Determining pose of a human face from a single monocular image

Jian-Gang Wang[1], Eric Sung[2], Ronda Venkateswarlu[1]

[1]Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
[2]Nanyang Technological University, Singapore 639798
{jgwang, vronda}@i2r.a-star.edu.sg, eericsung@ntu.edu.sg

Abstract

A new approach for estimating 3D head pose form a monocular image is proposed. Our 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. Eye-lines and mouth-line are assumed parallel in 3D space, and the vanishing point formed by the intersection of the eye-line and mouth-line in the image can be used to infer 3D orientation and location of human face. Perspective invariance of cross ratio and harmonic range is used to locate the vanishing point stably. The robustness analysis of the algorithm with synthesis data and real face images are enclosed.

1 Introduction

Two different transformations may be used for pose estimation from a single view: perspective or affine. The former one, e.g. [9], precisely models the actual projection of a 3D scene to the image plane. However the calibration of the camera is complex and could deliver up to a fourfold ambiguity in the estimation of the pose. Most existing real-time systems usually use affine transformation, e.g. [5], because it has simple calculations and is a good approximation of the perspective projection provided the depth of the object is small compared with the distance between the camera and object. This is usually the case in face tracking applications, but significant perspective distortion in the image will result when viewing the face from close range with a short focal length. Two-fold ambiguity resulted under affine assumption because it is based on the well-known three-point model developed by Huttenlocher and Ullman [11].

T. Horprasert et al [9] employs projective invariance of the cross-ratios of the four eye-corners and anthropometric statistics to estimate the head yaw, roll and pitch. Five points, namely the four eye corners and the tip of nose, are used. The four eye corners are assumed to be co-linear in 3D. This, however, is found not to be exactly true in general. Affine projection is assumed in [5, 4, 17, 8]. Gee et al [5] achieved a real-

BMVC 2003 doi:10.5244/C.17.11

time face tracker by utilizing simple feature trackers searching for the darkest pixel in the search window. The unique solution has to be searched by projecting both poses back into the image plane and measuring the goodness of the fit. In their earlier work, Gee and Cipolla [4] used five key feature points, nose tip, the outer eye corners and mouth corners, to estimate the facial orientation. The facial model is based on the ratios of four distances between these five relatively stable features, where the ratios were assumed not to change very much for different facial expressions.

The domain knowledge of human face structure can be advantageously used for pose estimation. In this paper, we study a novel approach that uses the vanishing point to derive a new and simple solution for measuring pose of human head from a calibrated monocular view. The vanishing point can be located using perspective invariance of cross ratio and the harmonic range. The 3D direction of eye-line and mouth-line can then be inferred from the vanishing point. Also, an analytic solution of orientation of the facial plane can be obtained when the ratio of the eye-line and mouth-line segments is given. Furthermore, the 3D coordinates of the four corners can be located if one of the lengths of the eye-line and mouth-line segments is given. Fischler and Bolles [2] revealed that a unique solution cannot be assured for the Perspective Four-Projection (P4P) problems. The P4P problem has a single theoretical solution [14, 6] when the coplanar points are in general configuration (no three co-linear scene points, non co-linear image points). Consequently, the solution of our algorithm is unique because the corner points we used are coplanar in 3D space, forming the facial plane. Vanishing point has been widely used in computer vision [10, 19, 15]. However, pose determination of the human face by using vanishing point is an unexplored approach.

2 Pose determination

2.1 Location of the vanishing point

In order to locate the vanishing point stably, we use perspective invariance of cross ratio and in particular the harmonic range.

An ordered set of collinear points is called a *range*, and the line passing through is called its axis. A range of four points, e.g. $\{A, B, C, D\}$ in Figure 1(a), is called a *harmonic range* if their cross ratio satisfies:

$$[A, B, C, D] = (AC/BC)/(AD/BD) = -1 \qquad (1)$$

Let $\{A, B, C, D\}$ be four image points in general position, see Figure 1(b). Let P be the intersection of AB and DC, Q the intersection of CB and DA. Such a set of six points $\{A, B, C, D, P, Q\}$ is called a *complete quadrilateral*.

