Example-Based Shape from Shading: 3D Heads form 2D Face Images

A.Z. Kouzani, F. He, and K. Sammut
School of Engineering, Flinders University of SA
GPO Box 2100, Adelaide, SA 5001, Australia
e-mail: egazk, egfh, egks@flinders.edu.au

Abstract— **Images taken of human faces vary in perspective depending on the view-point of the observer. Few of the face recognition methods reported in the literature are capable of recognising faces under varying pose. To improve the ability of face recognition systems to recognise faces under varying pose, a novel method is proposed in this paper that takes a face image of arbitrary pose and synthesises a standard frontal-pose image of the same face.**

I. Introduction

One of the key problems in automatic face recognition concerns the variability in a person's appearance that is present in an image due to changes in pose. The input face image of a person presented to a face recognition system is not usually taken from an identical view-point as that used for the face image of the same person in the database. Therefore, it is important that an automatic face recognition system is able to cope with possible view-point variations between different images taken of the same face.

Several of the methods that have been reported in the literature, have demonstrated some ability to reduce their sensitivity to variations in pose. The basic principle of these methods is to synthesise a front-view of the face image so that a direct comparison of the input face image with the corresponding image contained in the database could become feasible. The existing methods can be classified into two groups: example-based and 3D model-based.

The example-based methods use 2D views of prototype faces that sample different rotations out of the image plane. Different algorithms have been proposed for using the information in the prototype views to synthesise new views of a person. Ullman and Basri [14] have shown that a 2D view of an object under rigid 3D transformation can be obtained as a linear combination of a small set of 2D example views, where the 2D view representation is a vector of locations of a set of feature points. Poggio and Vetter [11] have discussed this linear-combination approach in the case where only one example view is available for an object. They employ the idea of using prior knowledge of the object class to generate virtual views. Two types of prior knowledge have been explored: knowledge of 3D object symmetry, and example images of objects of the same class. Beymer and Poggio [5] have used an automated labelling algorithm that computes the correspondence between every pixel of the two images rather than just for a hand-selected subset of feature points. Jones and Poggio [9] have employed a stochastic gradient descent algorithm to match the models of faces.

The 3D model-based methods use a 3D model of a head to predict the appearance of a face under different pose. The aspect of synthesising face images using 3D head models has been explored in [1]-[2], [6], [12]. In the 3D head model technique, a face shape is represented either by a polygonal model or by a more complicated multi-layer mesh that simulates the face tissue. Once a 2D face image texture is mapped onto the 3D head model, the face can be treated as a traditional 3D object.

Atick et al., [4] use linear combinations of 3D data (without correspondence) taken from a Cyberware scanner to build a 3D head model. Principal Components Analysis (PCA) has been used to derive a low-dimensional parameterisation of the head space. By assuming that an individual's head can be described using, for instance, 200 coefficients obtained from a projection of the head to the head space, a minimisation technique can then be used to estimate these coefficients. In each iteration of the minimisation process, the coefficients are estimated, the head is built, a 2D face image is rendered, and an error function of the distance between the rendered face image and the 3D head is calculated.

The hybrid example-model-based algorithm proposed in this paper is based on Atick et al.'s method, but differs from theirs in a number of ways; in particular, it does not require optimisation. A full description of the algorithm is provided in the next section.

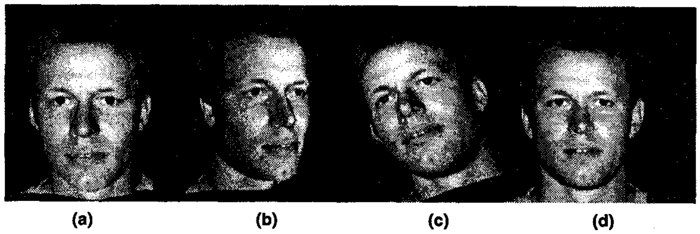

Fig. 1. Samples of 2D face images, taken from different view-points (a)-(c), that are rendered from a 3D head and the head's texture (d).

II. Method

Most algorithms, which attempt to extract shape-from-shading, are designed for images of arbitrary objects with smooth brightness variations [8]. These algorithms estimate shapes, present in a space which possesses a large number of degrees of freedom, from the limited information contained within an image. Therefore, the performance of these algorithms are restricted in their practical application. Since human faces are the only objects of interest in face recognition, the shape-from-shading method can be carried out more efficiently if the knowledge about the face class is taken into account.

