PERGAMON

Pattern Recognition 34 (2001) 2427–2445

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Pose determination of human faces by using vanishing points

Jian-Gang Wang, Eric Sung*

Division of Control and Instrumentation, School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798, Singapore

Received 29 December 1999; accepted 16 October 2000

Abstract

A new method for estimating 3D-head pose from a monocular image is proposed in this paper. The approach employs general prior knowledge of face structure and the corresponding geometrical constraints provided by the location of vanishing point to determine pose of human faces. The connection of the two far-eye corners and the two neighboring-eye corners, respectively, form the eye-lines. Connecting the two far-mouth corners forms the mouth-line. The eye-lines and the mouth-line are assumed parallel in 3D space, and their vanishing point on the image plane can be used to infer 3D pose of the human face. Perspective projection imaging model is used and an accurate analytic solution of pose (position and orientation) of human face is deduced. The orientation of the facial plane can be obtained when the ratio of the lengths of the eye-line segment (far-eye corners) and the mouth-line segment (far-mouth corners) is known. Furthermore, if one of the two lengths is known, then the absolute positions of the feature corners can be located. The robustness analysis of the algorithm with synthetic data and the experimental results of real face images are enclosed.

Keywords: Real-time face tracking; Pose estimation; Vanishing point; Projective geometry; Ratio; Human–machine interaction

1. Introduction

Being an intrinsically non-intrusive technology, vision-based facial pose estimation is becoming useful for the realization of a natural human–machine interface. Among the existing alternative approaches, the pose determination from a single view is one of the most intuitive and useful schemes for several tasks such as human–machine interfaces, face recognition, tracking, surveillance and virtual reality.

There are two different transformations that may be used for pose estimation from a monocular image: perspective or a sub-class of the affine transformation. The former one precisely models the actual projection of a 3D scene to the image plane. However, the calculation is complex and time consuming, and can deliver up to a fourfold ambiguity in the estimate of the pose. Most real-time systems usually use affine transformation. It is an approximation of the perspective projection as the depth of the object is very small compared with the distance between camera and object. Weak perspective model is not applicable when viewing the face from close range with a short focal length, significant perspective distortion in the image will result.

Thanarat Horprasert et al. [1] employed projective invariance of the cross-ratios of the eye-corners and anthropometric statistics to estimate the head yaw, roll and pitch. The five points, namely the four eye corners and the tip of nose, are used for head pose estimation. The four eye corners are assumed to be co-linear in 3D but this is found not to be exactly true in general. Gee and Cipolla [2] achieved a real-time face tracker by utilizing simple feature trackers searching for the darkest pixel in the search window. Weak-perspective (or scaled orthographic) imaging process is used there. This algorithm, and also some other similar work where weak perspective projection is assumed, has twofold ambiguity because it is based on the well-known three-point model developed by Huttenlocher and Ullman [3]. The

* Corresponding author. Tel.: + 65-790-5419; fax: + 65-792-0415.

E-mail address: eericsung@ntu.edu.sg (E. Sung).

0031-3203/01/$20.00
PII: S0031-3203(00)00154-0

two poses generated by the alignment algorithm are mirror images with respect to a plane parallel to the image plane and the unique solution has to be searched by projecting the two poses back into the image plane and measuring the goodness of the fit. In their earlier work, Gee and Cipolla [4] use five key feature points, nose tip and the far eyes and mouth corners, to estimate the facial orientation under weak perspective. The facial model is based on the ratio of four distances between five key relatively stable features. The ratios are assumed not to change very much for different facial expressions. Ho et al. [5] developed a pose determination algorithm where high constraints are required. Both the distance between the far-eye corners and the distance between far-mouth corners have to be known; however, the distances are not stable from face to face. Azarbayejani et al. [6] developed an interactive graphics system driven by vision input from a single CCD camera. They use extended Kalman filtering to recover the six rigid-body motion parameters of an object from a set of tracked visual feature points. The Hessian criterion is employed to detect feature points that include corners of eyes, pupils, nostrils, etc. Heinzmann and Zelinsky [7] suggested a three-level tracking system. The feature positions are measured using bitmap correlation in hardware and are further corrected using geometric constraints that are implemented with a Kalman filter network. The 2D feature positions are transferred to a 3D model of the feature locations and the pose can be determined finally. Affine projection is assumed in this work too.

The prior knowledge of human face structure, such as symmetry, can be utilized for pose estimation. We found the vanishing point, formed by parallel eye-lines and mouth-line, in the image also provides a clue for pose estimation of human face. Differing from Thanarat Horprasert's method [1], we use four relatively stable feature points. Connecting two far-eye corners and two neighboring-eye corners, respectively, forms eye-lines. Connecting two far-mouth corners forms the mouth-line. In this paper, we study a novel approach that uses vanishing point to derive a new and simple solution for measuring pose of human head from a calibrated monocular view. Vanishing line and vanishing point on image plane provide a cue for inferring 3D information and have been used in computer vision, such as pose estimation [8] camera calibration [9] and 3D shape reconstruction from a line drawing [10]. However, pose determination of human face by using vanishing point is an unexplored approach. The method is based on prior knowledge of face structure, namely that the two eye-lines and mouth-line are parallel in space. The vanishing point, formed on the image of these lines is found and the 3D direction of eye-lines and mouth-line can thus be inferred. An analytic solution of pose of human face is derived. The orientation of facial plane and the relative 3D position of the four corners can be obtained when the ratio of the lengths of the eye-line and mouth-line segments is given. Furthermore, the 3D absolute positions of the four corners can be located if one of the two lengths is given. The ratio of the lengths of the eye and mouth-line segments is verified to be stable from face to face in our experiment. The vanishing point can often be obtained from the image itself by standard techniques such as in Refs. [11,12,13], so making our algorithm practical. When the distance between the facial plane and camera plane is close, our algorithm which applies the full perspective model will result in a more accurate pose than the results obtained in most existing methods that are based on the affine or weak perspective camera assumption. This is because, under weak perspective model, they require the distance to be long enough compared with the depth variance on human face.

The robustness under noise conditions is discussed using synthetic data. Fischler and Bolles [14] revealed that a unique solution cannot be assured for the Perspective Four-Projection problems. The P4P problem has a single theoretical solution [15,16] when the coplanar points are in general configuration (no three colinear scene points, non-colinear image points). Consequently, the solution (include position and orientation) of our algorithm is unique if the ratio of the two lengths and one of the two lengths are given because the corner points we used are coplanar in 3D space, forming the facial plane. In the case only the ratio is assumed, the positional solution is not unique although the orientation is unique. Relative positions of the feature corners can be achieved in this case.

Pose determination using vanishing points is described in Section 2. The 3D positions of the corners and the orientation of the facial plane are discussed in Sections 2.1 and 2.2, respectively. Degenerate situation of the algorithm is analyzed in Section 2.3. Methods for measuring the ratio of the lengths of eye-line and mouth-line segments or lengths themselves are given in Section 2.4. The experimental results and conclusion are reported in Sections 3 and 4, respectively.

2. Pose determination using vanishing points

Pose information consists of position and orientation. The 3D positions (relative or absolute) of the feature corners (far-eye corners and far-mouth corners) are calculated first. Then the orientation of the facial plane can be obtained by taking the cross product of some two vectors formed by connecting two corners. We consider the pose solutions under two cases that the different constraints are set, respectively, i.e. the case when only the ratio is known and the case when both the ratio and one of the two lengths are known. Relative 3D position of the corners can be located for the first case whereas, absolute position can be obtained in the second case. The

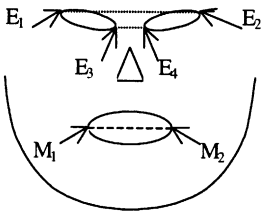


Fig. 1. Feature corners and the corresponding eye-lines and mouth-line.

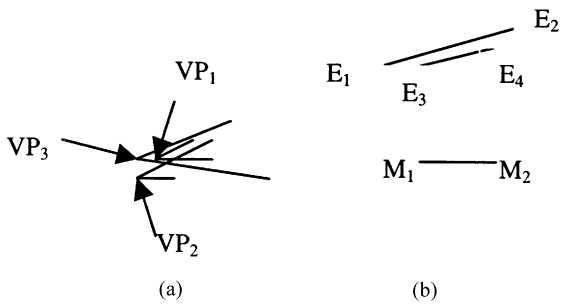


Fig. 2. Image of the eye-lines and the mouth-line: (a) three possible vanishing points; (b) the eye-lines and the mouth-line.

orientation solutions are the same for two case, i.e. obtained using the 3D position. We discuss the position and orientation solutions in following sub-sections, respectively. Some definitions are described first.

We define the two eye-lines. One is obtained by connecting the two far-eye corners (E_1E_2) while the other is obtained by connecting the two neighboring-eye corners (E_3E_4). The mouth-line is defined by connecting the two far corners of the mouth (We show the eye-lines and mouth-line in Fig. 1.).

We make a reasonable assumption that the eye-lines and the mouth-line are parallel in 3D space.

For convenient representation, image coordinate (x', y') are normalized using following formulas:

$$x = \frac{x' - x_0}{f_x},$$

$$y = \frac{y' - y_0}{f_y}, \tag{1}$$

where (x_0, y_0) are coordinates of the image-center point (principle point), f_x and f_y are the scale factors of the camera along the x- and y-axis, respectively. Using the normalized camera, i.e. the image plane is located at a unit distance from the optical center (focal length $f = 1$), we have

$$X_i = x_i Z_i, \tag{2}$$

$$Y_i = y_i Z_i, \tag{3}$$

$i = 1, 2, 3, 4.$

Here we assume $E_1(x_1, y_1)$ and $E_2(x_2, y_2)$ are image coordinates of the two far-eye corners (see Fig. 2), $E_1(X_1, Y_1, Z_1)$ and $E_2(X_2, Y_2, Z_2)$ being their 3D coordinates respectively (see Fig. 1); $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ are image coordinates of the two far-mouth corners (see Fig. 2), $M_1(X_1, Y_1, Z_1)$ and $M_2(X_2, Y_2, Z_2)$ being their 3D coordinates, respectively (see Fig. 1).