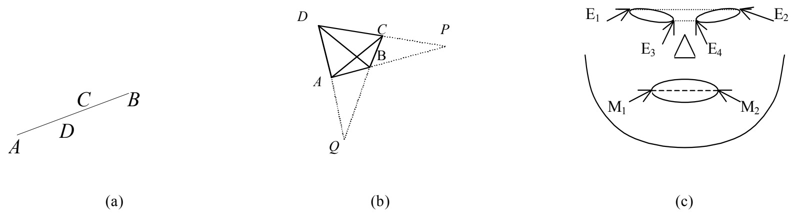


Figure 1: (a) Cross ratio; (b) complete quadrilateral; (c) eye and mouth lines.

Proposition Let I be the intersection of AC and BD, E and F be the intersections of IQ with AB and CD, respectively, see Figure 2. Then $\{P, F, D, C\}$ and $\{P, E, A, B\}$ are all harmonic ranges.

Proof of the proposition can be done based on the following propositions and theorem [12].

Proposition 1 A unique collineation is determined that maps four arbitrarily given image points in general position to four arbitrarily given image points in general position.

Proposition 2 Let D' and C' be distinct space point, and let F' be their midpoint, see Figure 2. If P' is the vanishing point of the space line passing through D' and C', then $\{P', F', D', C'\}$ is a harmonic range.

Theorem 1 The cross ratio is invariant under collineations.

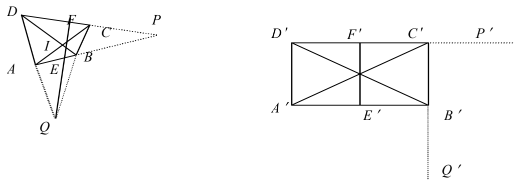


Figure 2: Mapping four arbitrarily given image points.

According to proposition 1, four points $\{A, B, C, D\}$ can be mapped to a rectangle by some collineation, see Figure 2. D', F', C', P' and Q' correspond to D, F, C, P and Q respectively. P' is a vanishing point, F' and E' is the midpoint of $D'C'$ and $A'B'$ respectively. $\{P', F', C', D'\}$ is harmonic range according to proposition 2, Hence, $\{P, F, D, C\}$ and $\{P, E, A, B\}$ become harmonic ranges according theorem 1, i.e.

$$[PFDC] = -1 \tag{2}$$

$$[PEAB] = -1 \tag{3}$$

Hence, the vanishing point P can be determined using (2) or (3).

The location of far-eye and mouth corners form the quadrilateral and its equivalent completion is depicted in Figure 1(c). Let E_1 corresponds to D, E_2 corresponds to C, M_1 corresponds to A, and M_2 corresponds to B. We can determine the vanishing point formed by the eye-line E_1E_2 and mouth-line M_1M_2.

Although two parallel lines are sufficient to compute the vanishing point, we use three sets of parallel lines. Three vanishing points will be detected due to image noise. Three points VP_1, VP_2, VP_3 are 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 Figure 3.

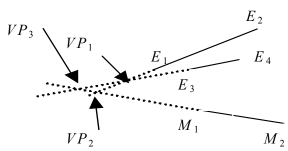


Figure 3: Three possible vanishing points formed by the intersections of the two eye lines and the mouth-line in the image.

Actually, we get a least-squares solution of the vanishing point using the above three parallel lines.

2.2 Pose determination

If the vanishing point is $P(u_\infty, v_\infty)$, (d_x, d_y, d_z) represent the 3D-direction vector of eye-lines and thus mouth-line (they are parallel, have the same direction vector). We have [7]:

$$(d_x, d_y, d_z) = \frac{1}{\sqrt{u_\infty^2 + v_\infty^2 + 1}} \begin{pmatrix} u_\infty \\ v_\infty \\ 1 \end{pmatrix} \tag{4}$$

