ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 2006. No. 8(72)
ELEKTRONIKA IR ELEKTROTECHNIKA

MEDICINE TECHNOLOGY
T 115
MEDICINOS TECHNOLOGIJA

Head Orientation Estimation using Characteristic Points of Face

D. Dervinis
Department of Electronics, Šiauliai University,
Vilniaus str. 141, LT – 76353 Šiauliai, Lithuania, e-mail: ddonatas@cr.su.lt

Introduction

Head pose determination represents an important area of research in human computer interaction (HCI). There are many researches in the area of estimation with monocular vision [1]. Methods for head pose estimation can be classified into two main categories: model-based and face property-based. Model-based use 3D model of the face and typically recover the face pose by first establishing 2-3D features correspondences and then solving for the face pose using the conventional pose estimation techniques. Property-based approaches assume there exists a unique causal-effect relationship between 3D face pose and certain properties of the facial image. Their goal is to determine the relationship from a large number of training images with known 3D face poses. They use neural networks to construct a mapping between 2D face images and 3D face poses [2], but, it cannot work well for previously unseen persons and backgrounds. Other proposed to use eigenspace for face detection and head pose estimation [3]. In summary, the property-based methods are simpler, but less accurate; many require a large number of training face image under different orientations.

Model-based methods usually start with feature detection, followed by matching 2D/3D corresponding features and determining face pose using the matched features. Among all facial features, the most commonly used eyes, nose, mouth [4,5].

In view this solution, we propose model-based method for estimation 3D head position. First, we created head model, where 3D rotation centre is located between eyes [Fig. 2]. Second, the algorithm for characteristic point of face detection is discussed. Third, the algorithm of detection of head rotation angles from the found points is discussed.

Head 3D pose estimation method

Head has six degrees of freedom: three rotation angles (Fig. 1) and three linear shifts. In general, a head rotation may be characterized by three Euler angles: roll, pitch and yaw (Fig. 1). Head 3D rotation matrixes **R** can be obtained from three single Euler rotations: angle around z (roll, $R_{z,\varphi}$) next around y (yaw, $R_{y,\Theta}$) and finally around x (pitch, $R_{x,\psi}$):

$$R_{z,\varphi}=\begin{bmatrix}\cos\varphi & -\sin\varphi & 0\\ \sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1\end{bmatrix};$$
$$R_{y,\Theta}=\begin{bmatrix}\cos\Theta & 0 & \sin\Theta\\ 0 & 1 & 0\\ -\sin\Theta & 0 & \cos\Theta\end{bmatrix}; \quad (1)$$
$$R_{x,\psi}=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\psi & -\sin\psi\\ 0 & \sin\psi & \cos\psi\end{bmatrix}.$$

Thus head 3D rotation matrixes **R** can be expressed:

$$\mathbf{R}=\begin{bmatrix}c\varphi\cdot c\Theta & c\varphi\cdot s\Theta\cdot s\psi-s\varphi\cdot c\psi & c\varphi\cdot s\Theta\cdot c\psi+s\varphi\cdot s\psi\\ s\varphi\cdot c\Theta & s\varphi\cdot s\Theta\cdot s\psi+c\varphi\cdot c\psi & s\varphi\cdot s\Theta\cdot c\psi-c\varphi\cdot s\psi\\ -s\Theta & c\Theta\cdot s\psi & c\Theta\cdot c\psi\end{bmatrix}, \quad (2)$$

here “s” is sin ; “c” is cos.

Local reference point and 3D rotation center of head $\{X_0,Y_0,Z_0\}$ centre is located between eyes. New local coordinate is recalculated from eyes coordinate (Z_0=0):

$$X_0=X_l+\frac{X_r-X_l}{2};\; Y_0=Y_l+\frac{Y_r-Y_l}{2}. \quad (3)$$

Head’s ground-start position is described by direction angle G, which is parallel to camera optical axis (Fig. 1). Coordinates are: $\{X_{pl}, Y_{pl}, Z_p\}$ – left eye, $\{X_{pr}, Y_{pr}, Z_{pr}\}$ – right eye and $\{X_{pn}, Y_{pn}, Z_{pn}\}$ – nose. After rotation characteristic points coordinates are $\{X_{fl}, Y_{fl}, Z_{fl}\}$, $\{X_{fr}, Y_{fr}, Z_{fr}\}$, $\{X_{fn}, Y_{fn}, Z_{fn}\}$.