If the 3D space distances between the two far-eye corners and the two far-mouth corners are represented as D_e and D_m, respectively, then we have

$$D_e = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2},$$

$$D_m = \sqrt{(X_4 - X_3)^2 + (Y_4 - Y_3)^2 + (Z_4 - Z_3)^2}. \tag{4}$$

From perspective projection, the eye-lines and the mouth-line should intersect at one point in the image, i.e. at the vanishing point. Although two parallel lines are sufficient, we can get a least squares solution of the vanishing point using the above three parallel lines. Three points VP_1, VP_2, VP_3 can be obtained from eye-line E_1E_2 and eye-line E_3E_4, eye-line E_1E_2 and mouth-line M_1M_2, eye-line E_3E_4 and mouth-line M_1M_2, respectively, as shown in Fig. 2. If the vanishing point is $V_p(u_\infty, v_\infty)$, let (d_x, d_y, d_z) represents the 3D-direction vector of the eye-lines and thus also for the mouth-line. We have

$$(d_x, d_y, d_z) = \frac{1}{\sqrt{u_\infty^2 + v_\infty^2 + 1}} \begin{bmatrix} u_\infty \\ v_\infty \\ 1 \end{bmatrix}. \tag{5}$$

2.1. 3D positions of feature corners

Generally speaking, the ratio of the length D_e to the length D_m is a more stable measurement than the lengths themselves for different people. So the case that the ratio of lengths D_e and D_m is given instead of the lengths themselves is considered first. The ratio is immediately available from a fronto-parallel view of the face. The orientation of the facial plane and relative positions of the corners can be located in the first case. In the second case, both the ratio and one of the lengths are given hence the absolute 3D positions of the four far corners and the orientation of the facial plane can be determined. The above two cases are discussed in Sections 2.1.1 and 2.1.2, respectively.

2.1.1. Case 1: The ratio is known

The ratio of the length of the eye-line segment and the length of the mouth-line segment is given.

In this case, the orientation of the facial plane and 3D relative positions of the feature corners can be calculated using the vanishing point and the ratio of the two lengths.

From formula (5), the 3D-direction vector of eye-line is (d_x, d_y, d_z). We arrive at

$$(X_2 - X_1)/D_e = d_x, \tag{6}$$

$$(Y_2 - Y_1)/D_e = d_y, \tag{7}$$

$$(Z_2 - Z_1)/D_e = d_z. \tag{8}$$

Similarly, because the 3D-direction vector of the mouth-line is (d_x, d_y, d_z) too, we have

$$(X_4 - X_3)/D_m = d_x, \tag{9}$$

$$(Y_4 - Y_3)/D_m = d_y, \tag{10}$$

$$(Z_4 - Z_3)/D_m = d_z. \tag{11}$$

The ratio (denoted as r) of the lengths of the eye-line and mouth-line segments is

$$r = D_e/D_m. \tag{12}$$

From formulas (6)–(12), we can obtain

$$b = \frac{d_y}{d_x}\, a, \tag{13}$$

$$c = \frac{d_z}{d_x}\, a, \tag{14}$$

$$d = \frac{a}{r}, \tag{15}$$

$$e = \frac{b}{r}, \tag{16}$$

$$f = \frac{c}{r}, \tag{17}$$

where

$$a = X_2 - X_1, \tag{18}$$

$$b = Y_2 - Y_1, \tag{19}$$

$$c = Z_2 - Z_1, \tag{20}$$

$$d = X_4 - X_3, \tag{21}$$

$$e = Y_4 - Y_3, \tag{22}$$

$$f = Z_4 - Z_3. \tag{23}$$

Let $a = 1$, so b, c, d, e, f can be determined in turn using formulas (13)–(17). This assumption means we can get only relative 3D positions of the feature corners. However, the orientation can be obtained uniquely.

Replace X_1 and X_2 in Eq. (18) with formula (2), we have

$$x_2 Z_2 - x_1 Z_1 = a. \tag{24}$$

Replace Y_1 and Y_2 in Eq. (19) with formula (3), we have

$$y_2 Z_2 - y_1 Z_1 = b. \tag{25}$$

Z_1 and Z_2 can be solved from Eqs. (24) and (25). Hence, X_1, X_2, Y_1 and Y_2 can be solved using Eqs. (2) and (3).

Similarly, Z_3 and Z_4 can be solved using the following formulas:

$$x_4 Z_4 - x_3 Z_3 = d, \tag{26}$$

$$y_4 Z_4 - y_3 Z_3 = e. \tag{27}$$

Hence, X_3, X_4, Y_3 and Y_4 can be solved using Eqs. (2) and (3).

2.1.2. Case 2: One of the two lengths and the ratio are known

In this case, the absolute 3D positions of the feature corners can be determined.

Formulas (6)–(8) described in the above case are still be used

$$(X_2 - X_1)/D_e = d_x, \tag{6}$$

$$(Y_2 - Y_1)/D_e = d_y, \tag{7}$$

$$(Z_2 - Z_1)/D_e = d_z. \tag{8}$$

Replace X_1 and X_2 in Equation (6) with formula (2), we have

$$(x_2 Z_2 - x_1 Z_1)/D_e = d_x. \tag{28}$$

From Eq. (8), we obtain

$$Z_1 = Z_2 - D_e d_z. \tag{29}$$

From Eqs. (28) and (29), we arrive at

$$(x_2 - x_1)Z_2 + D_e d_z x_1 = D_e d_x. \tag{30}$$

So, we obtain the following formula:

$$Z_2 = D_e(d_x - d_z x_1)/(x_2 - x_1). \tag{31}$$

Replace Y_1 and Y_2 in Eq. (7) with formula (3), we have

$$(y_2 Z_2 - y_1 Z_1)/D_e = d_y, \tag{32}$$

From Eqs. (29) and (32), we obtain

$$(y_2 - y_1)Z_2 + D_e d_z y_1 = D_e d_y. \tag{33}$$

Hence,

$$Z_2 = D_e(d_y - d_z y_1)/(y_2 - y_1), \tag{34}$$

where Z_1 and Z_2 can be calculated from Eqs. (31) and (34) and Eq. (29).

Hence X_1, X_2, Y_1 and Y_2 can be solved using Eqs. (2) and (3).

The 3D-space length of the mouth-line segment D_m can be obtained using the length D_e and the ratio r. Similarly, the 3D coordinate of far corners of mouth $M_3(X_3, Y_3, Z_3)$ and $M_4(X_4, Y_4, Z_4)$ also can be calculated using their correspondent image coordinates then.

2.2. Normal of the facial plane

From the four 3D coordinates of the far corners obtained under above two cases (relative coordinates for the first case and absolute coordinates for the second case), we can calculate the facial normal $\vec{N}$ as cross product of the two space vectors $\overrightarrow{M_2E_2}$ and $\overrightarrow{M_2M_1}$ (see Fig. 1):

$$\vec{N} = \overrightarrow{M_2E_2} \times \overrightarrow{M_2M_1} \tag{35}$$

The four points E_1, E_2, M_1 and M_2 are in general not expected to be coplanar due to noise. So instead, the facial normal line could be calculated as the average of following cross products of the pairs of space vectors: $\overrightarrow{M_2E_2}$ and $\overrightarrow{M_2M_1}$, $\overrightarrow{E_2E_1}$ and $\overrightarrow{E_2M_2}$, $\overrightarrow{E_1M_1}$ and $\overrightarrow{E_1E_2}$, $\overrightarrow{M_1M_2}$ and $\overrightarrow{M_1E_1}$.

2.3. Degenerate analysis

For frontal view, the facial plane is parallel to the image plane and so the estimation of the vanishing point on the image plane will be unstable but fortunately this does not equate to large directional errors in the pose estimation. The facial normal coincides with the view axis $(0, 0, 1)$. The 3D eye-lines and mouth-line have the same slope with their perspective projection lines, we have

$$d_x = (x_2 - x_1)/\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \tag{36}$$

$$d_y = (y_2 - y_1)/\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \tag{37}$$

$$d_z = 0. \tag{38}$$

Then the 3D coordinates of the corners of eyes and mouth are calculated similarly as above. This degenerate case can be detected when the eye-lines and mouth-line in the image are nearly parallel.

Next, we consider the degenerate case when the facial plane is orthogonal to the image plane, i.e. profile image. If the far-eye corners, the neighboring-eye corners, and the far-mouth corners are not occluded, respectively, the algorithm we developed has a unique solution in this case. However, if the far-eye corners, the neighboring eye corners, and the far-mouth corners are occluded, respectively, i.e. fully profile image, then the facial normal is orthogonal to the normal of the image plane and the 3D coordinate of corners of eyes and mouth are calculated same as above.

2.4. Measuring the ratio and the lengths

We found that the ratios of the length of eye-line and mouth-line segments are approximate invariant from face to face. In the following sub-sections we discuss the method to measure the ratio from face images. Two methods have been developed for this purpose. A simple method that measures the ratio from frontal face images using projective invariant will be discussed in Section 2.4.1. In Section 2.4.2, an alternative stereo vision-based method is briefly described, which can be applied to obtain the real lengths and consequently the ratio.

2.4.1. Measuring the ratio using projective invariant

If a facial plane is parallel to the image plane, i.e., frontal facial image, then the lengths of the eye- and mouth-line segments will be perspective projective invariant. So we can calculate the ratio directly from the projections of the two lengths. The distribution of the ratios has been measured over 300 frontal view face images. The experimental face images, include female, male, some of them with four expressions, are downloaded from the face database provided by University of Stirling [17]. The far-eye corners and the far-mouth corners are manually located on the image plane. The lengths of the eye- and mouth-line segments can be calculated. Hence, the ratio can be obtained. The average ratio of the 300 images is found to be 1.98, the standard derivation of the ratio is 0.02. Values of the ratio of the experimental images are shown in Fig. 3. Two examples that implement the variances of the ratio vs. the facial expressions are shown in Figs. 4 and 5 respectively. We

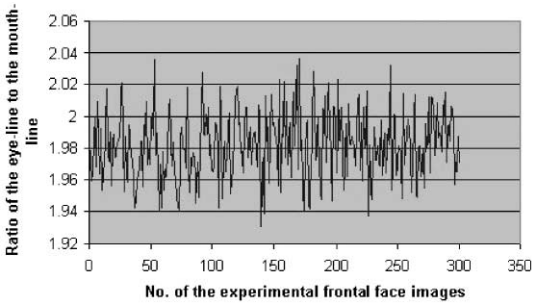


Fig. 3. Ratios and the average ratio of the experimental frontal face images.