Lets us assume (x_{e1}, y_{e1}) and (x_{e2}, y_{e2}) are normalized image coordinate ($f_x = f_y = 1$) of two far corners of eyes, $E_1(X_{e1}, Y_{e1}, Z_{e1})$ and $E_2(X_{e2}, Y_{e2}, Z_{e2})$ are their 3D coordinate respectively. (x_{m1}, y_{m1}) and (x_{m2}, y_{m2}) are normalized image coordinates of two corners of mouth, $M_1(X_{m1}, Y_{m1}, Z_{m1})$ and $M_2(X_{m2}, Y_{m2}, Z_{m2})$ are their 3D coordinate respectively.

$$X_{ei} = x_{ei}Z_{ei}, \qquad X_{mi} = x_{mi}Z_{mi} \tag{5}$$

$$Y_{ei} = y_{ei}Z_{ei}, \qquad Y_{mi} = y_{mi}Z_{mi} \tag{6}$$

$i = 1, 2..$

Generally speaking, the ratio (denoted as r) of the length of the eye-line segment (denoted as D_e) to the length of the mouth-line segment (denoted as D_m) is a more stable measure than the lengths themselves for different people. So the case that r is given instead of the lengths themselves is considered firstly. The ratio is immediately available from a frontal-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 corners and the orientation of the facial plane can be determined.

The above two cases are discussed as follows respectively.

Case 1: r is known

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 r. From (4), we arrive at

$$(X_{e2} - X_{e1})/D_e = d_x \tag{7}$$

$$(Y_{e2} - Y_{e1})/D_e = d_y \tag{8}$$

$$(Z_{e2} - Z_{e1})/D_e = d_z \tag{9}$$

$$(X_{m2} - X_{m1})/D_m = d_x \tag{10}$$

$$(Y_{m2} - Y_{m1})/D_m = d_y \tag{11}$$

$$(Z_{m2} - Z_{m1})/D_m = d_z \tag{12}$$

From the definition of r, we have,

$$r = D_e/D_m \tag{13}$$

From (7) to (13), we can obtain

$$Y_{e2} - Y_{e1} = (d_y/d_x)(X_{e2} - X_{e1}) \tag{14}$$

$$Z_{e2} - Z_{e1} = (d_z/d_x)(X_{e2} - X_{e1}) \tag{15}$$

$$X_{m2} - X_{m1} = (X_{e2} - X_{e1})/r \tag{16}$$

$$Y_{m2}\text{-}Y_{m1} = (Y_{e2}-Y_{e1})/r \quad (17)$$

$$Z_{m2}\text{-}Z_{m1} = (Z_{e2}-Z_{e1})/r \quad (18)$$

Let $X_{e2}-X_{e1} = 1$, so $Y_{e2}-Y_{e1}$, $Z_{e2}-Z_{e1}$, $X_{m2}-X_{m1}$, $Y_{m2}-Y_{m1}$, $Z_{m2}-Z_{m1}$ can be determined in turn using (14) to (18). This assumption means we can get only relative 3D positions of the feature corners. However the orientation can be obtained uniquely.

Replace X_{e1} and X_{e2} with (5), we have

$$X_{e2}-X_{e1} = x_{e2}Z_{e2}-x_{e1}Z_{e1} \quad (19)$$

Replace Y_{e1} and Y_{e2} with (6), we have

$$Y_{e2}-Y_{e1} = y_{e2}Z_{e2}-y_{e1}Z_{e1} \quad (20)$$

Z_{e1} and Z_{e2} can be solved from (19) and (20). Hence X_{e1}, X_{e2}, Y_{e1} and Y_{e2} can be solved using (5) and (6).

Similarly Z_{m1} and Z_{m2} are found using following equations:

$$x_{m2}Z_{m2}-x_{m1}Z_{m1} = X_{m2}-X_{m1} \quad (21)$$

$$y_{m2}Z_{m2}-y_{m1}3Z_{m1} = Y_{m2}-Y_{m1} \quad (22)$$

Hence X_{m1},X_{m2},Y_{m1} and Y_{m2} are found using (5) and (6).

