MOVING FACIAL IMAGE TRANSFORMATIONS BASED ON STATIC 2D PROTOTYPES

Bernard Tiddeman and David Perrett

Perception Laboratory,
Department of Psychology
University of St Andrews,
St Andrews, Fife KY16 9JU
UK

{bpt,dp}@st-and.ac.uk   http://www.perceptionlab.com/

ABSTRACT

This paper describes a new method for creating visually realistic moving facial image sequences that retain an actor's personality (individuality, expression and characteristic movements) while altering the facial appearance along a certain specified facial dimension. We combine two existing technologies, facial feature tracking and facial image transformation, to create the sequences. Examples are given of transforming the apparent age, race and gender of a face. We also create 'virtual cartoons' by transforming image sequences into the style of famous artists. The results show that static 2D face models can be used to create realistic transformations of sequences that include changes in pose, expression and mouth shape.

Keywords: facial image transformation, facial feature tracking, image processing.

1. INTRODUCTION

The use of computer graphics for altering the apparent age, sex or race of static facial images is not only entertaining but has found application in psychology, medicine, forensics and education. The extension of these methods to moving facial images has the potential to enhance and expand these applications. In this paper we show that an existing facial transformation method based on static 2D prototype ('average') facial images, combined with a suitable face tracking method, can produce excellent results. In addition, we extend the range of transformations to include various artistic styles, allowing the construction of 'virtual cartoons'.

2. PREVIOUS WORK

The synthesis and animation of facial images has long been of interest in computer graphics research. Early face models used geometrical methods [Parke82] [Duffy88] [Thalm89], which have been extended to include physics-based models of facial tissues [Terzo93] [Koch96] [Lee00] and statistical models of normal face variation [DeCar98] [Blanz99] [Tidde99] [Vette97]. The animation of geometrical models can be performed by morphing between predefined expression components [Parke82] [Duffy88] [Thalm89] [Pighn98]. For physics-based face models, animation can be performed using numerical simulation of muscle actions [Water87][Lee95].

Many face-tracking algorithms have been devised including those based on optical flow constraints [Mase91][DeCar00], active shape models (ASM) [Baumb96][Edwar98] or energy minimising point tracking techniques [Lucas81] [Lien00].

The reconstruction of tracked facial feature movements by 'virtual actors' has application in video telecommunication because of the potential for low-bandwidth communication [Choi91] [Choi94]. This kind of technology has also been used to lip-synch computer graphic animations with an actor's voice and movements for film entertainment or virtual avatars [Berge85] [Willi90] [Bregl97] [Essa96] [Guent98][Ezzat00]. The animated characters are

either a direct clone of the original (in the case of low-bandwidth communication) or are designed by a 3D-computer artist or computer algorithm. In this work, instead of employing a computer artist, we use facial transformations defined as the differences between populations of facial images, from groups of real individuals or sets of portraits in a particular artistic style. These can be used, for example, to alter the age or sex of an actor's face while retaining their personality, expression and typical movements.

3. METHOD

DEFINING FACIAL PROTOTYPES

In this work facial transformations are defined as the differences between two 'prototype' facial images e.g. a male prototype and a female prototype. Each prototype is constructed by averaging a set of facial images in terms of 2D shape, pixel colour [Bensn91] [Bensn93] and multiscale texture [Tidde00]. The shape of each face in the set is delineated with 179 points located along contours around the major facial features (eyes, nose and mouth) and the facial border. The average shape is found by averaging the position of each delineated point across the set i.e.

$$\overline{\mathbf{x}} = \frac{1}{N} \sum_{i=0}^{N} \mathbf{x}_i \qquad (1)$$

where $\mathbf{x}_i$ is the ith shape vector made from the x and y coordinates of the n delineated face points i.e.

$$\mathbf{x}_i = \left(x_0^i, y_0^i, x_1^i, y_1^i, \cdots, x_n^i, y_n^i \right)$$
$$(2)$$

and $\overline{\mathbf{x}}$ is the mean shape vector of N delineated faces.

