland, KY16 9JU, e-mail {dr, dp}@st-and.ac.uk, website http://www.st-and.ac.uk/~www_sa/academic/imaging/index.html.