First appears in *Imagina 97 - CONFERENCES - ACTES/PROCEEDINGS, Feb, Monte-Carlo, (1997), pp159-175.*
Later in *A Collection of Technical Publications, ATR-HIPS, Hikaridai, Kyoto 619-02, Japan, (1997), pp 453-469.*
(This is a pre-publication version)

Transforming Facial Images in 2 and 3-D

Rowland, D.A., Perrett, D.I., Burt, D.M., Lee, K J., and Akamatsu, S.*,

Psychological Laboratory, University of St Andrews, Fife, Scotland, KY16 9JU
*ATR HIPS, 2-2 Hikaridai, Seika-Cho, Soraku-gun,Kyoto 619-02,Japan

Abstract

We review image processing techniques for manipulating 2-D representations of facial images and show their potential for extension to 3-D. Techniques for the manipulation of 2-D images include those used for: (i) averaging images of different faces to form facial ‘prototypes’, (ii) automated caricature exaggeration of the way an individual face differs in shape and colour from a prototype and (iii) transformations manipulating perceived facial qualities while maintaining information about identity (e.g. modifying the apparent sex or age of a face). Each technique relies on 160 feature landmarks allocated to the original 2-D images (e.g. centre of left eye, tip of nose). These landmarks can also be found in the 3-D colour and depth maps obtained from Cyber-scans. Using 2-D image warping techniques to align colour and depth maps from different faces allows prototypes and caricatures and may similarly allow transformations of faces to be performed in 3-D.

Introduction

In this report we review the techniques that have previously been implemented to manipulate 2-D images of faces. These include warping and morphing and how they can be used in: the formation of a facial prototype (an average appearance of a group of faces, Galton 1879; Benson and Perrett 1994); a caricature exaggeration of the distinctive shape and colour of individual faces (Brennan 1985; Benson and Perrett 1991a,b; Lee and Perrett 1996); and transformations of face properties (Rowland and Perrett 1995; Burt and Perrett 1995). We show that each technique can be extended for use with 3-D image data.

Morphing

The basic technique of morphing involves 3 main stages: the location of similar feature points on two source images; the 'warping' of these images to intermediary shapes; the weighted blending of these warped images.

We will, briefly, describe a basic image warping algorithm (see Wolberg 1990 for a fuller discussion which can be used to implement the later defined transformational techniques. Figure 1 shows a facial image which has a standard set of feature points manually positioned to delineate the major facial landmarks (following conventions previously defined, notably by Brennan 1985). Further feature points are added around the perimeter of the image in order to form a limit for the warping. The feature points are used to tessellate the image surface into a set of triangles (Figure 2). A constrained Delaunay triangulation algorithm has been used to force maximum area triangles with all the feature lines included as triangle edges (Shewchuck 1995).

Each location (P) within a triangle can be specified using vectors (V_1 and V_2) defined by two different sides of its enclosing triangle (ABC).

$P=\lambda_1 V_1+\lambda_2 V_2$ where $\lambda_1>=0, \lambda_2>=0, \lambda_1+\lambda_2<=1, V_1=\overrightarrow{AB}$ and $V_2=\overrightarrow{AC}$

To warp an image it is thus simply the case of moving the defining feature points into the desired shape (making a target set of feature points) and then for each pixel in the generated image:

1) calculate values for λ_1 and λ_2 using the enclosing triangle and the target set of feature points;
2) look up the RGB value from the source image using λ_1 and λ_2 in the corresponding triangle. (Pixel interpolation maybe performed in order to avoid aliasing).

To calculate an image which is n% along the morph between image A and image B (where a value of n=0 would yield image A and n=100 would yield image B), we first calculate a set of feature points n% between the feature points from the delineated source images. Image A and image B are then warped to this same intermediary shape and blended together with the weighting specified by n (e.g. n=70% would be a blend with 70% of the warped image A and 30% of the warped image B).