Case 2: D_e or D_m and r are known

In this case, the absolute 3D positions of the feature corners can be determined. Let assume D_e is given. The Eqs of (7) to (9) described in the above case are still be used. Replace X_{e1} and X_{e2} in (7) with (5), we have

$$(x_{e2}Z_{e2}-x_{e1}Z_{e1})/D_e = d_x \quad (23)$$

From (9), we obtain

$$Z_{e1} = Z_{e2}-D_e d_z \quad (24)$$

From (23) and (24), we arrive at

$$(x_{e2}-x_{e1})Z_{e2}+D_e d_z x_{e1} = D_e d_x \quad (25)$$

Hence, we obtain

$$Z_{e2} = D_e(d_x-d_z x_{e1})/(x_{e2}-x_{e1}) \quad (26)$$

Replace Y_{e1} and Y_{e2} in (8) with (6), we have

$$(y_{e2}Z_{e2}-y_{e1}Z_{e1})/D_e = d_y \quad (27)$$

From (26) and (27), we obtain

$$(y_{e2}-y_{e1})Z_{e2}+D_e d_z y_{e1} = D_e d_y \quad (28)$$

Hence,

$$Z_{e2} =D_e(d_y-d_z y_{e1})/(y_{e2}-y_{e1}) \quad (29)$$

Z_{e2} and Z_{e1} can be calculated from (26) or (29) and (24).

Hence X_{e1}, X_{e2}, Y_{e1} and Y_{e2} are found using (5) and (6).

D_m can be obtained using r and D_e with (13). Similarly, the 3D coordinates of the mouth corners $M_1(X_{m1}, Y_{m1}, Z_{m1})$ and $M_2(X_{m2}, Y_{m2}, Z_{m2})$ can be calculated using their corresponding image coordinates then.

From the 3D coordinates of the four 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 line **n** as the cross product of the two space vectors $\mathbf{M_2E_2}$ and $\mathbf{M_2M_1}$ (see Figure 1(c)):

$$\mathbf{n} = \mathbf{M_2E_2}\times\mathbf{M_2M_1} \quad (30)$$

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: $\mathbf{M_2E_2}$ and $\mathbf{M_2M_1}$, $\mathbf{E_2E_1}$ and $\mathbf{E_2M_2}$, $\mathbf{E_1M_1}$ and $\mathbf{E_1E_2}$, $\mathbf{M_1M_2}$ and $\mathbf{M_1E_1}$.

2.3 Proof of the pose determination algorithm

Assume that the camera is fully calibrated and so the N-vectors of each of the corner points are known. (By definition, an N-vector is a unit vector that is in the direction starting from the optical center to the image point [12]). As such, we have $\mathbf{m}_{E1}$, $\mathbf{m}_{E2}$, $\mathbf{m}_{E3}$, $\mathbf{m}_{E4}$, $\mathbf{m}_{M1}$ and $\mathbf{m}_{M1}$ of the eye corners and mouth corners as in Figure 2, where $\mathbf{m}_{E1}$, $\mathbf{m}_{E2}$, $\mathbf{m}_{M1}$ and $\mathbf{m}_{M2}$ are N-vectors of the corners E_1, E_2, M_1, and M_2 respectively.

Four corners and the ratio are known. The direction of the eye-line and mouth line can also be inferred from the N-vectors of the far-eye and mouth corners. By hypothesis, lines E_1E_2, E_3E_4 and M_1M_2 are parallel in space. One can thus obtain the equations :

$$(k_{E1}\mathbf{m}_{E1} - k_{E2}\mathbf{m}_{E2}) = k_1\mathbf{n} \quad (31)$$

$$(k_{M1}\mathbf{m}_{M1} - k_{M2}\mathbf{m}_{M2}) = k_2\mathbf{n} \quad (32)$$

where k_{E1}, k_{E2}, k_{M1}, k_{M2}, k_1 and k_2 are scale constants. $\mathbf{n}$ represents the direction vector of the eye-line (and the mouth-line).
From (31) and (32), we obtain

$$(k_{M1}\mathbf{m}_{M1} - k_{M2}\mathbf{m}_{M2}) = (k_2/k_1)(k_{E1}\mathbf{m}_{E1} - k_{E2}\mathbf{m}_{E2}) \quad (33)$$