One way to incorporate the knowledge about the face class into the shape-from-shading method is to utilise an ensemble of shapes of human heads. Human head shape is consistent across people. The structure is invariably the same; all humans have eyes, noses, mouths, foreheads, and cheeks. The variability from one head to another is relatively small. Nevertheless, it is these small structural deviations that give a face its unique identity.

In the proposed approach, a database of 3D human heads, including 2D textures and 3D surfaces, are utilised for the construction of individual 3D heads from 2D face images. The head database consists of sixty-three 3D head models generated from stereo images obtained using the C3D system at The Turing Institute. Figure 1 shows samples of 2D face images taken from different view-points that are rendered from a 3D head and the head's texture.

A. Pose Estimation

The first step of synthesising a frontal view of a face from an arbitrary view is to estimate the pose of the face. Fuzzy logic and neural networks have been used to develop a face detection module [10] in which a face is located within an image of a scene. The face detection module also provides additional information concerning the size, location, and pose of each identified component of the detected face. The face detection module can handle large variations in face images due to changes in view-point of up to $\pm 45°$ about the up/down, left/right, and frontal directions. The system classifies the face image into one of $11 \times 11 \times 11$ possible poses. More detailed discussions of the face detection module are given in [10].

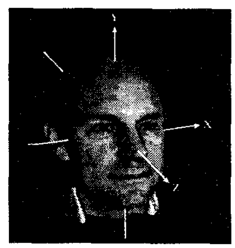

Fig. 2. The 3D head space.

B. Recovery of the 3D Structure of the Face

Using the information provided by the face detection module, the recovery of the 3D structure of the face is then performed. A total of 1331 2D full-face images of different views, comprising combinations of 11 rotations about each of the X, Y, and Z directions, are rendered by rotating the head model. Figure 2 shows the 3D head space. This rendering process produces face images with up to $\pm 45°$ up/down, left/right, and frontal directions of rotation (with a resolution of $9°$). Figure 3 displays 121 view-points of a person's face rotated about the X and Y axes. For each of these 121 view-points, 11 images are rendered as a result of rotation about the Z axis. Figure 4 shows 11 images of a person's face rotated about the Z axis. The images in both Figures 3 and 4 are all resized to equal sized square

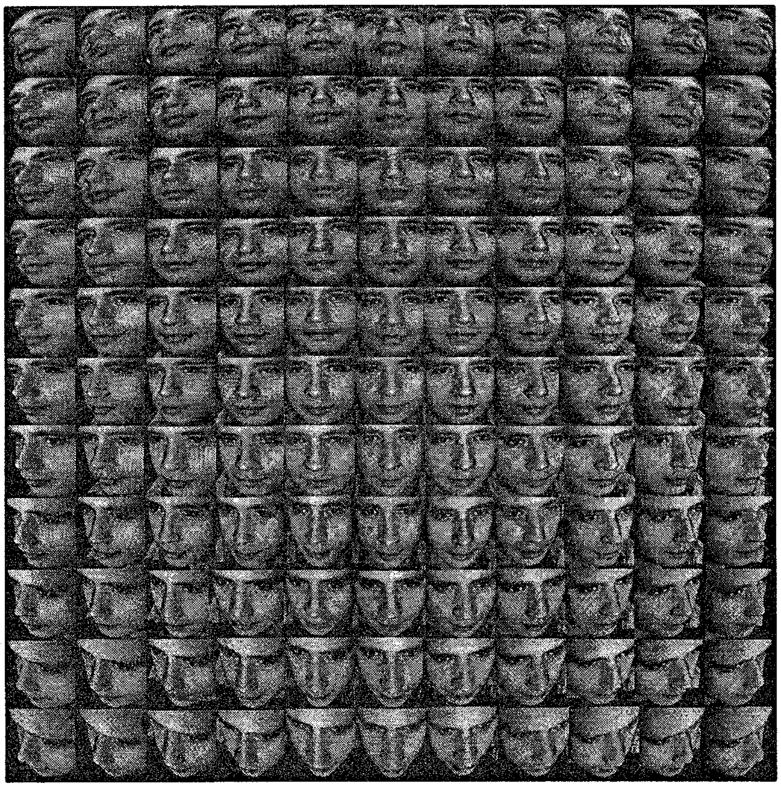

Fig. 3. 121 view-points of a person's face rotated about the X and Y axes.

images to allow better visualisation.