The colour of each pixel in the prototype image is found by warping each component image into the average shape and calculating the mean colour i.e.

$$\overline{\mathbf{c}}(x, y) = \frac{1}{N} \sum_{i=0}^{N} \mathbf{c}_i \left(W_x^i(x, y), W_y^i(x, y) \right) \quad (3)$$

where $\mathbf{c}_i(x, y)$ is the red, green and blue (RGB) colour vector of image i at point (x,y), (W_x^i, W_y^i) is the translation vector given by the warping function for image i and $\overline{\mathbf{c}}$ is the mean colour over the N

images. In this work we use linear warping over a Delaunay triangulation of the feature points, which can be combined with a one-to-one constraint algorithm to prevent the warped image folding [Tidde01].

The process given above leads to prototype images, which although realistic, do not always represent the underlying sample in terms of texture [Burt95]. By texture here we mean the intensity difference between nearby pixels e.g. fine wrinkles or the brush strokes of a particular artist. This is because the warping process only aligns the large scale facial features that have been delineated, whereas the fine details become blurred. To correct for this effect we also find the average magnitude of the intensity variations at multiple spatial scales by decomposing the sample facial images into a multiscale wavelet basis. For wavelet sub-band j of image i (w_i^j) we find the smoothed magnitude image (m_i^j) given by

$$m_i^j = H * \left| w_i^j \right|,$$
$$(4)$$

where H is a cubic B-spline smoothing filter and $*$ is the convolution operator. The mean of the smoothed magnitude images ($\overline{m}^j$) across the sample gives a measure of the average edge strength at pixel (x,y) and is given by

$$\overline{m}^j(x, y) = \frac{1}{N} \sum_{i=0}^{N} m_i^j(x, y). \qquad (5)$$

The wavelet coefficients in the 'untextured' (i.e. shape and colour only) prototype are amplified (in a spatially varying manner) according to

$$\overline{w}'^j(x, y) = \frac{\overline{w}^j(x, y)\overline{m}^j(x, y)}{n^j(x, y)} \qquad (6)$$

where

$$n^j = H * \left| \overline{w}^j \right| \qquad (7)$$

is the smoothed magnitude of the wavelets of the untextured prototype, and $\overline{w}'^j$ is the textured prototype's wavelet component. The textured prototype is reconstructed from these amplified wavelet components and the low-pass residual of the untextured prototype. In this work we have used a redundant wavelet representation with an exact reconstruction formula. This process leads to

prototypes that better represent the underlying sample in terms of texture. Example prototypes used in this paper are shown in Figure 1.
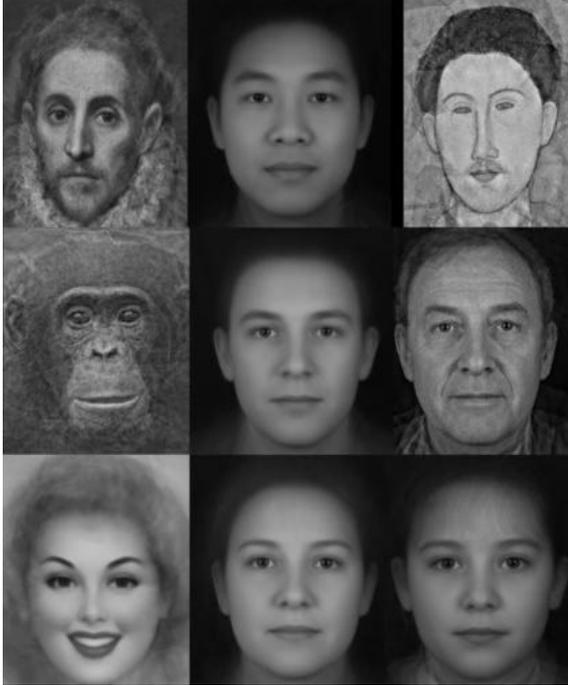
Figure 1. Example facial prototypes. From top left: El Greco adult male (textured), East-Asian adult male, Modigliani adult male (textured), Chimpanzee (textured), European adult male, European older male (textured), 1950's female 'pin-up', European adult female, European female child.