Let k_{E1}=1, we have

$$(\mathbf{m}_{M1}\ -\mathbf{m}_{M2}\ \ \mathbf{m}_{E2})((k_1/k_2)k_{M1}\ \ (k_1/k_2)k_{M2}\ \ k_{E2})^T = \mathbf{m}_{E1} \quad (34)$$

The ratio of the eye-line and mouth-line segments is r, we obtain

$$||k_{E1}\mathbf{m}_{E1} - k_{E2}\mathbf{m}_{E2}|| = r\, ||k_{M1}\mathbf{m}_{M1} - k_{M2}\mathbf{m}_{M2}|| \quad (35)$$

$$r = k_1/k_2 \quad (36)$$

We can get k_{E2}, k_{M1} and k_{M2} from (36) and (34).

We can calculate the facial normal $\mathbf{n}$ as the cross product of the two space vectors $\mathbf{M}_2\mathbf{E}_2$ and $\mathbf{M}_2\mathbf{M}_1$:

$$\mathbf{n} = \mathbf{M}_2\mathbf{E}_2 \times \mathbf{M}_2\mathbf{M}_1 \quad (37)$$

The orientation of the facial normal will be

$$(k_{E2}\mathbf{m}_{E2} - k_{M2}\mathbf{m}_{M2}) \times (k_{M1}\mathbf{m}_{M1} - k_{M2}\mathbf{m}_{M2}) \quad (38)$$

The relative position of the four feature corners can be determined using this method. This method can use a generic ratio of r and so is quite a general method. We do not need to measure r for each particular face. However, to justify our assumption, an experiment is conducted to understand the statistical nature of r. We need to ascertain that the standard deviation of r is indeed very small when a face image is with neutral or with slight expressions. This will be discussed in Section 2.4.

Six corners are known. Assume that the camera is fully calibrated and so the N-vector for each of the corner points is known. As such, we have $\mathbf{m}_{E1}$, $\mathbf{m}_{E2}$, $\mathbf{m}_{E3}$, $\mathbf{m}_{E4}$, $\mathbf{m}_{M1}$ and $\mathbf{m}_{M2}$ of the eye corners and mouth corners as in Figure 1(c). By hypothesis, lines E_1E_2, E_3E_4 and M_1M_2 are parallel in space. One can thus obtain the equations :

$$k_{E3}\mathbf{m}_{E3} - k_{E4}\mathbf{m}_{E4} = k_{E1}\mathbf{m}_{E1} - k_{E2}\mathbf{m}_{E2} \quad (39)$$

$$k_{M1}\mathbf{m}_{M1} - k_{M2}\mathbf{m}_{M2} = k_{E1}\mathbf{m}_{E1} - k_{E2}\mathbf{m}_{E2} \quad (40)$$

where the k values denote the (unknown) distances of the corner points to the optical center.

Being homogeneous, we can set $k_{E3} = k_{M1} = 1$. Thus we have 6 unknowns in 6 linear equations and hence a unique solution is obtained in general. Or one could obtain the same results as follows. For each of (39) and (40), take the cross product on both sides wrt $\mathbf{m}_{E1}$ and followed by the inner product wrt $\mathbf{m}_{E2}$ yielding :

$$(\mathbf{m}_{E3} \times \mathbf{m}_{E1})^T \mathbf{m}_{E2} - (k_{E4}\mathbf{m}_{E4} \times \mathbf{m}_{E1})^T \mathbf{m}_{E2} = 0 \quad (41)$$

$$(\mathbf{m}_{M1} \times \mathbf{m}_{E1})^T \mathbf{m}_{E2} - (k_{M2}\mathbf{m}_{M2} \times \mathbf{m}_{E1})^T \mathbf{m}_{E2} = 0 \quad (42)$$

Hence, k_{E4} and k_{M2} are solved. Consequently, the rest of the k values are retrieved from (39) and (40). This means all the corner points are found up to a scale factor. The normal to the plane is computed more robustly by finding the least squares plane from these six points. Euclidean position of the face is found once one of the k-values is known or, equivalently, the length of the mouth or eye corners are known. This could arise from stereo vision or by prior knowledge of person's face.