Next, PCA is utilised for the creation of a face space for each of the 1331 poses. In PCA, a dimensionality reducing linear projection that maximises the scatter of all projected samples is selected. The optimal basis is given by the eigenvectors of the total scatter matrix. Let the training set of n-dimensional face images be $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_N\}$. The average face of the set is defined by $\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{I}_i$. An $n \times m$ $(m < n)$ matrix B with orthonormal columns is sought to best describe the distribution of the data. If the total scatter matrix S_T is defined as

$$S_T = \sum_{i=1}^{N} (\boldsymbol{I}_i - \boldsymbol{\mu})(\boldsymbol{I}_i - \boldsymbol{\mu})^T, \quad (1)$$

then after applying the transformation B^T, the scatter of the transformed vectors is $B^T S_T B$. B contains m eigenvectors of S_T corresponding to the m largest eigenvalues.

Prior to the extraction of the 1331 face spaces for different views, the following procedure is performed.

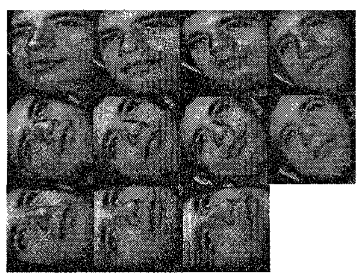

Fig. 4. Face images rendered from a 3D head with a fixed view-point along the X and Y axes, but rotated about the Z axis.

The size of each rendered image is set to 128×128. A reference image is chosen for each of the 1331 set of images. Within each set of face images, a *morphological transformation* [15] is performed on the reference image representing the set, and on the rest of images within the set. This transformation can be described as a point-to-point and area-to-area transition of one

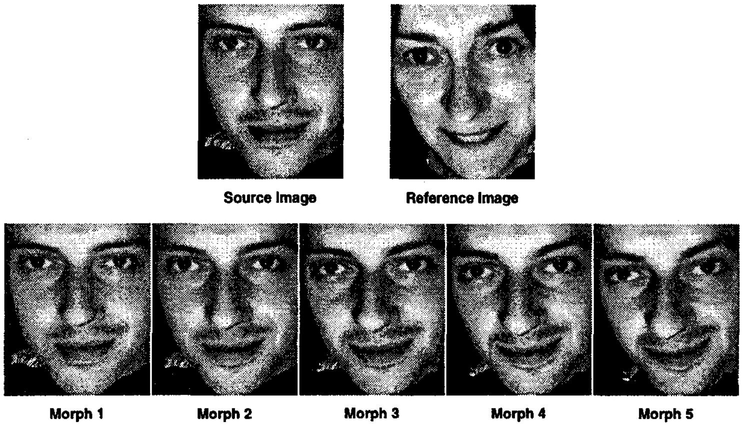

Fig. 5. Morphological transformation of two face images.

image into another. When one face transforms into another face, the transformation of the outline and components (nose, left eye, etc.) of the initial face to their corresponding outline and components of the destination face is also set. Figure 5 displays a morphological transformation of a source face image and a reference face image.

As stated earlier, the face detection module provides the pose information of the detected face, together with the location and size of each face component (i.e., left eye, right eye, nose, and mouth), to the stage where the 3D structure of the face is recovered. To adjust the scale, orientation, and position of the input face with that of the faces within the relevant pose set, the input face image is transformed so that the centres of the the face components in the input image and those in the reference image of the relevant pose set coincide with each other.