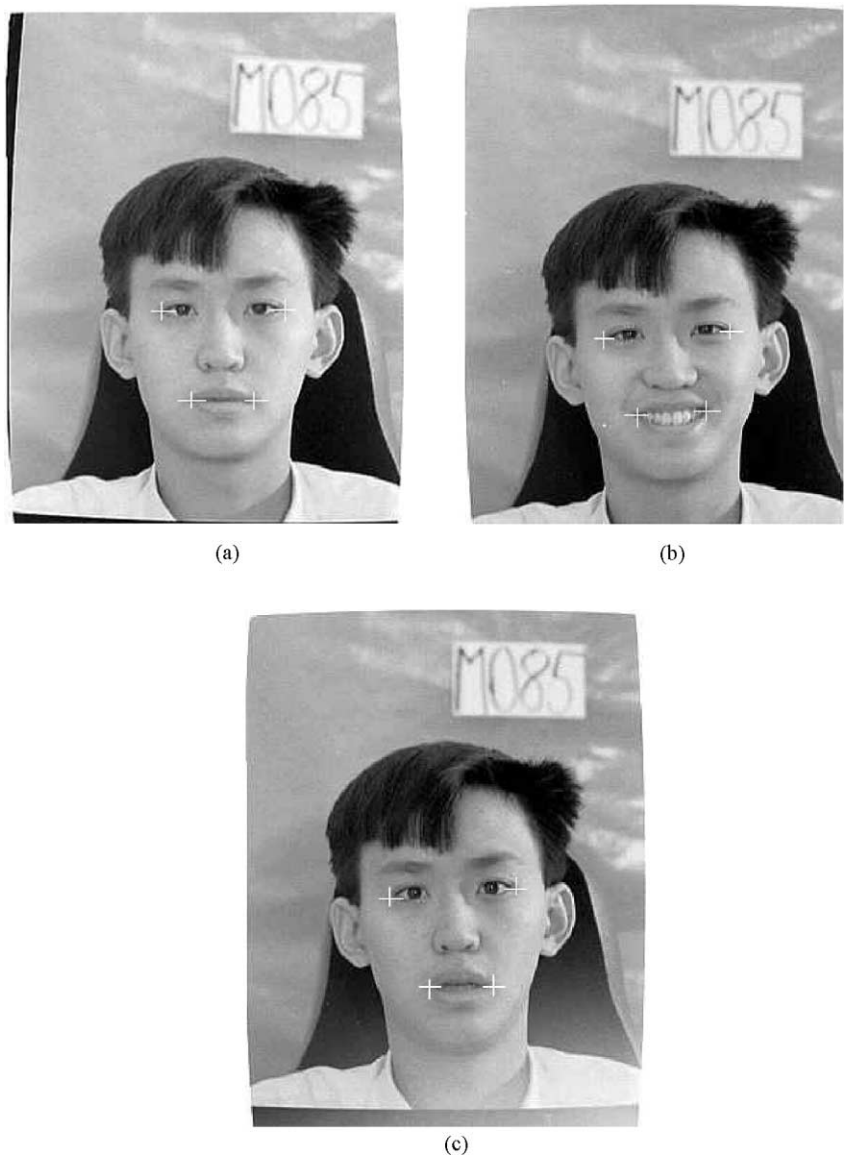


Fig. 4. Variance of the ratio vs. facial expressions: (a) male frontal neutral, 2.01; (b) male frontal smiling, 1.95; (c) male frontal surprise, 2.05.

can see that the ratio is stable vs. the facial expressions, which include smile, surprise, disgust, etc.

2.4.2. Measuring the lengths using stereo vision

If the actual lengths of the eye- and mouth-line segments are expected to locate the absolute 3D position of the head, then a method to construct partial 3D facial model is needed because the ratio and the lengths are the primitives of the partial 3D facial model.

A generic face model can be used to construct a special facial model. There are a number of papers dealing with fitting a generic 3D-face model to some specific person's face using several available 2D images. Aizawa et al. [18] originally adopted the generic face model, a geometric wire-frame model. Akimoto et al. [19] used calibrated front- and side-view images of a person's head to acquire the features need for fitting a 3D generic head model to the person's head, although the constraints were rather stringent. Huang et al. [20] extend Aizawa's method to allow using front and arbitrary side view, however they have to estimate the direction of the arbitrary side view in order to rotated model to that direction first. Zhang et al.

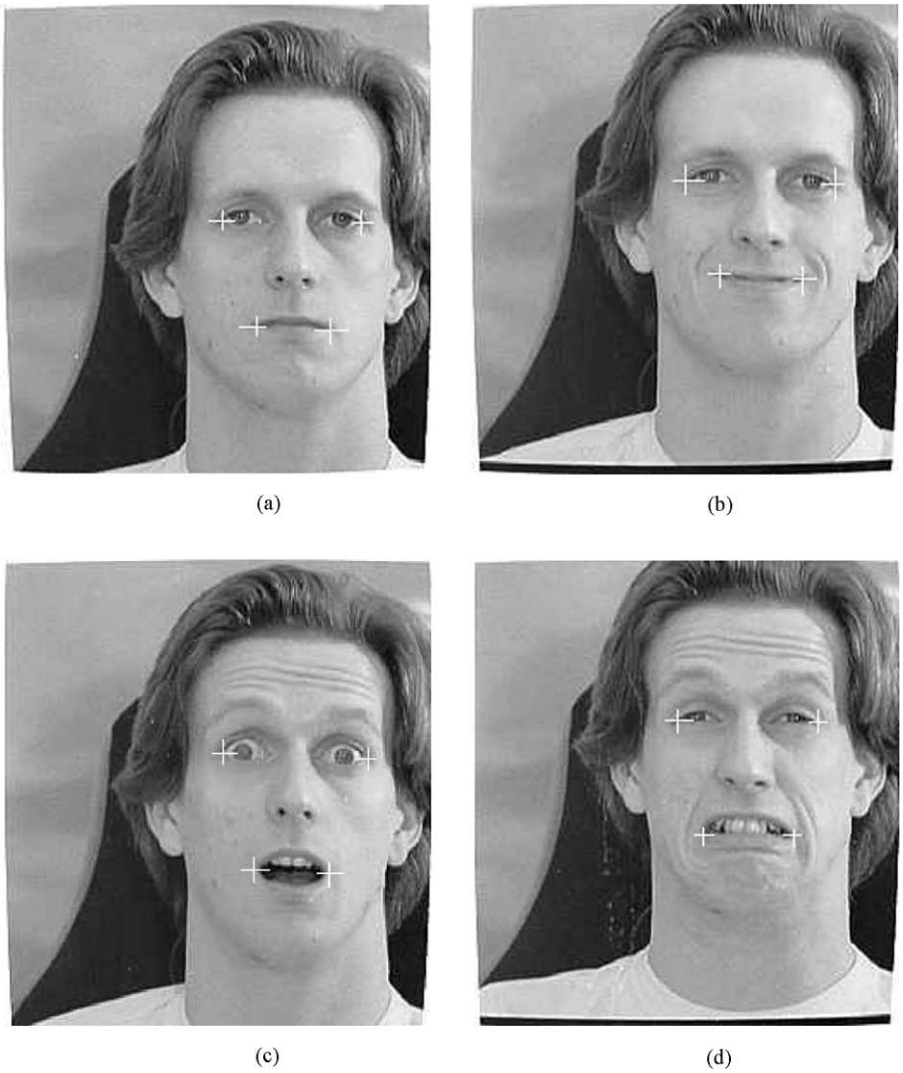
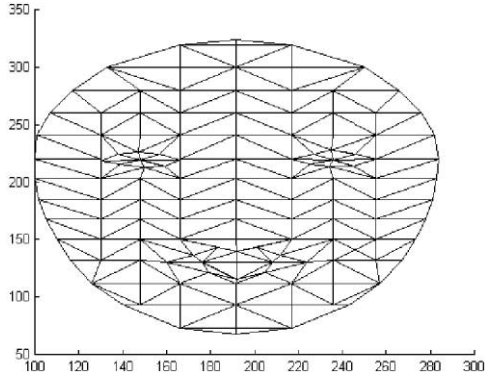


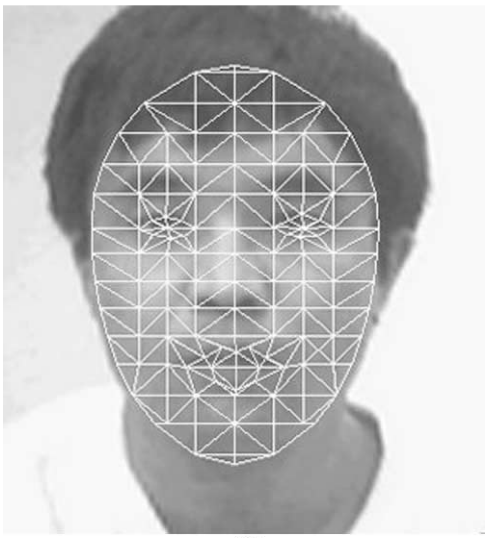
Fig. 5. Variance of the ratio vs. facial expressions: (a) male frontal neutral, 2.01; (b) male frontal smiling, 1.94; (c) male frontal surprise, 2.0; (d) male frontal disgust, 1.89.

[29] reconstruct Euclidean face structure from uncalibrated images and some prior knowledge of face structure such as distance and angle are used to solve the relationship between affine coordinates and Euclidean coordinates. Richard et al. [21] presented a global scheme for 3D face reconstruction and face segmentation from stereo images. Braccini et al. [22] model human heads based on an integrated 2D–3D approach.

In our approach, the reconstruction problem is formulated as a correspondence problem between two sets of vertices on the two face images. Here, we give a brief description of our calibration algorithm. A fully calibrated stereo vision system is utilized for the necessary calibration of each new face. The vertices are obtained by fitting general 2D facial model to the face image of a person on the first front images. Then, the vertex lying on the face model can be found in the other image by similarity measure (correlation-based) along the epipolar line. We use a front and an arbitrary side stereo color face images to construct 3D facial model. Our algorithm [23], which can detect the facial contour and the facial features using color and morphological operations, can be applied to the front-view face image first. Then the locations of the far-eye corners, far-mouth corners and the facial symmetry axis can be used for fitting a general 2D model onto the face image. The general 2D facial model we used is shown in Fig. 6(a). The result after fitting the 2D model onto a front face image is shown in Fig. 6(b).



(a)



(b)

Fig. 6. 3D modeling of human face using stereo vision technique: (a) a general 2D facial model; (b) fitting the 2D general model to a specific face.

The correspondent points of the vertices that lie on the first image are searched along the epipolar line using correlation-based technique on another image plane. Using the correspondent 2D coordinates on the two image planes, the 3D coordinate of these feature corners can be calculated using stereo vision algorithm [24] and hence, the ratio of the length of eye- and mouth-line segments can be determined.

3. Experimental results

We have tried our algorithm using synthetic data and real images. Its robustness has been tested by adding disturbances to the 2D positions of the facial corner points. The experiments show that our algorithm can provide a good estimation of pose of human head even up to a close distance.

3.1. Simulation using synthetic data

Computer simulations have been performed to analyze the errors in the presence of noise. The test utilizes synthetic image size of 512×512. We set practical camera parameters:

$$C_x = 0, \quad C_y = 0, \quad C_z = 50\,(\text{cm}), \tag{39}$$

$$f_x = f_y = 1000, \tag{40}$$

$$u_0 = v_0 = 255, \tag{41}$$

$$\theta = 0^0, \quad \phi = 90^0, \quad \varphi = 0^0, \tag{42}$$

where (C_x, C_y, C_z) is the 3D world coordinate of lens center of the camera. ϕ, θ, φ are the three angles that define the rotation matrix R such that

$$R = R_x(\phi)R_y(\theta)R_z(\varphi). \tag{43}$$

The optical axis is the z-axis of camera coordinate system, facing towards human face. The x-axis is horizontal axis, and y-axis is vertical axis. The transform matrix can be constructed using the above intrinsic and extrinsic parameters of the camera [26]:

$$T = \begin{bmatrix} -100/3 & 0 & -8.5 & 255 \\ 0 & 100/3 & -8.5 & 255 \\ 0 & 0 & -1/30 & 1 \end{bmatrix}. \tag{44}$$

In the world coordinate system, the initial coordinates of the four cornerpoints in the target face (Fig. 1) are set:

$$E_1(-5.25, -6, z), \quad E_2(5.25, -6, z),$$

$$M_1(-2.65, -1, z), \quad M_2(2.65, -1, z).$$

where z is the distance along the z-axis from the origin to the face. Hence $D_e = 10.5\,\text{cm}$, $D_m = 5.3\,\text{cm}$. The ratio of the eye- and the mouth-line is about 1.98.