2.4 Variance of the ratio

It should be noted that this ratio with respect to face expressions is not invariant. The mouth can deform significantly in the horizontal direction and thus leading to significant variation in the ratio. For example, the variances of the ratio of a male from Stirling database [19] with different expressions are found to be: 2.01 for neutral, 1.94 for smiling, 2.0 for surprise, 1.89 for disgust. However, if the facial expression is slight, this ratio is still quite close to the fixed ratio that is set in our algorithm. Our experiments show that it can tolerate some deviations from this fixed ratio. In general, the ratio is almost a generic constant for neutral expression of faces. Base on our experiments on some database, e.g. [19], we can say that our algorithm is workable on the face images with neutral or slight expressions. The average ratio of the 300 images with neutral expressions is found to be 1.98, the standard derivation of the ratio is 0.02.

3 Experimental results

We have tried our algorithm on synthesized data and real image. The experiments show that our algorithm can provide a good estimation of pose of human head within a close distance.

3.1 Simulation using synthesis data

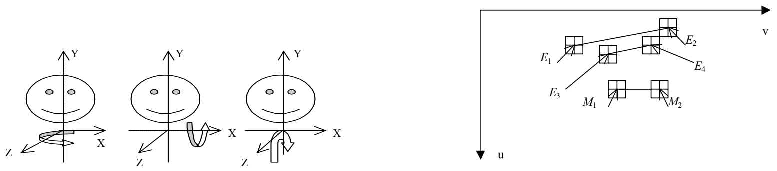
The synthetic data of different poses are produced as in Figure 4(a). The size of the test image is 512 × 512. The intrinsic and extrinsic parameters of the camera are as in Table 1, where (C_x, C_y, C_z) is the 3D world coordinate of the lens center, f_x and f_y are the scale factors of the camera along the x- and y-axis respectively. (u_0, v_0) are the coordinate of the principle point. We define a 3D rotation by three consecutive rotations around the coordinate axes, that is, a rotation by α degrees around the x-axis first, then a rotation by β degrees around the y-axis, and finally a rotation by γ degrees around the z-axis. The initial coordinates of the four corner points in the target face are set as:

$$E_1(-5.25, -6, Z),\ E_2(5.25, -6, Z),\ M_1(-2.65, -1, Z),\ M_2(2.65, -1, Z)$$

where Z is the distance along the z-axis from the origin to the face.

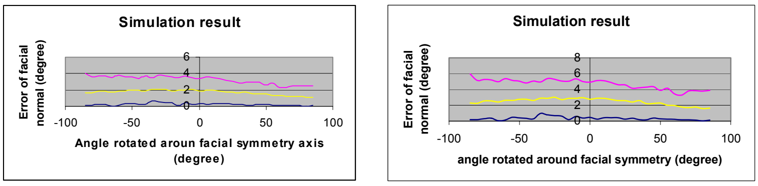
f_x	f_y	(u_0,v_0)	(C_x,C_y,C_z)	α	β	γ
1000	1000	(255, 255)	(0,0,50) cm	90^0	0^0	0^0

Table 1: Intrinsic and extrinsic parameters of the camera



(a) (b)

Figure 4: Simulations. (a) Simulations for different poses; (b) 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 within the corner-centered (2n+1)×(2n+1) window positions.



(a) (b)

Figure 5: Errors of the facial normal when the perturbation applied to the four corners is 1 pixel. Three curves, top: maximum errors; middle: mean errors; bottom: minimum errors. The distance between the original facial plane and the image plane is: (a) 50 cm; (b) 60 cm.