An *optical flow* algorithm is then employed to perform a pixel-wise correspondence between the detected face image $\boldsymbol{D}$ and the mean image $\boldsymbol{R}$ of the associated pose. The mean face is defined by

$$\boldsymbol{R} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{I}_i \tag{2}$$

where $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_N\}$ are N face images of the face set under consideration. A hierarchical correlation-based optical flow algorithm [3] is employed to determine the correspondence vectors. The algorithm is based on a *Laplacian pyramid* and a coarse-to-fine Sum-of-Squared-Difference (SSD) based matching strategy. This algorithm can handle the presence of large disparities in the images. The SSD measure is defined as

$$\begin{aligned} S(\boldsymbol{x},\boldsymbol{d}) \;=\; & \sum_{j=-n}^{n}\sum_{i=-n}^{n} \boldsymbol{W}(i,j)(\boldsymbol{R}(\boldsymbol{x}+(i,j)) - \\ & \boldsymbol{D}(\boldsymbol{x}+(i,j)))^2 \end{aligned} \tag{3}$$

where $\boldsymbol{W}(i,j)$ denotes a weighting function and $\boldsymbol{d}$ is restricted to the square neighbourhood of size $(2n+1)^2$ centred at $\boldsymbol{x}$. At the coarsest level, the correct displacements are assumed to be one spatial unit. SSD minima are first located to integer accuracy within small image regions. Sub-pixel displacements are then computed by finding the minimum of a quadratic approximation of the SSD surface about the integer location which best maximises $S(\boldsymbol{x},\boldsymbol{d})$.

Figure 6 displays the correspondence vector of two face images. The correspondence vector defines a position in the correspondence space in which image $\boldsymbol{D}$ should be located. A normalised detected face image $\boldsymbol{D_N}$ is obtained by warping image $\boldsymbol{D}$ along its correspondence vector.

Image $\boldsymbol{D_N}$ is then projected onto the associated face space to obtain a set of approximation coefficients as

$$\boldsymbol{a} = \boldsymbol{B}^T(\boldsymbol{D_N} - \boldsymbol{R}) \tag{4}$$

where $\boldsymbol{B}$ contains the eigenvectors of the total scatter matrix (see (1)).

The ensembles of 3D surfaces are manually normalised to adjust scales, orientations, and positions of heads. This is done by marking the key-points on the surface and the centre of the head, and by rotating, scaling, and translating each head to match the marked point between the head being normalised

Fig. 6. Correspondence vector of two face images.

and the reference head. PCA is again used to derive a dimensionally reduced presentation of surfaces. The corresponding coefficients, namely $\boldsymbol{s}$, are then calculated as for $\boldsymbol{a}$ in (4).

Having parametrised the space of head surfaces, an individual head can be constructed by specifying the coefficients, $\boldsymbol{s}$. The basic principle here is that the projection of the face image $\boldsymbol{D}_N$ onto the face space yields the same coefficients as the projection of its corresponding 3D surface onto the surface space. After projecting the face image onto the face space, the coefficients obtained are then used for the calculation of the 3D surface. A linear combination of the basis vectors that form the surface space produces the surface of the face image under consideration. Having calculated the 3D surface and the 2D face image, the 3D head is reconstructed using texture mapping [7].

C. Synthesis of the Front View Face

Once a 3D head is constructed from a 2D face image, 2D face images can be rendered under different pose and lighting effects. For face recognition, a standard front-view face image can be rendered because the known face database contains front-view images. Figure 7 displays a 2D face image, and three face images rendered from its calculated 3D head.

III. Results

In order to assess the performance of the proposed system, the following experiment is carried out. A test database of 30 people, collected by Bernard Achermann at the University of Bern, is used for this study. In this database, 10 images with variations of up to $\pm 45°$ of the head positions (facing the camera, right, left, down, and up) are taken for each person. Each image contains one well-illuminated face with a uniform background. Some face images may contain spectacles, moustache, and/or beard. The total number of face images is 300. By rotating each image randomly within the range of $\pm 45°$ twice, 600 extra images are produced and added to the data set, resulting in a total of 900 images. Figure 8 illustrates sample face images from this database.

From this database, 30 front-view images are copied into a known face database containing 236 face images of 236 different people under standard pose, thereby increasing the known face database to 266 front-view face images. The eigenfaces method [13] is then trained on the images of the known face database.

The images of the test database is presented to the eigenfaces method in two steps. In the first step, the face images of the test database are presented to the eigenfaces method without processing them. In the second step, however, the face images are presented to the proposed pose correction system, and the synthesised front-view face images are presented to the eigenfaces method. Table I displays the recognition results obtained from this experiment.

