Face Pose Estimation From a Face Ellipse

Qiang Ji
Department of Computer Science
University of Nevada at Reno
Reno, NV89557 U.S.A
qiangji@cs.unr.edu

Rong Hu
Department of Computer Science
University of Nevada at Reno
Reno, NV89557 U.S.A
rong@cs.unr.edu

Abstract

In this paper, we describe a new approach for estimating 3D pose of a human face from a monocular image. We assume that the shape of a face can be approximated by an ellipse and the ratio of the major to minor axis of the ellipse is given. We first detect a face ellipse and then estimate the face pose. Compared with the existing feature-based approaches for face pose estimation, our approach is more robust since an ellipse can be more reliably detected and embodies more compact information about 3D pose. Experimental study using a large number of synthetic and real images demonstrates that the approach is accurate and robust.

1 Introduction

Face pose determination represents an important research area in human computer interaction (HIC). There are many approaches of face pose estimation have been reported. Most of them model a face with certain facial features. In [4], six facial feature points including pupils, nostrils and lip corners are used to model a face. In [5], five feature points containing four eye corners and the tip of noise are used to model a face. Gee [1] proposed a facial model based on the ratios of lengths between some facial features. These feature-based approaches face a major challenge in detecting the required facial features under varying illuminations and different head orientations.

In this paper, we propose a novel model-based approach for estimating the 3D pose of a human face from a monocular image. Our approach models the shape of a face with an ellipse. The use of an ellipse has the following advantages: First, human face can be rather accurately modeled with an ellipse. Second, ellipses can be preserved under projective transformations. Third, ellipses contain compact global information of the face, they are expected to be more robust to noise than facial feature points. Fourth, the 2D/3D correspondences can be established much more easily.

In [3], a technique is introduced to reconstruct a 3D conic and its pose from two images obtained from different views. To overcome the problem that a single ellipse image is not sufficient to recover 3D face pose, we assume that the ratio of the major to minor axis of the 3D face ellipse is given and that the face pose rotation movement is limited to two angles: tilt (around vertical axis) and slant (around horizontal axis). Both assumptions are reasonable. The ratio between major and minor of axes of a face ellipse can be obtained from the face image without any slant and tilt. The rotation around the Z axis (optical axis) is rare for many applications in HCI.

2 Mathematical Model

Let $[X\ Y\ Z]^T$ represent the coordinates in the object frame, $(u\ v)^T$ be the coordinates in the image frame, $[X_c\ Y_c\ Z_c]^T$ be the coordinates in the camera frame, and $[c\ r]^T$ be the coordinates in the row-column frame(see Fig 1).

2.1 Projection equation of points

Between row-column frame and object frame, we have

$$\lambda \begin{pmatrix} c \\ r \\ 1 \end{pmatrix} = \mathbf{WM} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \tag{1}$$

where λ is a scalar factor, $\mathbf{W}$ is the camera intrinsic matrix, $\mathbf{M}$ is camera extrinsic matrix, and We have

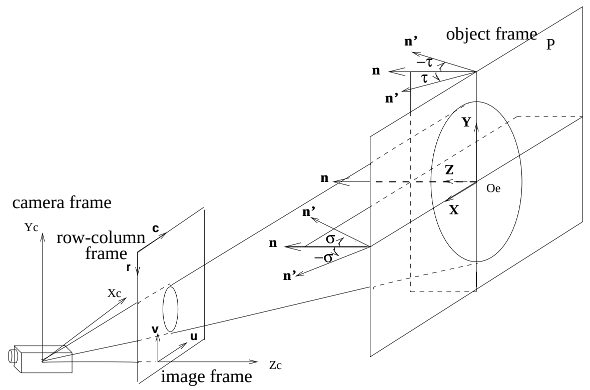


Figure 1. Camera perspective projection model

$\mathbf{M} = [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{T}]$, and

$$\mathbf{W} = \begin{pmatrix} s_x f & 0 & u_0 \\ 0 & s_y f & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

s_x, s_y is scale factor (pixels/mm) in image u, v axis, f is the camera focal length, u_0, v_0 is the coordinates of the principle point.