We show an example of the simulation pose results generated by rotating the facial plane about the facial symmetry (see Figure 4(a) for details), and the perturbation applied to the four eye and mouth corners of 1 pixel deviation each as in Figure 4(b). Four corner points rotated about the face symmetry axis from left -80^0 to 80^0 in steps of 5^0. 100 simulation results are generated and averaged. The errors of the facial normal estimation are shown in Figure 5. We can see also that the closer the camera gets to the human face, the more accurate the estimations and this can be seen by comparing Figure 5(a) and Figure 5(b). The error is found to be less than 2^0 when the distance between the original facial plane and the image plane is 60cm. The errors of the 3D positions of the four corners are shown in Figure 6. The degenerate case (when the facial plane is roughly parallel to image plane) occurs in the angle range of (-2^0, 2^0). The degenerate case is detected when the eye-lines and mouth line are nearly parallel in the image.

3.2 Experiment with real images

We have extensively evaluated our algorithm on video sequences of face images, as a person is moving his head in various directions in front of a PC. The experiments show that our method has a good performance and is robust for pose estimation of human head. The algorithm runs on a Pentium III 733M Hz PC in a speed about 25 frames per second. Some frames from a sequence of a subject are shown in Figure 7, where the facial normals of a face image are represented as arrows shown under the face image. Eye and mouth corners are tracked using template-matching technique.

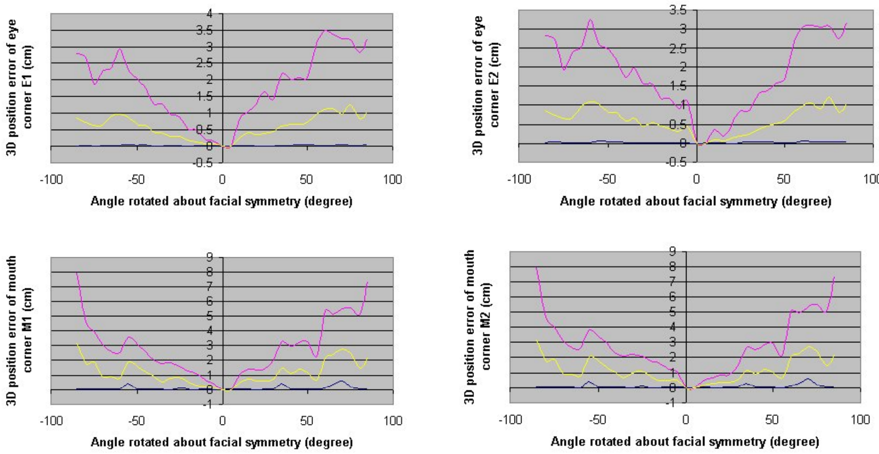


Figure 6: 3D position errors of the four corners, top left: E_1; top right: E_2; bottom left M_1; bottom right: M_2.

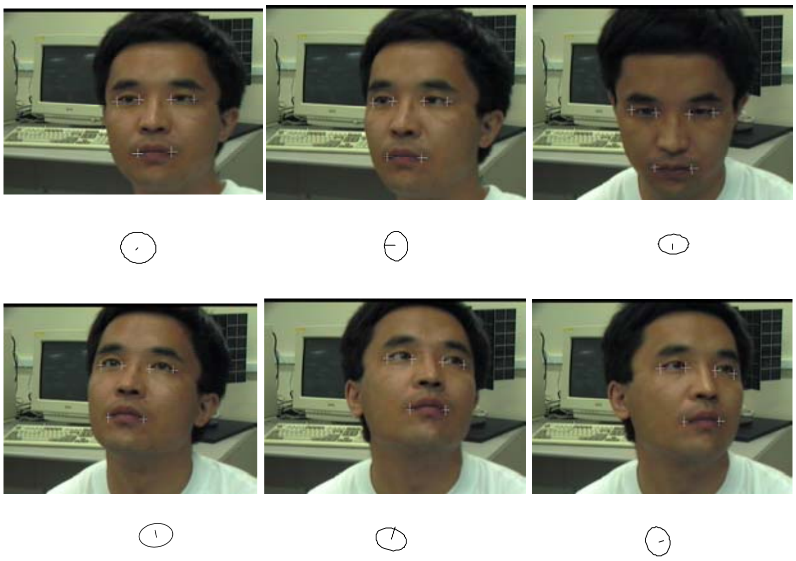


Figure 7: Pose determination results of some frames from a sequence.