TABLE I
Recognition results.

Method	Recognition Rate
Eigenfaces	36.7%
Proposed	98.2%

As indicated by Table I, the proposed pose correction system significantly increases the performance of a standard face recognition system.

IV. Conclusions

A method for synthesising a front-view face image from the image of the same face with an arbitrary pose, has been presented in this paper. Using a 3D head database, 2D face images of different views are rendered. The face images are allocated to different sets; each set represents a particular pose. A ref-

(a) (b) (c) (d)

Fig. 7. A 2D face image (a), and the three face images rendered from its calculated 3D head (b)-(d).

erence image is chosen for each set of face images. Morphological transformations are performed on the reference image and the images of each set. PCA is utilised for the creation of a face space for each set of face images. The ensembles of 3D surfaces are manually normalised to adjust the scales, orientations, and positions of heads. PCA is then used for deriving a dimensionally reduced presentation of surfaces. A pixel-wise correspondence between the input face image and the mean image of the associated pose is obtained. The input face image is projected onto the associated face space, and the obtained coefficients are used for the calculation of the 3D surface. A linear combination of the basis vectors that form the surface space produces the surface of the face image under consideration. Finally, the 3D head is reconstructed using texture mapping. Once the 3D head is constructed, a front-view 2D face image can be rendered using computer graphics techniques.

Simulation results reveal that the recognition success rate for the proposed methods is higher than that for the traditional eigenfaces method.

Fig. 8. Sample face images from the test database.

Acknowledgements

The first author would like to thank DEETYA, Australia, and FUSA, for providing him with scholarships to undertake this research. The authors would also like to acknowledge the support of CSSIP.

References

[1] A.C. Aitchison and I. Craw, "Synthetic images of faces - an approach to model-based face recognition", in *Proc. British Machine Vision Conference*, 1991, pp. 226–232.

[2] T. Akimoto, Y. Suenaga, and R.S. Walace, "Automatic creation of 3D facial models", *IEEE Computer Graphics and Applications*, vol. 13, no. 5, pp. 16–22, 1993.

[3] P. Anandan, "A computational framework and an algorithm for the measurment of visual motion", *IJCV*, , no. 2, pp. 283–310, 1989.

[4] J.J. Atick, P.A. Griffin, and A.N. Redlich, "Statistical approach to shape from shading: Reconstruction of 3D face surfaces from single 2D images", *Neural Computation*, vol. 8, no. 6, pp. 1321–1340, 1996.

[5] D. Beymer and T. Poggio, "Face recognition from one example view", Tech. Rep. A.I. 1536, Artificial Intelligence Lab., MIT, 1995.

[6] I.A. Essa and A. Pentland, "A vision system for observing and extracting facial action parameters", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Seattle, WA, 1994, pp. 76–83.

[7] James D. Foley, *Computer Graphics : Principles and Practice*, Addison-Wesley, 2 edition, 1990.

[8] B.K.P. Horn and J. Brooks, Eds., *Shape from Shading*, MIT Press, Cambridge, Mass., 1989.

[9] M.J. Jones and T. Poggio, "Model-based matching by linear combinations of prototypes", Tech. Rep. A.I. 1583, Artificial Intelligence Lab., MIT, November 1996.

[10] A.Z. Kouzani, F. He, and K. Sammut, "Commonsense knowledge representation and reasoning and its application to face detection", *International Journal of Advanced Computational Intelligence*, vol. 2, no. 3, pp. 1–8, 1998.

[11] T. Poggio and T. Vetter, "Recognition and structure from one 2D model view: Observations on prototypes, object classes, and symmetries", Tech. Rep. A.I. 1347, Artificial Intelligence Lab., MIT, 1992.

[12] D. Terzopoulos and K. Waters, "Analysis of facial images using pysical and anatomical models", in *Proc. Int. Conf. on Computer Vision*, Osaka, Japan, December 1990, pp. 727–732.

[13] M. Turk and A. Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[14] S. Ullman and R. Basri, "Recognition by linear combinations of models", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 13, no. 10, pp. 992–1006, 1991.

[15] G. Wolberg, *Digital Image Warping*, IEEE Computer Society Press, Los Alamitos, CA, 1990.