The initial face plane is parallel to the image plane, distance between the two planes:

$$z - C_z.$$

The methods for generating simulation data are shown in Fig. 7. The left figure of Fig. 7 shows pose data generated by rotating facial plane about the facial symmetry, the middle figure shows the pose data generated by rotating the facial plane about the horizontal axis; the right figure shows the pose data generated by rotating the facial plane about the optical axis.

The simulations on a set of different poses have demonstrated the performance of our method. The rotated four corner points' 3D coordinate (accurate) can be

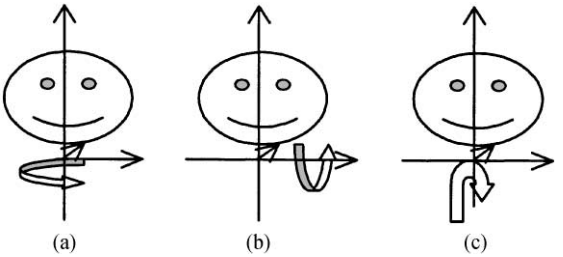


Fig. 7. Simulations on different pose, rotating the facial plane about: (a) facial symmetry: pan; (b) horizontal axis: pitch; (c) optical axis: tilt.

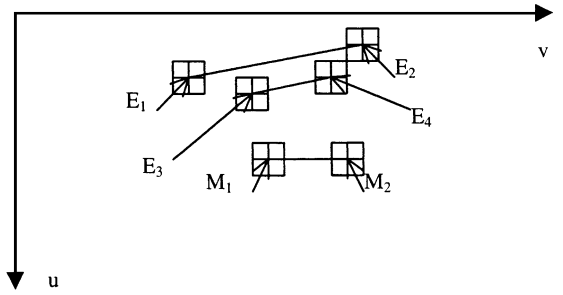


Fig. 8. Adding perturbations to the facial corners on the image plane, perturbations n pixels to a corner point means the new position of the corner will lie at random positions within the corner-centered $(2n+1)\times(2n+1)$ window.

calculated using the rotation angles around the axes and stored as actual 3D positions. The perspective projection of the calibration target in the image is computed using the above imaging model. The vanishing point can thus be located by using projected eye-lines and mouth-line. Least-squares errors fitting was used to find the vanishing point from the projected two eye-lines and one mouth-line. Subsequently, the four corner points' 3D coordinates can be re-calculated using our method via the vanishing point. So the errors between actual 3D position and the values calculated using our method can be obtained. The orientation errors can be obtained too.

In order to test the robustness of our method, we perturb the corners image positions before calculating the vanishing point. The image coordinates (u, v) of all corners are disturbed by adding a Gaussian noise generated by a modified random number function. The output of the random function lie at random the range: $-x_{max}$ to x_{max}, where the magnitude value correspond to the maximum number of disturbance pixels we applied to the image. Perturbations by a magnitude not more than n pixels distance to a corner point means its new position will lie at random positions within the corner-centered $(2n+1)\times(2n+1)$ window positions. For example, one of the six corner points on the image is perturbed by 1 pixel displacement, the new position will be lie at a random position within the corner centered 3×3 window, see Fig. 8. For each noise deviation, 100 simulation results are generated and averaged.

We show here an example of the simulation on the poses generated by rotating facial plane around the facial symmetry. Simulations on other poses have the similar performances. Four corner points rotated about the facial symmetry axis from left -85° to right 85° in steps of 1°. The orientation errors of the facial normal are shown in Fig. 9 where we assume only the ratio (1.98 as above) of the two lengths is known. The errors of the 3D positions of the four corner points are shown in Figs. 10 and 11 where the length of eye-line segment D_e (10.5 cm as above) and ratio (1.98 as above) are assumed known.

The experiment results have shown good performance of the method. The closer the camera is to the face, the more accurate the calculation is, and this can be seen by comparing Figs. 10 and 11 for 3D position errors of the feature corners, and Fig. 9 or Fig. 12 for the orientation errors of the facial plane. When the distance between the

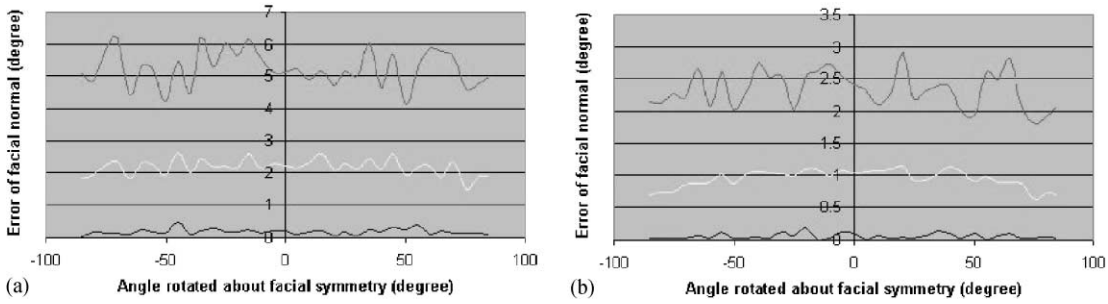


Fig. 9. Errors of the facial normal. Only the ratio (1.98) is used. The perturbation applied to the points is 1 pixel. Three curves, top: maximum error; middle: average error; bottom: minimum error. One-hundred simulation results are generated and averaged. The distance between the original facial plane and the image plane is: (a) 60 cm; (b) 50 cm.

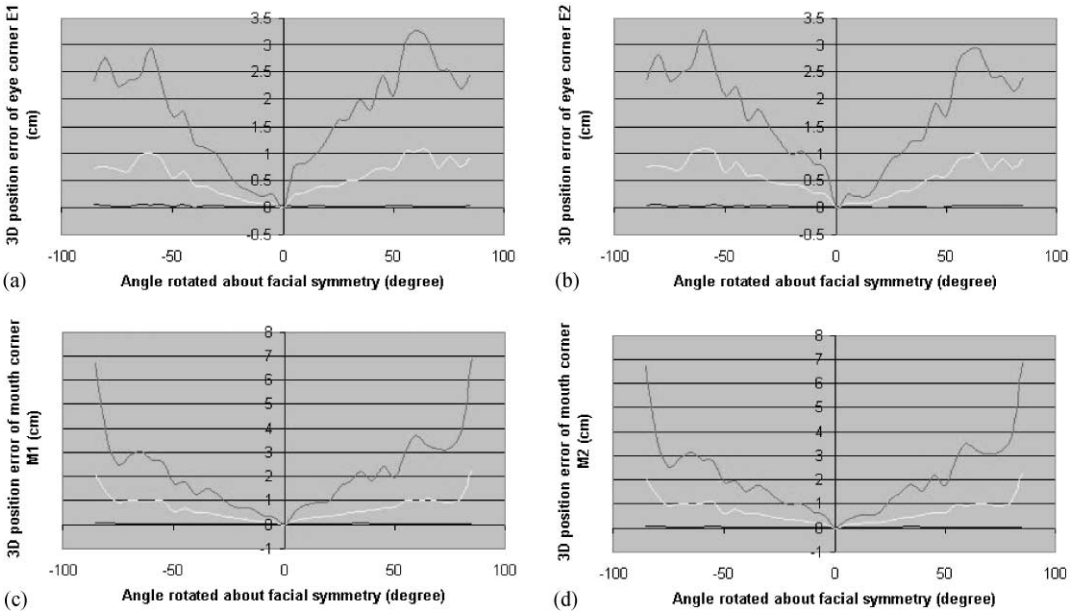


Fig. 10. 3D position errors of the four corners. The ratio (1.98) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 60 cm; 100 simulation results are generated and averaged; top: maximum error; middle: average error; bottom: minimum error. (a) Point E_1, (b) point E_2, (c) point M_1, (d) point M_2.

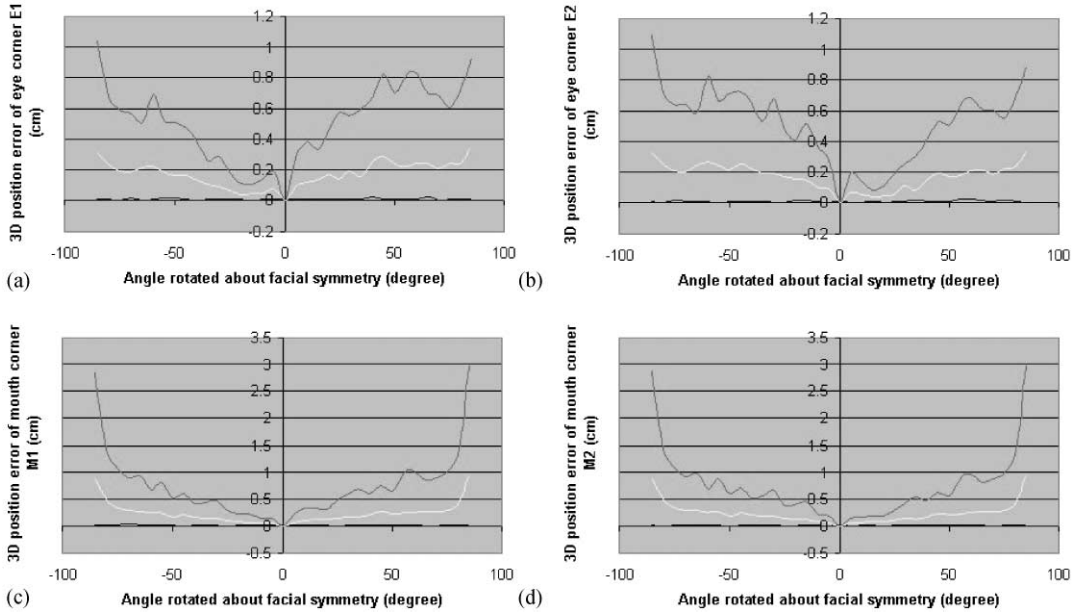


Fig. 11. 3D position errors of the four corners. The ratio (1.98) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 50 cm; 100 simulation results are generated and averaged; three curves, top: maximum error; middle: average error; bottom: minimum error. (a) Point E_1, (b) point E_2, (c) point M_1, (d) point M_2.

initial facial plane and the image plane is 60 cm, the average position errors are less than 1 cm within the visible range (Fig. 10) and the average orientation errors are less than 2.5° (Fig. 9). The algorithm here requires visibility of the eye corners (at least two far-eye corners or two neighboring-eye corners are required) and two

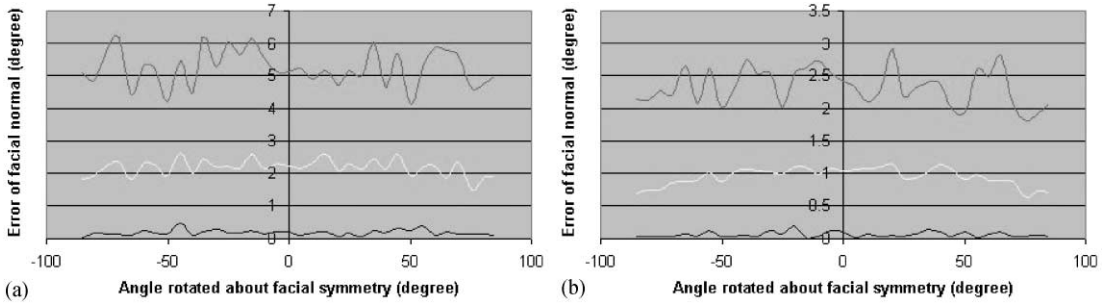


Fig. 12. Errors of the facial normal line. The ratio (1.98) and one of the two lengths are used. The perturbation applied to the points is 1 pixel. Three curves, top: maximum error; middle: average error; bottom: minimum error. One-hundred simulation results are generated and averaged. The distance between the original facial plane and the image plane is: (a) 60 cm and (b) 50 cm.