Forming facial prototypes

A prototype can be defined as being a representation of the consistencies across a collection of faces. For example, a prototypical male Caucasian face (Figure 3) would contain all that is consistent about Caucasian faces and can be generated by calculating a mean face from a set of Caucasian faces. In our derivation of prototypes we have used images from a variety of sources. Some collections of faces have been specifically captured under standard lighting, with adornments such as jewellery/make-up removed. Other collections have been gleaned from magazines, and various other less constrained sources.

To derive the prototypical shape for a group of faces, the delineation data for each face are first "normalised", making the faces nominally of the same size and orientation. The left and right pupil centres provide convenient landmark points for this process. The first step is to calculate the average left and right eye positions for the whole population. The next step is to apply a uniform translation, scaling, and rotation to the (x, y) positions of all the feature points, thus normalising 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 standardises face size and alignment. We can now calculate the average positions of each remaining template point (after alignment), the resulting data constitutes the mean shape for the given population.

The prototypical colour information can be obtained by warping each face in the set to the population mean shape. These images are then blended with equal weighting. Since each face has been aligned for all features to coincide across the normalised faces, this resultant blend contains the population mean colouration as well as the population mean face shape.

The technique of blending images together is essentially the same as that pioneered by Galton over a hundred years ago. Galton (1878, 1979) combined faces photographically using multiple exposure of film. He manipulated the orientation and size of component photographs so that the centre of the left and right eyes were constant across the set of faces. Galton’s technique has been replicated digitally (Langlois and Roggman 1990; Burson 1992). These digital versions, however, align component images on only 2 image points (the eye centres). With the technique described here, the shape of the component images is modified (to average) so that the images are aligned not just on the eye centres but on each of the 160 landmarks defined manually.

As Galton noted, the features of the component faces that are consistent across the group can be visualised in the blend or prototype. Prototypes formed by this process have been used to study the facial characteristics related to age (Burt and Perrett 1994), gender (Brown and Perrett 1995), beauty (Langlois and Roggman 1990; Perrett et al 1994) and artistic style (Perrett, Grogan, Dodd, et al. unpublished studies).

Caricaturing
Once a prototype has been formed for a collection of faces it is possible to generate caricatures by accentuating the difference between an individual face and a relevant prototype. After normalising the feature location data from the prototype to the eye positions of an example face, all feature points on the example face can be shifted away from their counterparts on the prototypical face by a given percentage. This percentage is the amount of caricature and can be thought of as extrapolating a morph between a prototype and the example face. (If the percentage is -100% then the product of the manipulation will be the prototype, if the percentage is -50% then the result will be halfway along the morph between the prototype and the example face, if the percentage is 0% then the example face is returned, if it is +50% then a caricature of the original face is the result.

Recognition of face images (of photographic quality) is improved by caricaturing facial shape, though the improvement has so far only been apparent with highly familiar faces (Rhodes et al 1987; Benson & Perrett, 1991b; 1993 and unpublished studies). The caricature technique is useful for psychological studies because it allows the systematic and quantitative manipulation of facial traits. For example, the degree of fear in a face can be modified in steps from a neutral to normal expression and beyond, to a caricature exaggeration of an expression. This quantitative manipulation of expression allowed correlation with the degree of neural activity in different brain regions that results from viewing fearful expressions (Morris et al 1996).

Caricaturing intensity and colour information can be performed in an analogous way to that for manipulating shape. To caricature colour the difference in colour between corresponding pixels in the original image and the relevant prototype is exaggerated. To do this the prototype is first re-mapped to the shape of the original target image. Then the difference in red, green and blue values of corresponding pixels is calculated and a percentage of this difference added to each of the original pixels. Caricaturing intensity and colour information has been found to facilitate human recognition of famous faces in perception time paradigms where the face images are presented briefly (Lee and Perrett 1996). Caricaturing intensity information can also improve performance of artificial face recognition systems (e.g. Craw et al 1996).