2.2 Projection equation of ellipses

Let Q be a 3×3 matrix representing the 3D ellipse in object frame, A be a 3×3 matrix for the image ellipse, we have

$$\begin{pmatrix} c \\ r \\ 1 \end{pmatrix}^T \mathbf{A} \begin{pmatrix} c \\ r \\ 1 \end{pmatrix} = 0 \qquad (2)$$

$$\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}^T \mathbf{Q} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = 0 \qquad (3)$$

Substituting (1) to (2), we have

$$\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}^T \mathbf{M}^T\mathbf{W}^T\mathbf{AWM} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = 0 \qquad (4)$$

From (3) and (4), and denote $\mathbf{W}^T\mathbf{AW} = \mathbf{B}$, this yields

$$\mathbf{Q} = \lambda \mathbf{M}^T\mathbf{BM} \qquad (5)$$

Let the length of major and minor axis of the 3D ellipse be a and b respectively, the equation 5 can be rewritten

$$\lambda \begin{pmatrix} 1/a^2 & 0 & 0 \\ 0 & 1/b^2 & 0 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1^T\mathbf{Br}_1 & \mathbf{r}_1^T\mathbf{Br}_2 & \mathbf{r}_1^T\mathbf{BT} \\ \mathbf{r}_2^T\mathbf{Br}_1 & \mathbf{r}_2^T\mathbf{Br}_2 & \mathbf{r}_2^T\mathbf{BT} \\ \mathbf{T}^T\mathbf{Br}_1 & \mathbf{T}^T\mathbf{Br}_2 & \mathbf{T}^T\mathbf{BT} \end{pmatrix} \qquad (6)$$

Note that $\mathbf{B}$ is known for a calibrated camera. In equation (6), there are six constraints, however, there are nine unknowns: three rotation angles, three translation variables, ellipse major axis length a, minor axis length b, and scale factor λ. To solve these unknowns, additional information is necessary.

2.3 Face pose characterization

We will characterize the 3D face pose by a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{T}$.

2.3.1 The rotation matrix R

We assume there is no rotation around the optical axis of the camera, thus, face orientation can be characterized by two angles: slant and tilt. $\mathbf{R}$ can be obtained by rotating the object frame around Y axis by 180 degree, around X by σ degree, and finally around Y axis again by τ degree. Thus, $\mathbf{R}$ can be expressed as follows,

$$\mathbf{R} = \begin{pmatrix} -\cos\tau & 0 & -\sin\tau \\ \sin\sigma\sin\tau & \cos\sigma & -sin\sigma\cos\tau \\ \cos\sigma\sin\tau & -\sin\sigma & -\cos\sigma\cos\tau \end{pmatrix} \qquad (7)$$

2.3.2 The constraint equations

Let $c = a^2/b^2$, from the first 2×2 sub-matrix in equation (6), two equations are obtained

$$\begin{aligned} \mathbf{r}_1^T\mathbf{Br}_2 &= 0 \\ \mathbf{r}_1^T\mathbf{Br}_1 - c\mathbf{r}_2^T\mathbf{Br}_2 &= 0 \end{aligned} \qquad (8)$$

where $\mathbf{r_1}$ and $\mathbf{r_2}$ are the first and second column of $\mathbf{R}$.

We can solve for slant and tilt from equation 8. Moreover, from equation (6), we have

$$\begin{pmatrix} \mathbf{r}_1^T\mathbf{BT} \\ \mathbf{r}_2^T\mathbf{BT} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad (9)$$

where $\mathbf{T} = (\mathbf{t}_x \quad \mathbf{t}_y \quad \mathbf{t}_z)^T$, $\frac{\mathbf{t_x}}{\mathbf{t_z}}$, $\frac{\mathbf{t_y}}{\mathbf{t_z}}$ can be solved by the above two constraints.

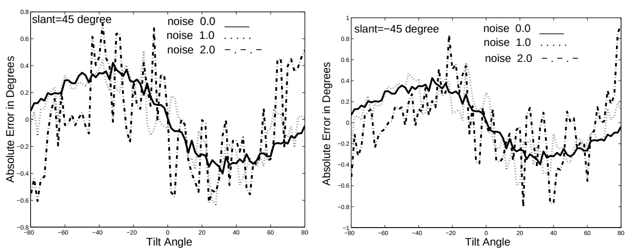
3 Experimental Results

The proposed algorithm has been tested extensively using synthetic ellipse data, actual ellipse data, and human face image data.