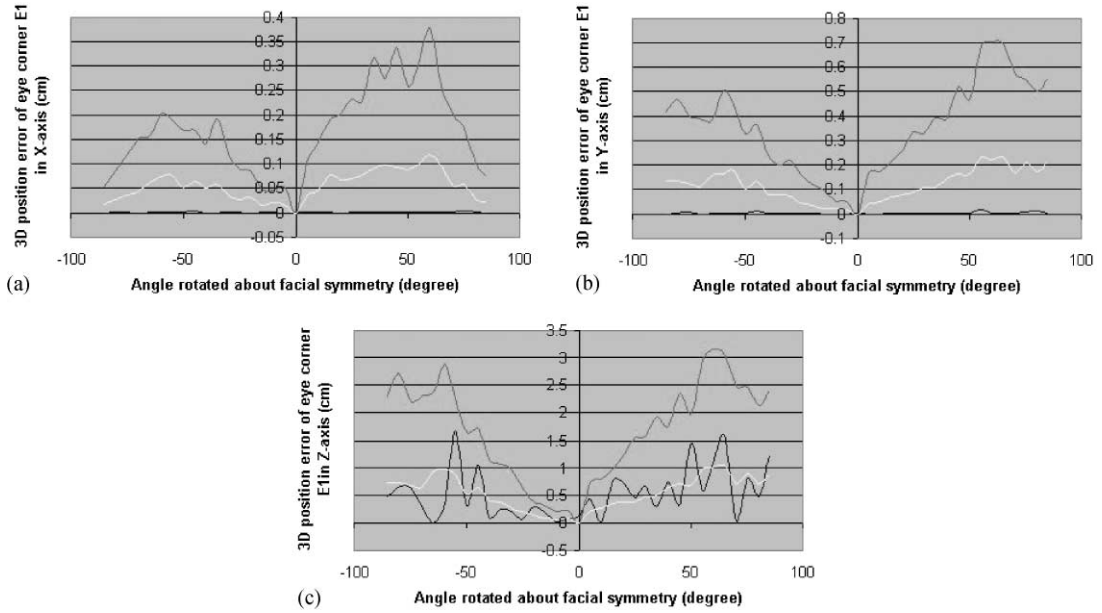


Fig. 13. 3D position error (absolute) of eye corner E_1 (see Fig. 10(a)) in X-, Y- and Z-axis separately. The ratio (1.98) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 60 cm; 100 simulation results are generated and averaged; Top: maximum error; middle: average error; bottom: minimum error. (a) Error in X-axis; (b) error in Y-axis; (c) error in Z-axis.

far-mouth corners. Although this is a serious problem for near-profile views of the face, these points are visible for a wide range of poses (visible range). When the points are obscured, their positions could be estimated using the locations of the other facial features.

This experiments have shown that the largest error lies in the Z-axis. An example in Fig. 13 shows the 3D position errors of Fig. 10(a) in X-, Y- and Z-axis separately.

To test how the variation of ratio affects the estimated pose, we test the accuracy by letting the ratio change from 1.94 to 1.98, corresponding to the bound we obtained from 300 face images shown in Fig. 3. We keep the length of eye-line at 10.5 cm as above, while estimating the pose with varied ratios (lengths of the mouth-line are thus calculated according the different ratio). This is just an approximation of the variation because both the length of the eye-line and the length of the mouth-line will change according a complex relation for different face expressions. The perturbations (one pixel) are still applied to the eye- and mouth-corners in this ratio variation experiment.

The orientation and 3D position error analysis are shown in Figs. 14–16 (ratio of 1.94) and Figs. 17–19 (ratio of 2.03). We can see the pose results are somewhat sensitive to the variation of the ratio. Fortunately, the

variations of the ratio from person to person are stable and this can be seen from our measurements of 300 frontal face images including different persons and different expressions shown in Fig. 3. So the practicality of the algorithm is good. The accuracy is low for special large variations of the ratio, for example, ratio 1.89 for disgust expression shown in Fig. 5(d). From Figs 20–22, we show the error measurement ratio of 1.89, of the orientation and 3D position. This drawback can be overcome using the continuous constraint in real time pose estimation applications.

Our algorithm assumes the eye-lines are parallel to mouth-line. The pose of the human face can be estimated based on the vanishing point formed by the eye-lines and mouth-line. If the four corners E_1, E_2, M_1 and M_2 are not coplanar, we can still obtain the pose by applying our algorithm. The precision in this case is tested by the simulations. Using the same assumption mentioned in Eqs. (39)–(44), we add 0.5 cm in Z coordinate of eye corner E_1 while keep E_2, M_1 and M_2 are coplanar, the orientation and 3D position errors are show in Figs. 23(a) and 24, respectively. Then we add 0.5 cm in Z coordinate of mouth corner M_1, the orientation and 3D position errors are show in Figs. 23(b) and 25, respectively.

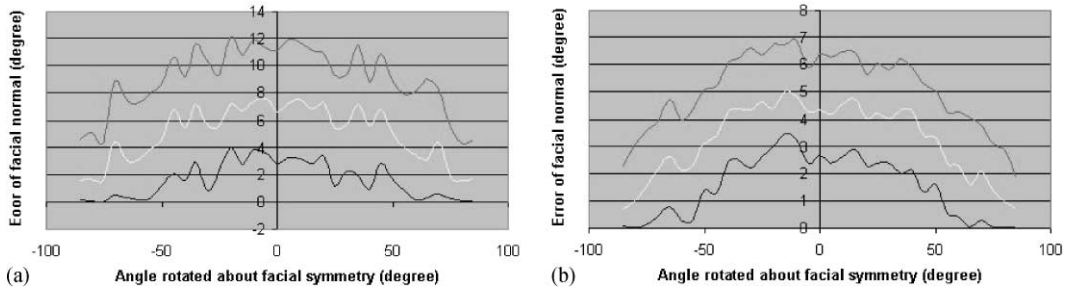


Fig. 14. Errors of the facial normal. Only the ratio (1.94) is used. The perturbation applied to the points is 1 pixel. Three curves, top: maximum error; middle: average error; bottom: minimum error. One-hundred simulation results are generated and averaged. The distance between the original facial plane and the image plane is: (a) 60 cm and (b) 50 cm.

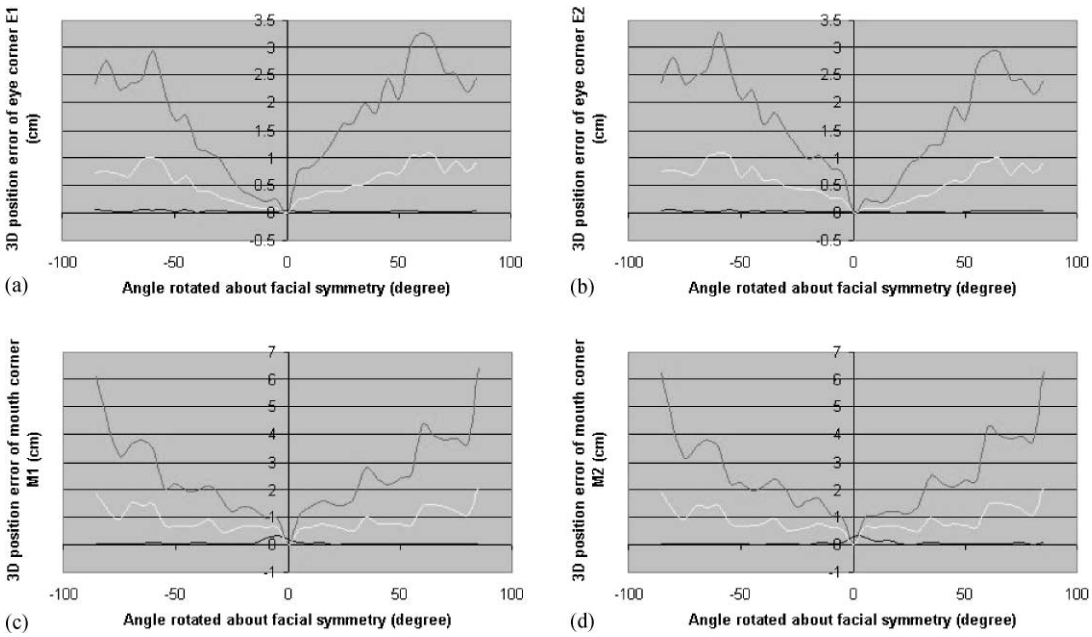


Fig. 15. 3D position errors of the four corners. The ratio (1.94) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 60 cm; 100 simulation results are generated and averaged; top: maximum error; middle: average error; bottom: minimum error. (a) Point E_1, (b) point E_2, (c) point M_1, (d) point M_2.