Transformation based on prototypes
Information typical to a class of face (e.g. young female) is maintained in prototypes, but any information relating to specific identity is lost (since the prototype has many individuals constituting its make up). The difference between two prototypes can thus be used to manipulate apparent class membership for faces while maintaining identity characteristics. That is, one can change the apparent gender of a female facial image adding the difference in shape and colour between male and female prototypes to an original face (Rowland and Perrett 1995).

To perform the transform, we first normalise 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. We can then add a percentage of this vector difference to each corresponding feature point from the source face. The source face is then warped into this new shape and the shape transform is completed.

The shape transformation is limited in impact because a lot of gender information is missing, namely the colour differences between male and female faces. We can perform a similar transformation in colour space. Figure 4 shows the impact of a colour 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 colour difference between prototypes at each pixel is then calculated and subsequently added to the corresponding value from the original image.

Combining these two transformations (see Figure 4d) is more effective than either shape or colour manipulations alone (Figures 4b and 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 colour difference.

Psychological investigations of visual cues to age shows that the prototyping process extracts information relevant to perception and that this can be applied in the transformational process to increase or decrease perceived age in a realistic manner (Burt and Perrett, 1995).

3-D representations of faces

The 3-D data from the face is collected using a Cyber-scanner. This device moves a camera around a stationary head. To minimise distortions the centre of the head needs to be positioned at the origin of the camera's rotation. The vertical line of pixels recorded at the centre of the camera image is stored for 512 equal intervals during rotation. Given a vertical resolution of the camera of 512, this yields a 512 square 'anamorphic' image which is a projection of the head surface onto a surrounding cylinder. The 512 square image is recorded with a colour resolution of 24 bits (8 bits, red, green and blue). It can be delineated (see Figure 5) in the same way as normal facial images.

The head has 3 degrees of freedom for translation and 3 degrees of freedom for rotation. During scanning, head position and posture may not be precisely aligned with the scanner centre. Moreover posture varies from person to person. Normalisation can be performed on the 3-D data for each head using 3 facial landmarks (e.g. centre of the two eyes and the centre of the mouth). This is equivalent to translating, rotating and scaling all the heads (as virtual objects) so that the 3 landmarks align. For the purposes of this paper the scan data are treated as normalised.

As the scanner rotates, the contour shape (and depth profile) is recorded at 512 intervals spaced equally around the 360 degrees of revolution. Again the vertical resolution of the recording camera is 512, yielding a 512 square depth map with depth information stored as a 16 bit number.

3-D Prototypes

Prototypes are formed in part using procedures equivalent to those described above for 2-D face images. First, the colour maps for each face in a set need to be delineated. The average shape of the 2-D delineation data is then calculated and each original image colour map warped into this average shape. The aligned colour maps are then blended together to yield the average colour map (Figure 6, middle, right). The depth maps for each face in the set are aligned to the average shape of the 2-D delineation data and blended in the same way to yield the average depth map (Figure 6, middle, left). The average depth information can be reconstructed as a 3-D virtual object with depth expressed as the distance from the centre of the head (or centre of the scanner). Finally, the average colour information can be texture rendered onto the surface of the 3-D virtual head to synthesise a 3-D prototype (Figure 6, bottom).

3-D Caricatures

These are made in two stages. The initial steps follow the procedures described above for exaggerating the shape information of the 2-D colour representations. The first stage proceeds by calculating the difference in shape between the delineation structure of the target face and that of the chosen prototype. Here, shape refers to the x-y position on the cylindrical projection of the face. Shape differences are then amplified by the desired percentage. The depth and colour maps for the target face and depth map of the prototype are then warped to the caricatured x-y shape. The depth map for the target face is now compared to the depth map of the prototype (in the new caricatured x-y shape). Differences in depth at each pixel in the 512 by 512 array are exaggerated by the same percentage. This yields a 3-D caricatured depth map that has deviations from the prototype caricatured in x,y and z (Figure 7). Since the colour map information from the re-mapped target face (warped into the caricatured x-y shape) lies in register with the new 3-D caricatured depth map, this colour information can be texture-mapped onto the 3-D caricatured shape to create a texture-rendered 3-D caricature.