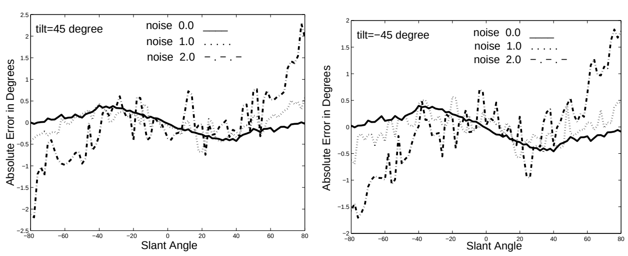
3.1 Experiment results with synthetic ellipse data

It is difficult to obtain the ground truth of the rotation angle of a face. To evaluate our algorithm, we used about 1,153 artificial images of a computer generated ellipse viewed from every possible direction. Some experimental results are shown in Fig.4.

To study the sensitivity of the algorithm to image noise, the imaged ellipse locations were corrupted by zero-mean Gaussian noise with standard deviations 0.5, 1.0, 1.5, 2.0 pixels respectively. Fig.2 shows the estimation errors for slant and tilt respectively .



Pose estimation error with tilt rotations only



Pose Estimation error with slant rotation only

Figure 2. Slant and tilt estimation error for synthetic ellipses

As we can see from Fig2, without Gaussian noise, the maximum estimation error is less than 0.5°. When noise is increased to 2 pixels, the maximum estimation error is about 2°, which demonstrates that the algorithm is fairly robust to noise.

Fig.3 shows the estimation errors of ellipse pose in every possible directions. When slant or tilt is close to -80° or -80°, the head pose is a near-profile pose or a deeply face-up-down pose, estimation error considerably increases with the noise level. This is due to the fact that the ratio is extremely small and any perturbation with the image can lead to a significant increase with the estimated orientations.

3.2 Actual ellipse data

We drew a model ellipse on a planar board. We then obtain images of the model ellipse while it was

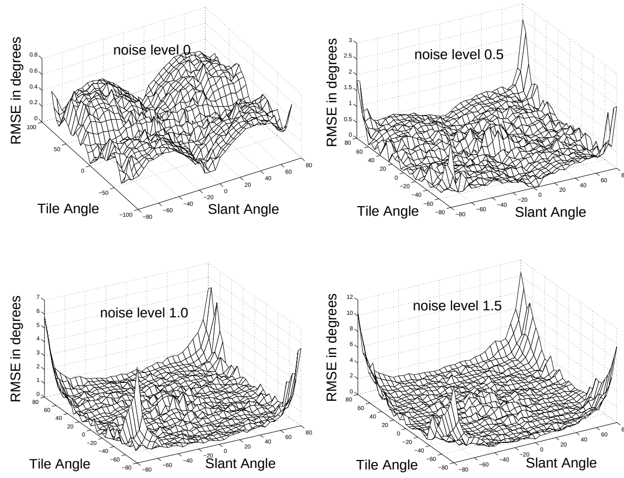


Figure 3. Pose estimation error for synthetic ellipses

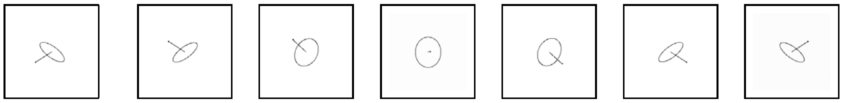
rotated around the X and Y axes respectively. These images are shown in Fig.5. The estimated results with the proposed algorithms are consistent with perceived orientations.

3.3 Human face image data

We captured image sequences of a male with different slants and tilts. An ellipse detection [2] is then performed on each face image to detect face. The detected ellipse is used to estimate the pose of the human face. These results are shown in Fig.6. Through visual inspection, we can conclude the estimated face poses are in good agreement with actual face orientations.