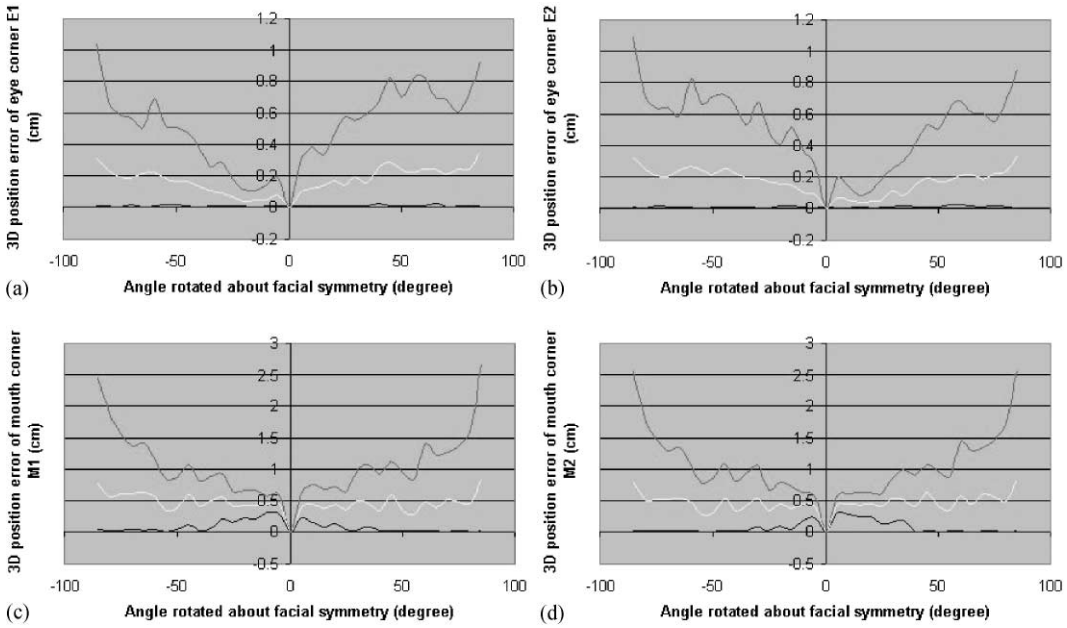


Fig. 16. 3D position errors of the four corners. The ratio (1.94) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 50 cm; 100 simulation results are generated and averaged; Three curves, top: maximum error; middle: average error; bottom: minimum error. (a) Point E_1, (b) point E_2, (c) point M_1, (d) point M_2.

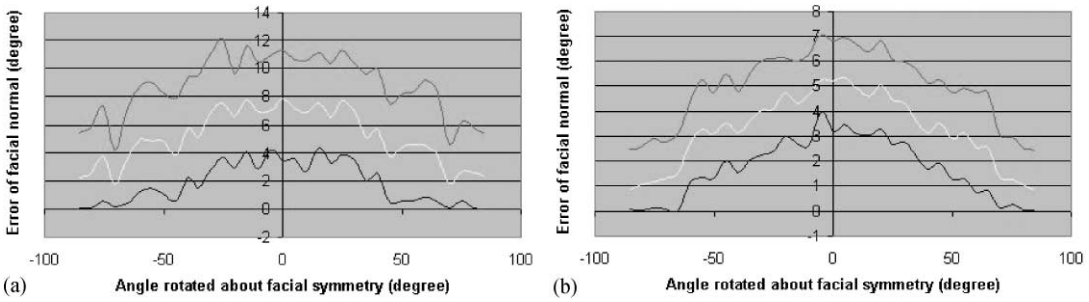


Fig. 17. Errors of the facial normal. Only the ratio (2.03) is used. The perturbation applied to the points is 1 pixel. Three curves, top: maximum error; middle: average error; bottom: minimum error. One-hundred simulation results are generated and averaged. The distance between the original facial plane and the image plane is: (a) 60 cm and (b) 50 cm.

Comparing Figs. 9(a) and 23(a), Figs. 9(b) and 23(b), Figs. 10 and 24, Figs. 10 and 25, we can see our algorithm is robust even the four corners E_1, E_2, M_1 and M_2 (see Fig. 1) are not coplanar.

3.2. Experiment with real image

We have implemented the algorithm using a Pentium 300 PC with 128 M RAM. A frame grabber and a HUNT HTC-460 CCD camera are installed. The camera is calibrated and the perspective projection matrix, and hence the intrinsic parameters, can be obtained. A preliminary set of images has been tried using our method. In this paper, we show the results of the calibrated image of Herve's head *n* Fig. 26. The size of the image is 512×512. The perspective projection matrix of the camera is given in [25]

$$T = \begin{bmatrix} 0 & 485.552 & 308.161 & -5.68434e-14 \\ 25.1169 & -27.787038 & 292.906 & 482.815 \\ -1.20063 & 0.948344 & 0.802503 & 1 \end{bmatrix}. \quad (46)$$

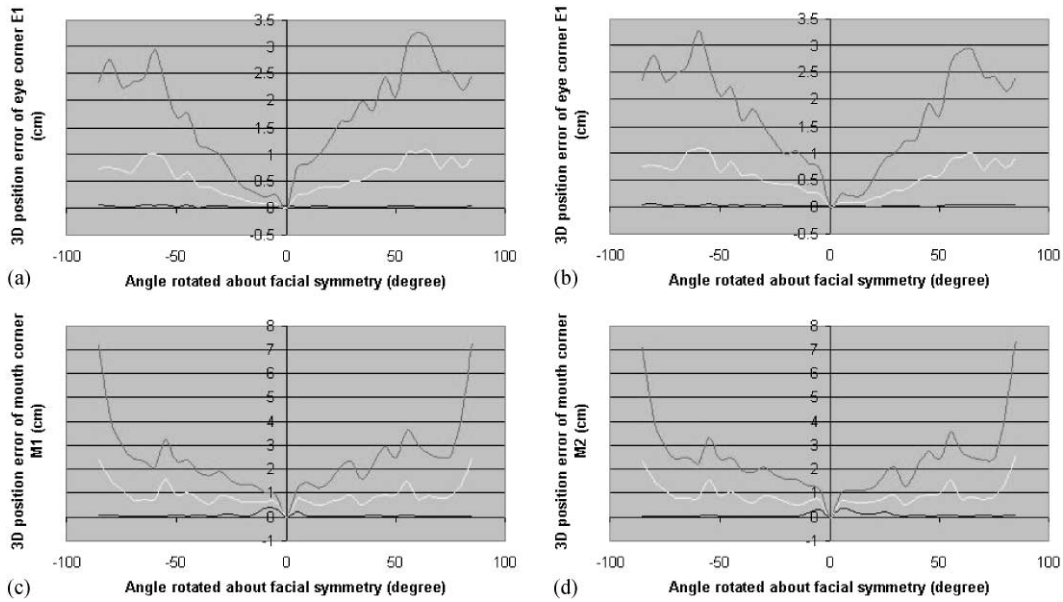


Fig. 18. 3D position errors of the four corners. The ratio (2.03) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 60 cm; 100 simulation results are generated and averaged; top: maximum error; middle: average error; bottom: minimum error. (a) Point E_1, (b) point E_2, (c) point M_1, (d) point M_2.

The intrinsic parameters of the camera, include the focal length and center of the image, can be obtained by decomposing the transform matrix [26].

The ratio of the 3D distance between the far-eye corners and the distance between the far-mouth corners is assumed as 1.98 (average of the ratios obtained from 300 face images, see Section 2.4.1). The 2D coordinates of the far-eye, neighboring-eye corners, and far-mouth corners can be detected by applying a standard procedure for facial feature extraction to the image, described in Refs. [23,27].

Far-eye corners:

$E_1(256, 138), \quad E_2(258, 375).$

Neighboring-eye corners:

$E_3(256, 210), \quad E_4(254, 298).$

Mouth corners:

$M_1(388, 194), \quad M_2(384, 304).$

The vanishing point of the two eye-lines and the mouth-line can be located on the image plane, then the 3D relative coordinates of left-eye corner, right-eye corner, left-mouth corner and right-mouth corner are obtained by using our algorithm:

$E_1(62.43, 0.98, -1.69)$(cm),

$E_2(59.56, 105.73, -179.20)$(cm),

$M_1(57.84, 96.15, -171.02)$(cm),

$M_2(56.99, 99.81, -176.11)$(cm).

Hence, the facial normal vector can be calculated using formula (35):

$\vec{N}(0.469129, -0.503611, 0.464343).$

The two eye corners, the two mouth corners and the facial normal are re-projected onto the image, as shown in Fig. 26. The 3D midpoints of the far eye corners E_1 and E_2, eye corner E_1 and mouth corner M_1, eye corner E_2 and mouth corner M_1, two mouth corners M_1 and M_2 are calculated respectively and also re-projected onto the face image. The facial normal is marked as a vector that starts at the center of the facial plane in Fig. 26.

The experiments show that our method has reasonable accuracy and is robust for pose estimation of the human head.

4. Conclusion

Vanishing point and vanishing line are important cues that can be exploited to infer 3D-object information from its image. In this paper, we have presented a new finding for computation of head pose by using vanishing point.

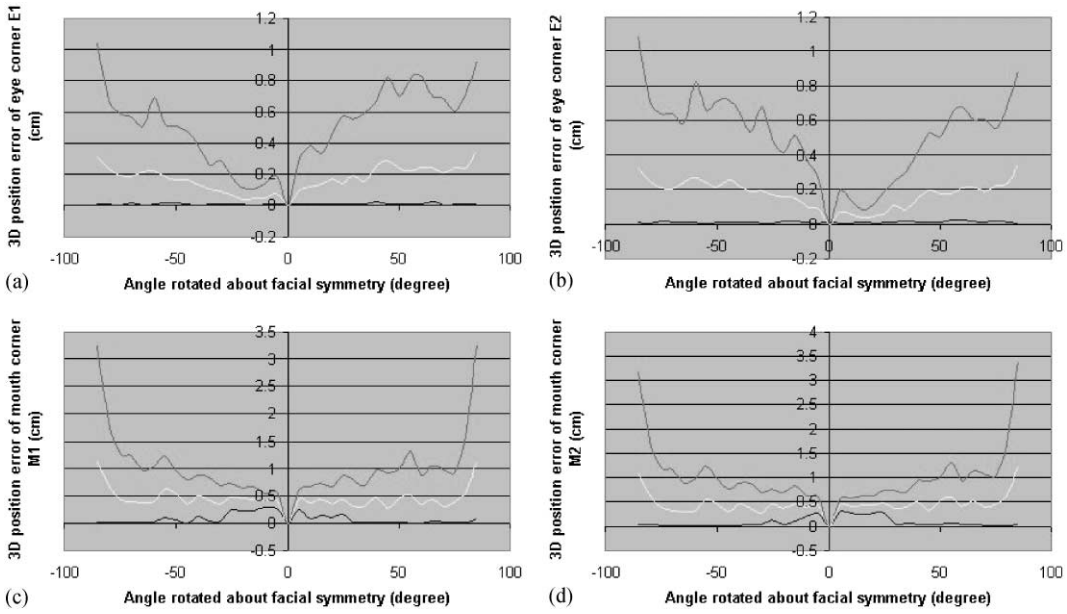


Fig. 19. 3D position errors of the four corners. The ratio (2.03) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 50 cm; 100 simulation results are generated and averaged; top: maximum error; middle: average error; bottom: minimum error. (a) Point E_1, (b) point E_2, (c) point M_1, (d) point M_2.