4 Conclusion

In this paper, we have presented a new methodology for computation of head pose by using projective property of the vanishing point. The computation is quite light. An analytic solution has been inferred and the pose can be determined uniquely when the ratio of the length of the eye-line segment to the length of the mouth-line segment is known. Our algorithm is reliable because the ratio is found to be stable from face to face with neutral or slight expressions. 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 viable 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 than the use of the lengths themselves from face to face. Two alternative proofs of the vanishing-point method have been provided.

Accuracy of the vanishing point computation plays an important role on performance of the proposed method. The vanishing point can often be obtained from the image itself by some standard techniques [1, 13, 3, 16], and so making our algorithm practical. In situations where the distance between the face and camera is close, the full perspective model that we used can provide more accurate pose estimation than other existing methods that are based on the affine assumption.

References

[1] B. Brillault and O.Mahony. New method for vanishing points detection. *CVGIP: Image Understanding*, 54(2): 289-300, 1991.

[2] M. A. Fischler, R. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6): 381-395, 1981.

[3] A.C. Gallagher. A ground truth based vanishing point detection algorithm. *Pattern Recognition*, 35: 1527-1543, 2002.

[4] A. Gee and R. Cipolla. Estimating gaze from a single view of a face. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 758-760, 1994.

[5] A. Gee and R. Cipolla. Fast visual tracking by temporal consensus. *Image and Vision Computing*, 14: 105-114, 1996.

[6] R. M. Haralick. Determining camera parameters from the perspective projection of a rectangle. *Pattern Recognition*, 22(3): 225-230, 1989.

[7] R. M. Haralick, L. Shapiro. *Computer and Robot Vision*. Addison-wesley publishing company, 1993

[8] J. Heinzamann and Alexander Zelinsky. 3-D facial pose and gaze point estimation using a robust real-time tracking paradigm. In *Proc of the IEEE Conference on Automatic Face and Gesture Recognition*, pages 142-147, 1998

[9] T. Horprasert, Y. Yacoob and L. S. Davis. Computing 3-D Head orientation from a monocular image sequence. In *Proc of IEEE Conference on Automatic Face and Gesture Recognition*, pages 242-247, 1996.

[10] J.-B. Huang, Z. Chen and J.-Y. Lin. A study on the dual vanishing point property. *Pattern Recognition*, 32: 2029-2039, 1999.

[11] D. P. Huttenlolocher and S. Ullman. Recognizing solid objects by allignment with an image. *International Journal of Computer Vision*, 5(2): 195-212, 1990.

[12] K. Kanatani, *Geometric computation for machine vision*, Clarendon Press, Oxford, 1993.

[13] M. J. Magee and J. K. Aggarwal. Determining vanishing points from perspective images. *Computer Vision, Graphics and Image Processing*, 26: .256-267, 1984.

[14] D. Oberkampf, D. F. Dementhor and L. S. Davis. Iterative pose estimation using coplanar feature points. *CVGIP: Image Understanding*, 63(3): 1996.

[15] P. Parodi and G. Piccioli. 3D shape reconstruction by using vanishing points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(2): 211-217, 1996.

[16] L. Quan and R. Mohr. Matching perspective image using geometric constraints and perceptual grouping. In *International Conference on Computer Vision*, Tampa, FL, pages 679-684, Dec. 1988.

[17] B.Tordoff, W.W.Mayol, T.E.de Campos and D.W.Murry, Head pose estimation for wearable robot control. In *Proc of British Machine Vision Conference,* pages 807-816, 2002.

[18] University of Stirling, Face database: http://pics.psych.stir.ac.uk/cgi-bin/PICS/find-image.cgi

[19] L.-L. Wang and W.-H. Tsai. 'Camera calibration by vanishing lines for 3-D computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4): 370-376, 1991.