4 Conclusions

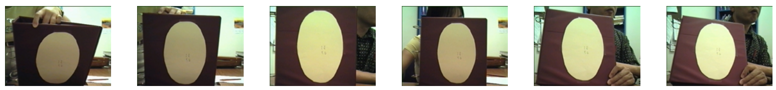
In this paper, we describe a new approach for estimating 3D pose of a human face from a monocular image. The algorithm achieves good experimental results. From synthetic data, without noise, the estimation errors are less than 0.5°, even if Gaussian noise with standard deviation 2.0 pixels is added, the estimate error is only 2.5°. For actual ellipse data and real human face data, the estimated pose results are in good agreement with their perceived orientations. The experimental results demonstrated the proposed facial model is reasonable and that the proposed approach is reliable and stable.



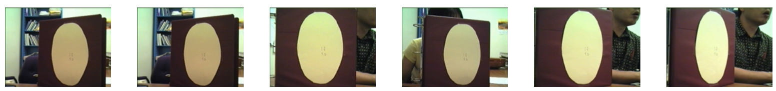
$\sigma: -49.6^\circ$ 49.60° 24.0° 1.8° -24.0° -49.6° 49.6°

$\tau: -50.6^\circ$ -50.6° -26.1° 9.7° 26.1° 50.6° 50.6°

Figure 4. Pose estimation from simulated ellipse. The line in the ellipse represents the estimated ellipse normal. The angle below each image is the estimated slant and tilt.

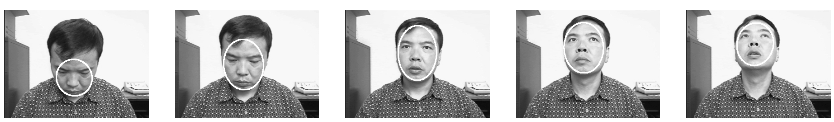


$\sigma: -45^\circ$ -18° -2° 3° 9° 21°



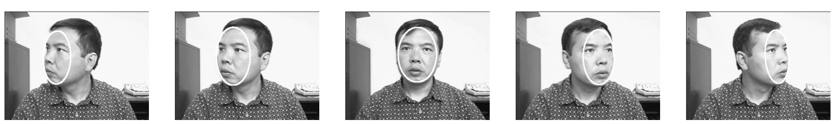
$\tau: -46^\circ$ -20° 11° 22° 27° 40°

Figure 5. The estimation of 3D poses of an actual ellipse with consecutive slant rotation(above) and tilt rotation (below). The angle is the estimated slant and tilt.



$\sigma = -48^\circ$ $\sigma = -28^\circ$ $\sigma = -2^\circ$ $\sigma = 30^\circ$ $\sigma = 42^\circ$

$\tau = -4^\circ$ $\tau = -8^\circ$ $\tau = -8^\circ$ $\tau = -1^\circ$ $\tau = -10^\circ$



$\tau = -49^\circ$ $\tau = -39^\circ$ $\tau = 8^\circ$ $\tau = 39^\circ$ $\tau = 48^\circ$

$\sigma = 11^\circ$ $\sigma = 4^\circ$ $\sigma = -2^\circ$ $\sigma = 5^\circ$ $\sigma = 19^\circ$

Figure 6. The estimation of 3D face pose with consecutive slant rotations (above) and consecutive tilt rotations (below). The angle is the estimated slant and tilt.

References

[1] A. H. Gee and R. Cipolla. Determining the gaze of faces in images. *CUED/F-INFENG/TR 174*, 1994.

[2] Qiang Ji and Robert M. Haralick. A statistically efficient method for ellipse detection, in 1999 ieee international conference on image process ing. *IEEE International Conference on Image Processing, Kobe, Japan*, 1999.

[3] S. D. Ma. Conics-based stereo, motion estimation, and pose determination. *International Journal of Computer Vision*, 10(1):7–25, 1993.

[4] I. R. Stiefelhagen, J. Yang, and A. Waile. A model-based gaze tracking system. 1999.

[5] A. T.Horprasert, Y. Yacoob, and L. S. Davis. Computing 3-d head orientation from a monocular image sequence. 1999.