These procedures caricature only the 3-D shape information. The colour information can, of course, be caricatured as well, by defining the difference in red, green and blue intensity values between corresponding pixels between the original and prototype and adding a percentage of this colour difference back to the original image. This can be done only when the x-y shapes of the original and prototype lie in register. If colour differences are to be calculated after shape caricaturing, then the colour map of the prototype needs to be warped to the caricatured x-y shape of the individual. Texture rendering employing both the 3-D caricatured shape and the enhanced colour map yields a caricature which exaggerates all information available from the scan data.

Transformations in 3-D

As noted above, prototypes can be used to as a basis for changing the apparent characteristics of a face while maintaining identity information (Rowland and Perrett 1995; Burt and Perrett 1995). To perform these operations in 3-D, a two stage process is needed. The initial operations are equivalent to those described for 2-D. First, the difference in 2-D delineation shape for the 2 prototypes is calculated. A percentage of this is then added or subtracted from the shape of the delineation for the target face. To modify the apparent gender of a face, for example, to induce a more masculine appearance to a female face, the difference between male and female prototypes is added to the original shape of the target face. To increase the femininity of the target female face, a percentage of the difference between male and female prototypes is subtracted. Once the new x-y shape of the target face is calculated, the colour and depth maps for the original target face and the prototypes are all warped to this shape.

This establishes the desired shape transformation in only x and y. To calculate the shape transformation in the depth plane (z), the difference in depth maps between the male and female prototypes (aligned to the new x-y shape) is found. A percentage of this difference is then added or subtracted from the corresponding pixels in the depth map of the target face (aligned to new x-y shape). This establishes a 'sex transform' transformation in 3-D shape for the target face, producing a head shape with enhanced masculine or feminine structure but maintaining the identity or family characteristics of the individual.

Acknowledgements
This work was sponsored by project grants from Unilever Research, the Economic and Social Research Council (grant no. R000234003) and an invited researcher collaboration with ATR HIPS, Kyoto, Japan. We thank Keith May for help developing the shape manipulations and Oda-san and Isono-san for help developing the extensions to 3-D.

Figures

The Perception Lab,
University of St Andrews,
Fife,
Scotland,
KY16 9JU.

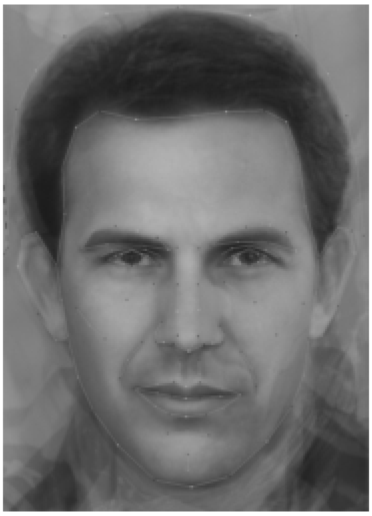

Figure 1. Delineation of a features in a 2-D face image. 160 feature points have been manually defined. Feature points are joined to form a line drawn representation.

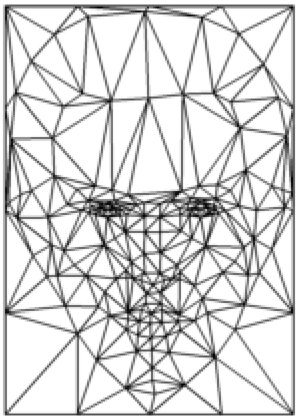

Figure 2. Constrained Delaunay tessellation of a face. The triangles are those with the shortest sides given the constraint that no triangle can bisect a line feature present in Figure 1.