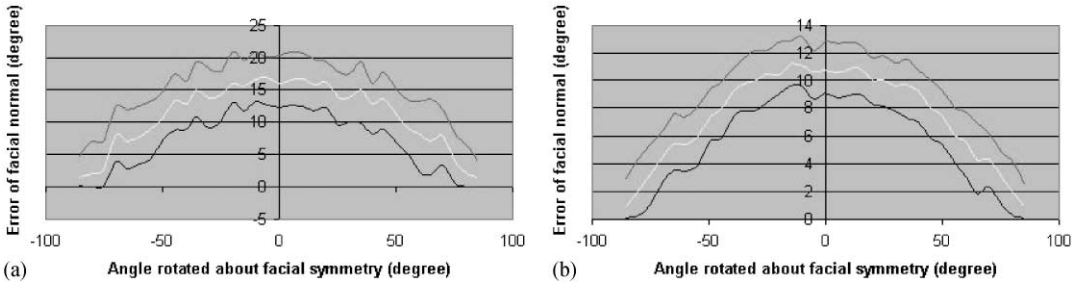


Fig. 20. Errors of the facial normal. Only the ratio (1.89) is used. The perturbation applied to the points is 1 pixel. Three curves, top: maximum error; middle: average error; bottom: minimum error. One-hundred simulation results are generated and averaged. The distance between the original facial plane and the image plane is: (a) 60 cm and (b) 50 cm.

A good and reasonable assumption about face structure, namely that eye-lines are parallel to mouth-line in 3D space, is used and the corresponding vanishing point on the image plane can be used to infer 3D information of the human faces. An analytic solution has been inferred and the pose can be determined uniquely when the ratio of the distance of the far-eye corners and distance of the far-mouth corners is known. The ratio is found to be stable from face to face so our algorithm is reliable. The approach is a new one where perspective projection and fully calibrated camera imaging model are employed. The robustness analysis shows that it is an alternative approach for estimating 3D pose (position and orientation) from a single view, especially, when an automatic method of finding the vanishing point is possible. On the other hand, our algorithm is reliable because the ratio we used here is more stable then the actual lengths themselves from face to face. Accuracy of the vanishing point plays an important role on performance of the proposed method. It should be attempted to correct the coordinates of the vanishing point such that it can be located more accurately. Fortunately, the occurrence of the lowest error corresponds to the corner visible range. Some techniques, such as the method developed by Quan and Mohr [28], can be used to correct the location of the vanishing point in order to

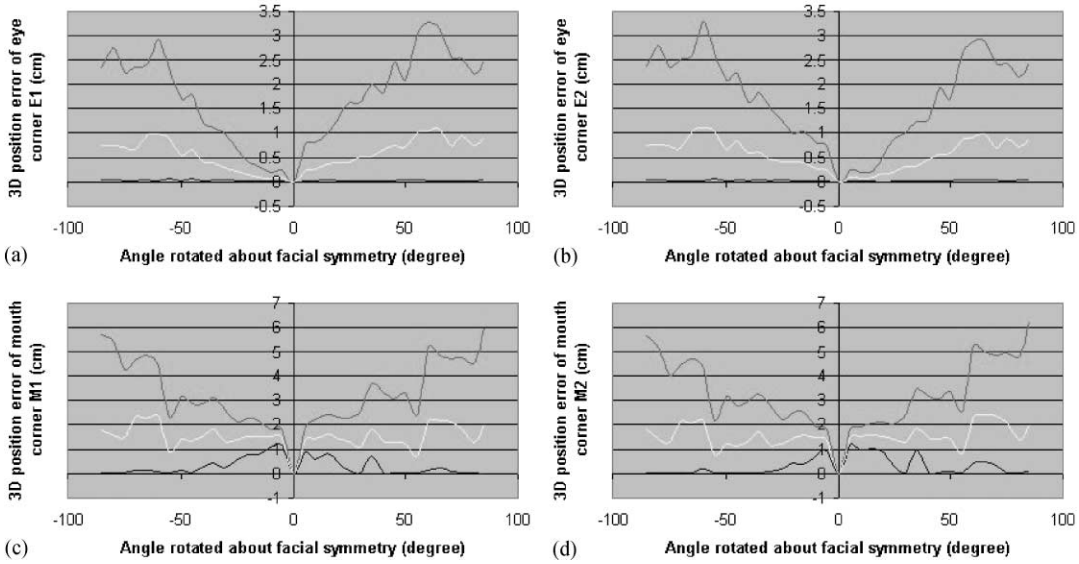


Fig. 21. 3D position errors of the four corners. The ratio (1.89) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 60 cm; 100 simulation results are generated and averaged; top: maximum error; middle: average error; bottom: minimum error. (a) Point E_1, (b) point E_2, (c) point M_1, (d) point M_2.

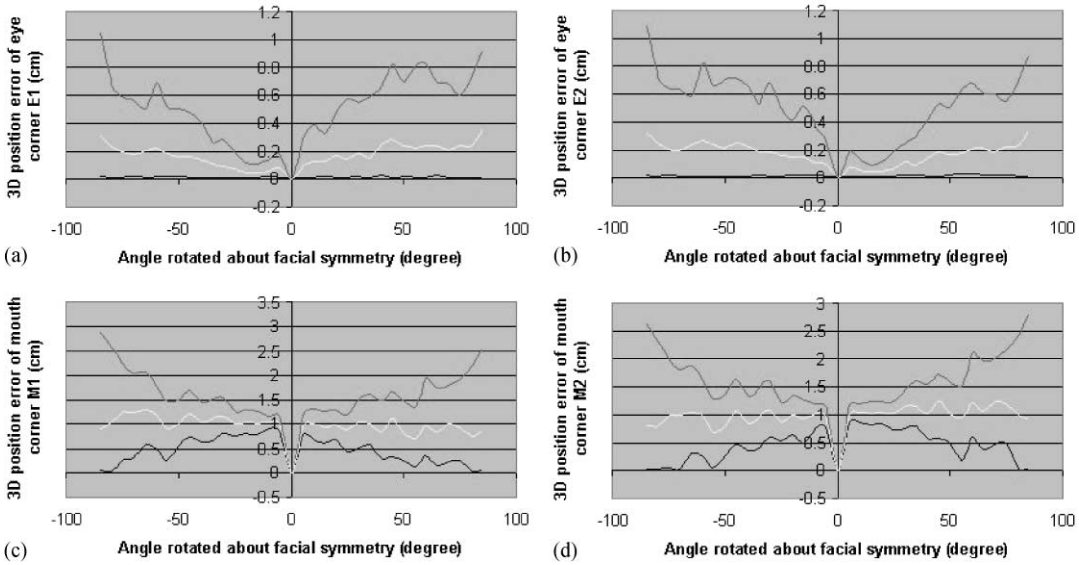


Fig. 22. 3D position errors of the four corners. The ratio (1.89) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 50 cm; 100 simulation results are generated and averaged; top: maximum error; middle: average error; bottom: minimum error. (a) Point E_1, (b) point E_2, (c) point M_1, (d) point M_2.

improve the precision of our algorithm. The calculation involved in this method is little and thus useful for real-time pose determination. The method we proposed here can also provide a good initial solution for some complex pose estimation approaches where iterative procedure may be needed. Another future research will focus on applying this method for real-time human–machine interaction.

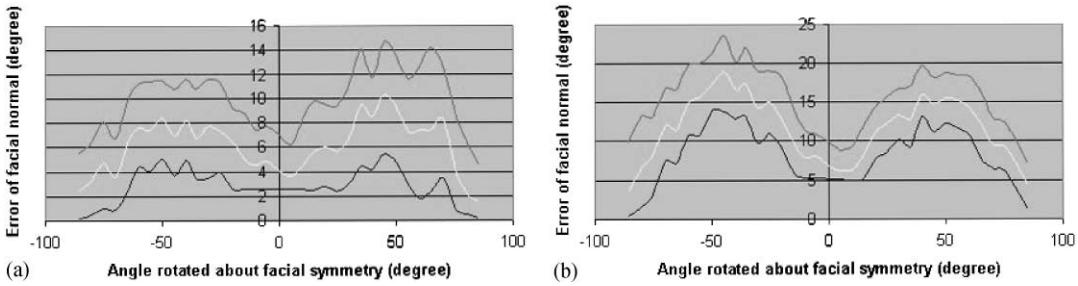


Fig. 23. Errors of the facial normal. Only the ratio (1.98) is used. The perturbation applied to the points is 1 pixel. Three curves, top: maximum error; middle: average error; bottom: minimum error. One-hundred simulation results are generated and averaged. The distance between the original facial plane and the image plane is 60 cm. (a) Adding 0.5 cm to the eye corner E_1. (b) Adding 0.5 cm to the mouth corner M_1.

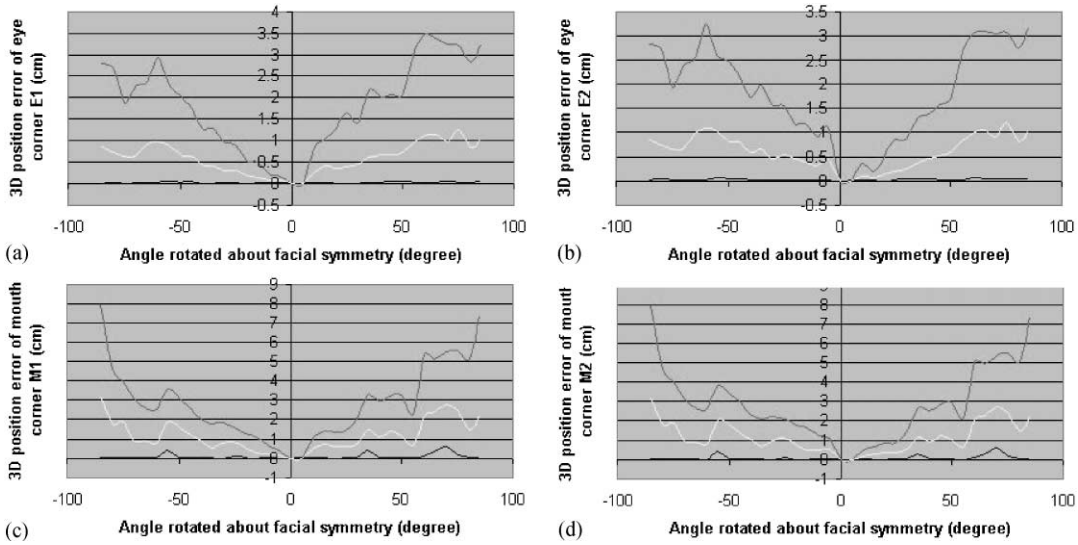


Fig. 24. 3D position errors of the four corners. Adding 0.5 cm to the eye corner E_1. The ratio (1.98) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 60 cm; 100 simulation results are generated and averaged; top: maximum error; middle: average error; bottom: minimum error. (a) Point E_1, (b) point E_2, (c) point M_1, (d) point M_2.

5. Summary

The pose of a human face is a potential tool for a natural user interface. There are two different transformations may be used for the pose estimation from a monocular image: perspective or a sub-class of the affine transformation. The former one precisely models the actual projection of a 3D scene to the image plane. However, the calculation is complex and time consuming and can deliver up to a fourfold ambiguity in the estimate of the pose. Most real-time systems usually use affine transformation, such as in the work by Gee and Cipolla, [1], because it has simple calculations and only a twofold ambiguity. It is an approximation of the perspective projection as the depth of the object is very small compared with the distance between the camera and object.