APPLYING FACIAL TRANSFORMATIONS

The facial prototypes described above define the typical differences between two sets of images. These differences can be applied to a delineated individual static face image or individual face frame from a sequence in five steps [Rowla95] (Figure 2). The first step normalises the position of the two prototypes to the subject. In this work we use a least squares rigid body fit (translation, scaling and rotation) [Arun87]. Secondly the translation of each feature point from the source prototype (usually of the same class as the image undergoing the transform) to the destination prototype is calculated, scaled if desired, and added to the subject's facial feature point, i.e.

$$\mathbf{x}' = \mathbf{x} + \boldsymbol{a}(\mathbf{d} - \mathbf{s}) \qquad (8)$$

where $\mathbf{x}$ is the original shape vector, $\mathbf{x}'$ is the transformed shape vector, $\mathbf{d}$ and $\mathbf{s}$ are the destination and source prototype shapes respectively and $\boldsymbol{a}$ is the shape change scale factor.

This defines the target shape for the transform. The subject's image and the source and destination prototype images are all then warped into this new shape. In order to prevent the transformation being swamped by a very bright or highly saturated prototype a colour normalisation step can be included. The average brightness and saturation of the prototypes (across only face points) are shifted to be the same as the subject's. Finally the colour shift of each pixel from the source prototype to the destination prototype is calculated, scaled if desired and added to the colour of the subject's pixel i.e.

$$n(\mathbf{t}) = o\left(\mathbf{W}_{\mathbf{t}}^{o}\right) + \boldsymbol{a}\left[d\left(\mathbf{W}_{\mathbf{t}}^{d}\right) - s\left(\mathbf{W}_{\mathbf{t}}^{s}\right)\right] \qquad (9)$$

where $n(\mathbf{t})$ is the (new) transformed image at pixel $\mathbf{t}=(x,y)$, o is the original subject's image, d and s are the destination and source prototype images respectively and $\mathbf{W}_{\mathbf{t}}^{k}$ is the position vector given by the warping function $\mathbf{W}$ for image k to the new shape at point $\mathbf{t}$.

This completes a shape and colour transformation of the subject image. This process is similar to image morphing [Beier92][Lee96] but the identity of the subject does not change through the sequence. An additional texture transformation can be included, but in this paper additional texture is introduced by performing a shape and colour transformation using a textured prototype as the destination image when appropriate (e.g. when adding wrinkles to age a face).
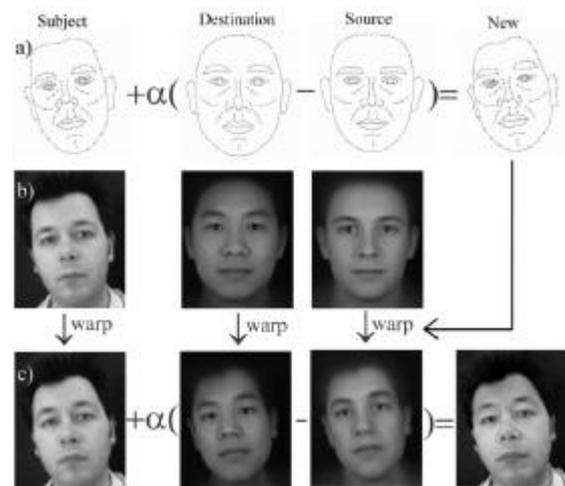

Figure 2. An illustration of the transformation process: (a) Define new shape, (b) warp subject and prototypes into new shape and (c) transform colours at each pixel. The prototype shape and colour normalisation steps are not shown.

TRACKING FACIAL FEATURES

The steps above can be applied to every frame of a moving sequence to construct a moving facial transform. Manual delineation of each frame of a sequence is not only tedious and time consuming, but is likely to cause the sequence to jitter because of small variations in the delineation from frame to frame. A better solution is to use one of the many tracking algorithms now available.

In this work we use an active shape model (ASM) [Coote95][Lanit97] based face tracking method [Baumb96][Edwar98]. ASMs need to be trained with a large and varied database containing faces of a variety of people with different poses and expressions. Alternatively, an ASM can be built for tracking an individual sequence by using a subset of the frames from that sequence. This requires quite a large amount of manual delineation but it has the advantage that the tracking is robust against apparent changes in identity (which can also be included in a general ASM [Edwar98]). The delineated facial data can be added to a general face tracking ASM as more subjects are filmed.

Figure 3 shows selected frames from three example sequences. The tracking copes well with changes in pose and expression, including mouth shape changes and eye blink.