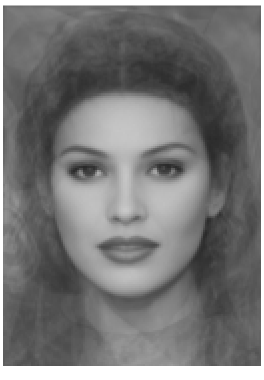

Figure 3. 2-D prototype of a Caucasian female. The image is the prototype made from 60 Caucasian females.

Figure 4. Sex Swap. Applying shape and colour transformations for gender differences. a) Original face, b) shape transformed, c) colour transformed, d) shape and colour transformed.

Figure 5. Delineation of the facial features from a 3-D facial scan. The colour information is stored as a 512 by 512 image with 24 bit red, green, blue information per pixel. x and y dimensions of the image correspond to the cylindrical co-ordinates of the scan (x, the angular position around the 360 degrees of the scan and y, the height).

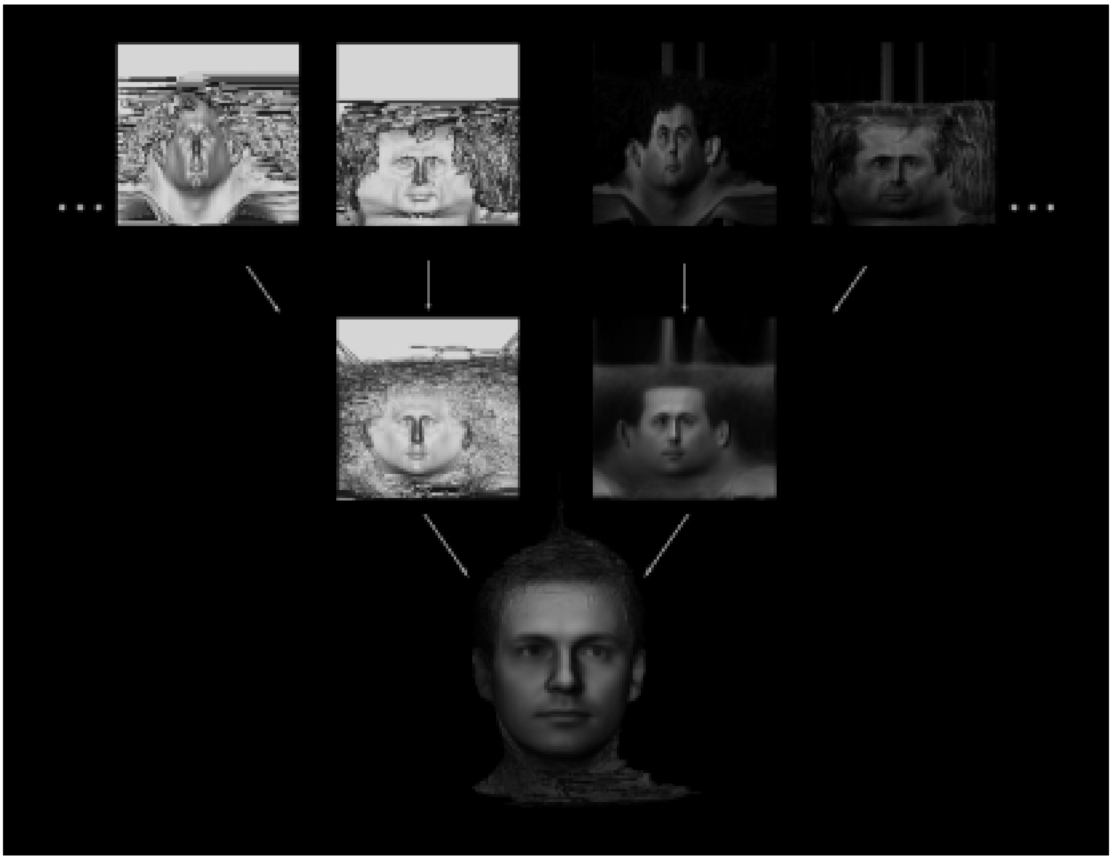

Figure 6. Forming a 3-D facial prototype. Upper left: the depth maps for 2 faces from a set of 20 male Caucasian faces. Upper right: the corresponding colour maps for the same faces. Mid right: the average colour map formed from the 20 component colour maps. Each original colour map was delineated and the average 2-D shape of the delineation data defined. The original colour maps were then aligned by warping them the average 2-D delineation shape and the aligned colour maps were finally blended together. Mid left: the average depth map formed by warping the 20 component depth maps into the average 2-D delineation shape and then averaging the depth values across the aligned maps at each pixel. Bottom: 3-D prototype for the set combining average depth and colour maps. The depth map is now expressed as radial distance from the centre of the head and the colour has been texture mapped on to the surface.

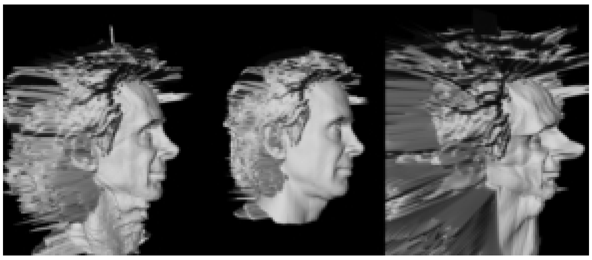

Figure 7. 3-D Caricature of identity. Centre: Original face. Right: 3-D shape caricatured by 50% against and the Caucasian male prototype. Left: 3-D caricature of identity against an inappropriate race (Japanese male) prototype.

References

Benson PJ, Perrett DI, 1991a "Synthesising continuous-tone caricatures" *Image and Vision Computing* **9** 123-129

Benson P J, Perrett D I, 1991b "Perception and recognition of photographic quality facial caricatures: implications for the recognition of natural images" *European Journal of Psychology* **3** 105-135