A new approach for estimating 3D-head pose from a monocular image is proposed in this paper. Our algorithm is different from previous work by Horprasert et al. [1] that employs projective invariance of the cross-ratios of the eye-corners and anthropometric statistics to estimate the head yaw, roll and pitch. Five points, namely the four eye corners and tip of nose, are used there for head orientation estimation. They also made an assumption that four eye corners are co-linear but this is not true in general. In this paper, we use four relatively stable feature points. Our approach employs general prior

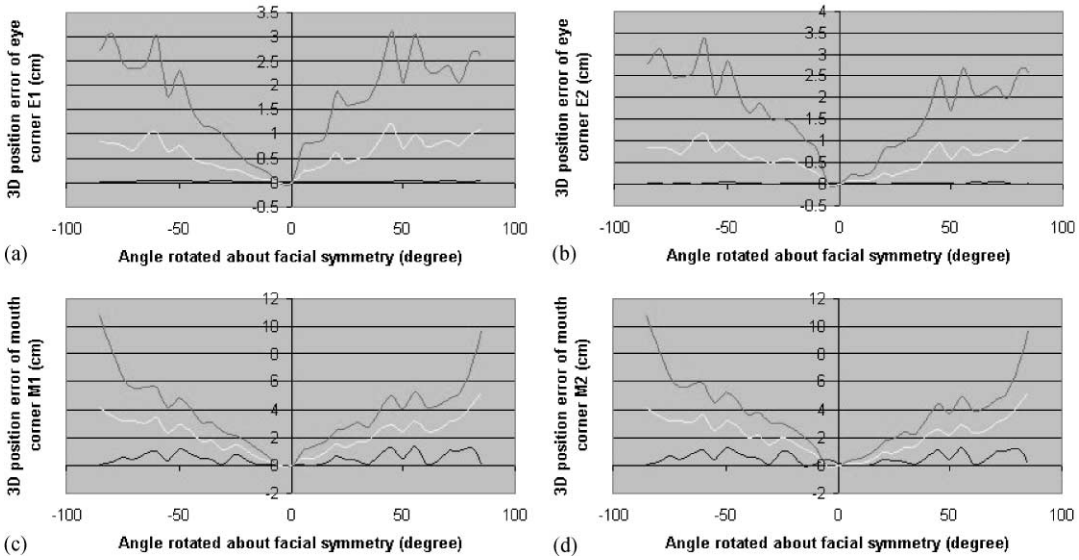


Fig. 25. 3D position errors of the four corners. Adding 0.5 cm to the mouth corner M_1. The ratio (1.98) and one of the two lengths are used. The value of the perturbation is 1 pixel; the distance between the original facial plane and the image plane is 60 cm; 100 simulation results are generated and averaged; top: maximum error; middle: average error; bottom: minimum error. (a) Point E_1, (b) point E_2, (c) point M_1, (d) point M_2.

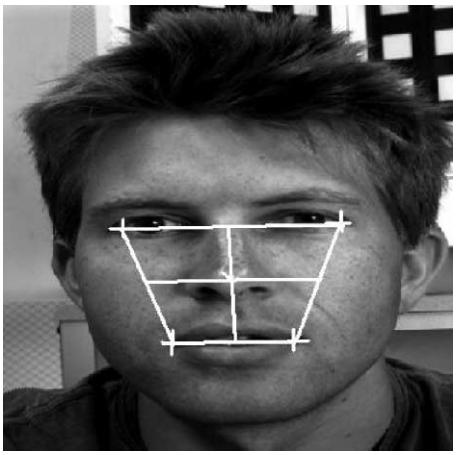


Fig. 26. Experiment results on the image of Herve's head, re-projecting the resulting 3D coordinates of the four corners, midpoints of arbitrary grouped two neighboring corners and the facial normal onto the original image. The facial normal is marked as a vector that starts at the center of the facial plane formed by the four corners.

knowledge of face structure and the corresponding geometrical constraints provided by the location of vanishing point to determine pose of human faces, both the 3D position and orientation of the head are determined from a monocular view. Eye-lines and mouth-line are assumed parallel in 3D space, and the vanishing point formed by them in the image can be used to infer 3D pose of human face. Perspective projection imaging model is used and an accurate analytic solution of 3D poses (position and orientation) of human faces is deduced. The orientation of the facial plane and relative 3D positions of the corners can be determined when the ratio of the length of eye-line and mouth-line segments is known. Furthermore, 3D absolute position of the feature corners can be determined if one of the two lengths is given. The experiments on a set of frontal images shown that the ratio we used is a stable value for the variances of people, also stable for the variances of the facial expression. The robustness analysis of the algorithm with synthetic data and real face image are enclosed.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful comments.

References

[1] T. Horprasert, Y. Yacoob, L.S. Davis, Computing 3-D head orientation from a monocular image sequence, Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, 1996, pp. 242–247.

[2] A. Gee, R. Cipolla, Fast visual tracking by temporal consensus, Image Vision Comput. 14 (1996) 105–114.

[3] D.P. Huttenlolocher, S. Ullman, Recognizing solid objects by allignment with an image, Int. J. Comput. Vision 5 (2) (1990) 195–212.
[4] A. Gee, R. Cipolla, Estimating gaze from a single view of a face, Proceedings of the IEEE International Conference on Pattern Recognition and Computer Vision, 1994, pp. 758–760.
[5] Shinn-Ying Ho, Hui-Ling Huang, An analytic solution for the pose determination of human face from a monocular image, Pattern Recognition Lett. 19 (1998) 1045–1054.
[6] A. Azarbayejani, T. Staarner, B. Horowitz, A. Pentland, Visually controlled graphics, IEEE Trans. Pattern Anal. Mach. Intell. 15 (6) (1993) 602–605.
[7] J. Heinzamann, A. Zelinsky, 3-D facial pose and gaze point estimation using a robust real-time tracking paradigm, Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, 1998, pp. 142–147.
[8] J.-B. Huang, Zen Chen, Jenn_yin Lin, A study on the dual vanishing point property, Pattern Recognition 32 (12) (1999) 2029–2039.
[9] L.-L. Wang, W.-H. Tsai, Camera calibration by vanishing lines for 3-D computer vision, IEEE Trans. Pattern Anal. Mach. Intell. 13 (4) (1991) 370–376.
[10] P. Parodi, G. Piccioli, 3D shape reconstruction by using vanishing points, IEEE Trans. Pattern Anal. Mach. Intell. 18 (2) (1996) 211–217.
[11] B. Brillault, O. Mahony, New method for vanishing points detection, CVGIP: Image Understanding 54 (2) (1991) 289–300.
[12] M.J. Magee, J.K. Aggarwal, Determining vanishing points from perspective images, Comput. Vision Graphics Image Process. 26 (1984) 256–267.
[13] M. Straforni, C. Coelho, M. Campani, V. Torre, Recovery and understanding of a line drawing from indoor scenes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 298–303.
[14] M.A. Fischler, R. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.
[15] D. Oberkampf, D.F. Dementhor, L.S. Davis, Iterative pose estimation using coplanar feature points, Comput. Vision and Image Understanding 63 (3) (1996) 495–511.
[16] R.M. Haralick, Determining camera parameters from the perspective projection of a rectangle, Pattern Recognition 22 (3) (1989) 225–230.
[17] WWW-b, University of Stirling, Face database: http://pics.psych.stir.ac.uk/cgi-bin/PICS/find-image.cgi.
[18] K. Aizawa, H. Harashima, Model-based analysis synthesis image coding (MBASIC) system for a person face, Signal Process.: Image Commun. 1 (1989) 139–152.
[19] T. Akimoto, Y. Suenaga, R.S. Wallace, Automatic creation of 3D facial models, IEEE Comput. Graphics Appl. 13 (1993) 16–22.
[20] T.S. Huang, L. Tang, 3-D face modeling and its applications, Int. J. Pattern Recognition Artif. Intell. 10 (5) (1996) 491–520.
[21] R. Lengagne, J.-P. Tarel, O. Monga, From 2D images to 3D face geometry, Proceedings of IEEE Second International Conference on Automatic Face Gesture Recognition (FG'96), Killington, USA, October 14–16 1996.
[22] C. Braccini, S. Curinga, A. Grattarola, F. Lavagetto, 3D Modeling heads from multiple views, in: G. Vernazz, A.N. Venetsanopoulos, C. Braccini (Eds.), Image Processing: Theory and Applications, Elsevier Science, Amsterdam, 1993, pp. 131–136.
[23] Jian-Gang Wang, E. Sung, Frontal view face detection and facial feature extraction using color and morphological operations, Pattern Recognition Lett. 20 (10) (1999) 1053–1068.
[24] Jian-Gang Wang, Y.F. Li, Human assisted environment modeling for robots, Autonomous Robots 6 (1) (1999) 89–103.
[25] WWW-a, Demo for stereo calibration, http://www-sop.inria.fr/robotvis/personnel/zzhang/CalibEnv/CalibEnv.html.
[26] S. Ganapathy, Decomposition of transformation matrices for robot vision, Proceedings of the IEEE Conference on Robotics and Automation, Atlanta, CA, 1984, pp. 130–139.
[27] A.L. Yuille, P.W. Hallinan, D.S. Cohen, Feature extraction from faces using deformable temples, Int. J. Comput. Vision 8 (2) (1992) 99–111.
[28] L. Quan, R. Mohr, Matching perspective image using geometric constraints and perceptual grouping, Proceedings of the International Conference on Computer Vision, Tampa, FL, December, 1988, pp. 679–684.

About the Author—JIAN-GANG WANG received the B.E. degree in Computer Science in 1985 from Inner Mongolia University, China. In 1988, he received the M.E. degree in Pattern Recognition & Machine Intelligence from Shenyang Institute of Automation, Chinese Academy of Sciences. From 1988 to 1997, he was with the Robotics Laboratory, Shenyang Institute of Automation, Chinese Academy of Sciences, where he was appointed associate professor in 1995. During the academic year 1997 to 1998, he was a research assistant at the Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong. In June 1998, he joined the School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore, where he is now pursuing the Ph.D. degree in Computer Vision. His research interests include 3D computer vision, face recognition, autonomous robots, human–machine interaction and virtual reality.

About the Author—ERIC SUNG graduated from the University of Singapore with a B.E. (Honours Class 1) in 1971 and then obtained his MSEE in 1973 from the University of Wisconsin. In 1999, he obtained his Ph.D. from NTU and his thesis is on structure from motion from image sequences. He lectured in the Electrical Engineering Department of the Singapore Polytechnic from 1973 to 1978. Subjects taught include Control Engineering and Industrial Electronics. In 1975, he was sent on a one-year industrial attachment at the Singapore Senoko Power Station. From June 1978 till April 1985, Eric Sung worked in design laboratories in multinational organisations such as Philips (Video), Luxor and King Radio Corporation designing television and microprocessor-based communication products. Joining Nanyang Technological University in April 1985, he is presently an associate professor in the Division of Control and Instrumentation of the School of Electrical and Electronic Engineering. His research interests include stereo vision, structure from motion, face recognition, digital image watermarking and robot-vision for autonomous applications.