Figure 3. Frames from three face tracking examples.

4. RESULTS

The following results show a number of transformations of the three tracked sequences of Figure 3. The images illustrate how the sequence maintains the identity, expression and pose of the individual but is transformed along a single facial parameter i.e. age, race, gender or art-style. Figure 4 shows frames from the first sequence transformed across race (from Ethnically European to Ethnically East-Asian), species (from human to chimpanzee) and into the artistic style of El Greco.

Figure 5 shows the second sequence transformed into an Ethnically East-Asian, an older man and into the style of the artist Modigliani. Comparison of the race transformations between Figure 4 and Figure 5 illustrates how identity is maintained across the transformation. The ageing transformation of Figure 5 illustrates how wrinkles e.g. forehead lines behave, moving in an apparently natural manner, driven by the other nearby facial features such as the eyebrows and mouth.

Figure 6 shows three transformations of the third sequence of Figure 3: a gender transform (female to male), an age transform (adult to child) and an art transform (female photograph to 1950's female 'pin-up' style). This final transform demonstrates the effect of using prototypes with different expressions. The changing expressions from the sequence are still clearly visible, but have been added to the open mouth smile of the pin-up prototype (Figure 1).
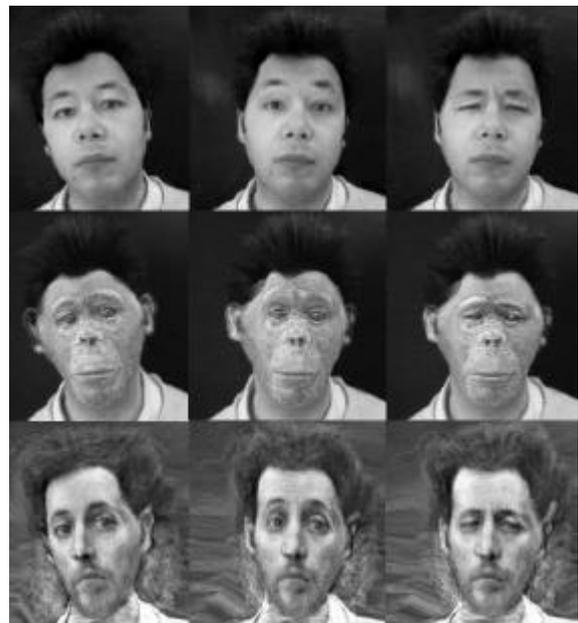

Figure 4. Example transformations of the first sequence of Figure 3. Top row: Ethnic East-Asian, Centre row: Chimpanzee (75% transform), Bottom row: El Greco's portrait style.
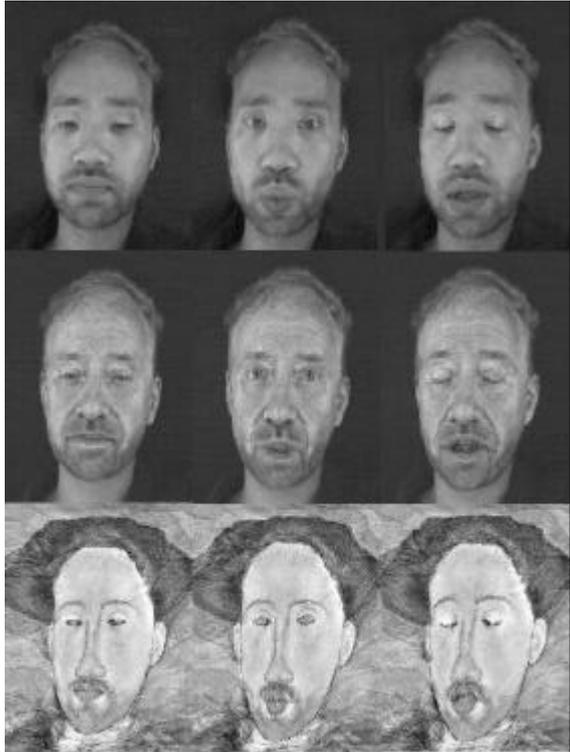
Figure 5. Example transformations of the second sequence of Figure 3. Top row: Ethnic East-Asian, Centre row: aged, Bottom row: Modigliani's portrait style.