Benson P J, Perrett D I, 1993 "Extracting prototypical facial images from exemplars" *Perception* **22** 257-262

Benson P J, Perrett D I, 1994 "Visual processing of facial distinctiveness" *Perception* **23** 75-93

Brennan S E, 1985 "The caricature generator" *Leonardo* **18** 170-178

Brown E, Perrett D I, 1993 "What gives a face its gender?" *Perception* **22** 829-840

Burson, N, 1992 *Faces* (New York: William Morrow & Co.)

Burt D M, Perrett D I, 1995 "Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour information" *Proceedings of the Royal Society of London, Series B* **259** 137-143

Craw I, Costen N, Kato T, Robertson G, Akamatsu S, 1996 "Automatic face recognition: combining configuration and texture" (in press)

Galton F J, 1878 "Composite portraits" *Nature* **18** 97-100

Galton F, 1879 "Composite portraits, made by combining those of many different persons into a single, resultant figure" *Journal of the Anthropological Institute* **8** 132-144

Langlois J H, Roggman L A, 1990 "Attractive faces are only average" *Psychological Science* **1** 115-121

Lee K, Perrett D I, 1996 "Perception time measures of the effects of manipulations in colour space on discrimination of famous faces" (in submission)

Morris J S, Frith C D, Perrett D I, Rowland D, Young A W, Calder A J, Dolan RJ, 1996 "A differential neural response in the human amygdala to fearful and happy facial expressions" *Nature* , **383** 6603 812-815

Perrett D I, May K, Yoshikawa S, 1994 "Attractive characteristics of female faces: preference for non-average shape" *Nature* **368** 239-242

Rhodes G, Brennan S E, Carey S, 1987 "Identification and ratings of caricatures: implications for mental representations of faces" *Cognitive Psychology* **19** 473-497

Rowland D A, Perrett D I, 1995 "Manipulating facial appearance through shape and colour" *IEEE Computer Graphics and Applications* **15** 70-76

Shewchuck J R, 1995 “Delaunay triangulator” *Unpublished Manuscript.* Comp Science, Carnegie Mellon.

Wolberg, G, 1990 “Digital Image Warping” :ISBN 0-8186-8944-7