Figure 6. Example transformations of the third sequence of Figure 3. Top row: Male, Centre row: child, Bottom row: 1950's pin-up style.

5. CONCLUSION

The results above show that moving facial transformations, based on static 2D prototypes combined with active shape model based tracking, can achieve realistic results. The transformations can accommodate small head rotations (approx 15º) and changing mouth and eye shapes (e.g. open and closed). In addition, the properties added to the subject during the transformation e.g. wrinkles, appear to move naturally, even though those features are not delineated or tracked. The technique described in this paper has potential for application in the film industry, for example for the virtual ageing of actors instead of makeup, and for creating cartoons in a particular artistic style from footage of real actors. They also have potential for application in psychological research, for example for investigating the effects of age, gender or race on subconscious stereotypic attitudes.

REFERENCES

[Parke82] Parke F.I.: Parameterised models for facial animation, *IEEE Computer Graphics Applications*, Vol. 2, No. 9, pp61-68, 1982.

[Duffy88] Duffy N.D. and Yau J.F.S: Facial image reconstruction and manipulation from measurements obtained using a structured lighting technique, *Pattern Recognition Letters*, Vol. 7, pp 239-243, 1988.

[Thalm89] Magneneat-Thalmann N., Minh H., Angelis M. and Thalmann D.: Design, transformation and animation of human faces, *Visual Computer,* Vol. 5, pp32-39, 1989.

[Terzo93] Terzopoulos D. and Waters K.: Analysis and synthesis of facial image sequences using physical and anatomical models, *IEEE Trans. on PAMI*, Vol. 15, No. 6, pp569-579, 1993.

[Koch96] Koch R.M., Gross M.H., Carls F.R., von Büren D.F., Frankhauser G. and Parish Y.I.H.: Simulating facial surgery using finite element methods, in *SIGGRAPH99 Conference Proceedings*, pp421-428, 1996.

[Lee00] Lee W-S., Magneneat-Thalmann N.: Fast head modelling for animation, *Image and*

Vision Computing, Vol. 18, No. 4, pp355-364, 2000.

[DeCar98] DeCarlo D., Metaxas D. and Stone M.: An anthropometric face model using variational techniques, in SIGGRAPH98 Conference Proceedings, pp67-74, 1998.

[Blanz99] Blanz V. and Vetter T.: A morphable model for the synthesis of 3d faces, In *SIGGRAPH99 Conference Proceedings*, pp187-194, 1999.

[Tidde99] Tiddeman B., Rabey G. and Duffy N.: Synthesis and transformation of three-dimensional facial images, *IEEE Engineering in Medicine and Biology*, Vol. 18, No. 6 pp 64-69, 1999.

[Vette97] Vetter T., Jones M.J. and Poggio T.: A bootstrapping algorithm for learning linear models of object classes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, Puerto Rico, USA, IEEE Computer Society Press, 1997.

[Pighn98] Pighin F., Hecker J., Lischinski D, Szeliski R. and Salesin D.: Synthesizing realistic facial expressions from photographs, in *SIGGRAPH98 Conference Proceedings*, pp75-84, 1998.

[Water87] Waters K.: A muscle model for animating three-dimensional facial expression, *in SIGGRAPH87 Conference Proceedings*, pp17-24, 1987.

[Lee95] Lee Y., Terzopoulos D. and Waters K.: Realistic face modelling for animation, in *SIGGRAPH95 Conference Proceedings*, pp55-62, 1995.

[Mase91] Mase K. and Pentland A.: Lipreading by optical flow. *Systems and Computer*s, Vol. 22, No. 6, pp67–76, 1991.

[DeCar00] DeCarlos D. and Metaxas D.: Optical flow constraints on deformable models with applications to face tracking, *International Journal of Computer Vision*, Vol. 38, No. 2, pp99-127, 2000.

[Edwar98] Edwards G.J., Taylor C.J. and Cootes T.F.: Learning to identify and track faces in image sequences, *International Conference on Face and Gesture Recognition*, pp260-265, 1998.

[Baumb96] Baumberg A and Hogg D: Generating spatiotemporal models from examples, *Image and Vision Computing*, Vol 14, No. 8, pp525-532, 1996.

[Lucas81] Lucas B. and Kanade T.: An interactive image registration technique with an application in stereo vision. In *The 7th International Joint Conference on Artificial Intelligenc*e, pp674–679, 1981.

[Lien00] Lien J.J.J., Kanade T., Cohn PJF and Li C-C.: Detection, tracking and classification of action units in facial expression, *Journal of Robotics and Autonomous Systems*, Vol. 31, No. 3, pp131-146, 2000.

[Choi91] Choi C.S., Okazaki, Harashima H. and Takeba T.: A system of analyzing and synthesizing facial images, in *Proc. IEEE Int. Symposium of Circuits and Systems (ISCAS91)*, pp2665-2668, 1991.

[Choi94] Choi C.S., Aizawa K., Harashima H. and Takebe T.: Analysis and synthesis of facial image sequences in model based image coding, *IEEE Trans. on Circuits and systems for video technology*, Vol. 4, No. 3, pp257-275, 1994.

[Berge85] Bergeron P. and Lachapelle P.: Controlling facial expressions and body movements in the computer-generated short 'Tony de Peltrie', in *SIGGRAPH85 Conference Proceedings*, 1985.

[Willi90] Williams L.: Performance driven facial animation, in *SIGGRAPH90 Conference Proceedings*, pp235-242, 1990.

[Bregl97] Bregler C., Covell M. and Slaney M.: Video rewrite: driving visual speech with audio, in *SIGGRAPH97 Conference Proceedings*, pp353-360, 1997.

[Essa96] Essa I., Basu S., Darrell T. and Pentland A.: Modelling, tracking and interactive animation of faces and heads using input from video, in *Computer Animation Conference*, pp 68-79, 1996.

[Guent98] Guenter B., Grimm C., Wolf D., Malvar H. and Pighin F.: Making faces, in *SIGGRAPH98 Conference Proceedings*, pp55-66, 1998.

[Ezzat00] Ezzat T. and Poggio T.: Visual speech synthesis by morphing visemes, *International Journal of Computer Vision*, Vol. 38, No. 1, pp45-57, 2000.

[Bensn91] Benson P.J. and Perrett D.I.: Synthesizing continuous-tone caricatures, *Image and Vision Computing*, Vol. 9, pp123-129, 1991.

[Bensn93] Benson P.J. and Perrett D.I.: Extracting prototypical facial images from exemplars, *Perception*, Vol. 22, pp257-262, 1993.

[Tidde00] Tiddeman B.P. Perrett D.I. and Burt D.M.: A wavelet based method for prototyping and transforming the textural detail of images, submitted work preprint available from http://psy.st-and.ac.uk/people/personal/bpt/publications.html.

[Tidde01] Tiddeman B., Duffy N. and Rabey G.: A General Method for Overlap Control in Image Warping, *Computers and Graphics*, Vol. 25, No 1, to be published February 2001.

[Burt95] Burt D.M. and Perrett D.I.: Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour

information, *Proceedings of the Royal Society of London B*, Vol. 259, pp 137-143, 1995.

[Rowla95] Rowland D. A. and Perrett D. I.: Manipulating Facial Appearance through Shape and Color, *IEEE Computer Graphics and Applications,* Vol. 15, No. 5, pp70-76, 1995.

[Arun87] Arun K.S., Huang T.S. and Blostein S.D.: Least-squares fitting of two 3-D point sets, *IEEE Trans. on PAMI*, Vol. 9, No. 5, pp698-700, 1987.

[Beier92] Beier T. and Neely S.: Feature based image metamorphosis, in *SIGGRAPH92 Conference Proceedings,* pp35-42, 1992.

[Lee95] Lee S-Y., Chwa K-Y., Shin S.Y. and Wolberg G.: Image metamorphosis using snakes and free-form deformations, in SIGGRAPH95 Conference Proceedings, pp439-448, 1995.

[Coote95] Cootes T., Taylor C. Cooper D. and Graham J.: Active shape models - their training and application, *Computer Vision, Graphics and Image Understanding*, Vol. 61, No. 1, pp38-59, 1995.

[Lanit97] Lanitis A., Taylor C.J. and Cootes T.F.: Automatic interpretation and coding of face images using flexible models, *IEEE Trans. on PAMI*, Vol. 19, No. 7, pp743-756, 1997.