Relation between ground-start position and final head position can be described by matrix **R:**

$$\mathbf{R}\cdot\begin{bmatrix}X_{px}\\Y_{px}\\Z_{px}\end{bmatrix}=\begin{bmatrix}X_{fx}\\Y_{fx}\\Z_{fx}\end{bmatrix}, \qquad (4)$$

here and forward $x \in$ *(l* – coordinate left eye*, r* – coordinate right eye*, n* – coordinate nose*).*

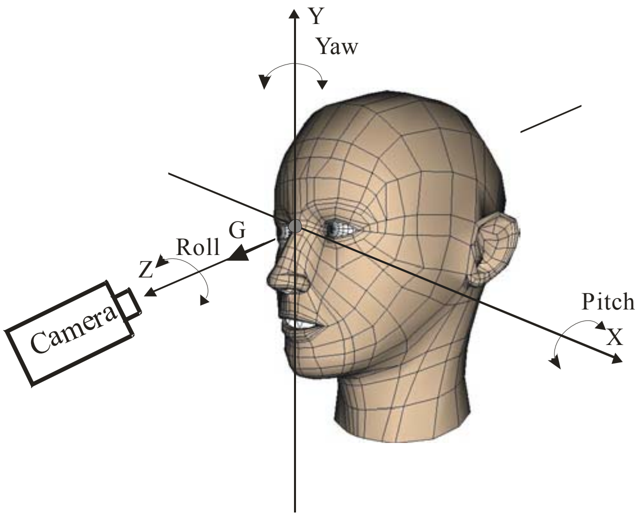

Fig. 1. The definition of three head rotation angles: roll, yaw, pitch and camera position

After rotation we have shifting between head and camera. Because of this reason we calculated head drift coefficient *d* by equation:

$$d=\frac{\sqrt{(x_{pr}-x_{pl})^2+(y_{pr}-y_{pl})^2}}{X_E}, \qquad (5)$$

here X_E – parameter, described in next section.

All coordinates are multiplied by *d*. Using obtained parameter *d* we can rewrite (4) as:

$$\mathbf{R}\cdot\begin{bmatrix}X_{px}\cdot d\\Y_{px}\cdot d\\Z_{px}\cdot d\end{bmatrix}=\begin{bmatrix}X_{fx}\cdot d\\Y_{fx}\cdot d\\Z_{fx}\cdot d\end{bmatrix}. \qquad (6)$$

If coordinates define over vector $\mathbf{v}_p = (X_{px}\cdot d, Y_{px}\cdot d, Z_{px}\cdot d)^T$ and , $\mathbf{v}_f = (X_{fx}\cdot d, Y_{fx}\cdot d, Z_{fx}\cdot d)^T$. Nine equations are naturally formed, but we have eliminated equations with final positions Z_{fl}, Z_{fr}, Z_{fn} – deepness coordinates, which are impossible to get from the final image. Equation (6) can be rewritten as:

$$\varepsilon_1=\sum_{j=1}^{3}\mathbf{R}_{1,j}\cdot\mathbf{v}_{pj}^{(i)}-\mathbf{v}_{f1}^{(i)}\,;\;\varepsilon_2=\sum_{j=1}^{3}\mathbf{R}_{1,j}\cdot\mathbf{v}_{pj}^{(i)}-\mathbf{v}_{f2}^{(i)}, \quad (7)$$

where i=1..3.

System of six equations can be solved using minimization least squares method:

$$E=\sum\varepsilon_1^2+\sum\varepsilon_2^2\,. \qquad (8)$$

After minimization least squares method (8) we obtained three angles *φ, Θ, ψ,* which characterized head 3D rotation.

Head model

To use our algorithm, we need to measure few face parameters and to create head model. It's start procedure of our method.

We selected next parameters: Z and Y coordinate of nose tip Z_n, Y_n , depth of eyes - Z_l, Z_r $(Z_l = Z_r)$ and distance between eyes - X_E (Fig. 2).

There was analyzed about 300 different head pictures. The needed parameters were evaluated statistically: $Z_n = 21$ mm, $Y_n = 41$ mm, $Z_{l,r} = 19.5$ mm, $X_E = 70$ mm.

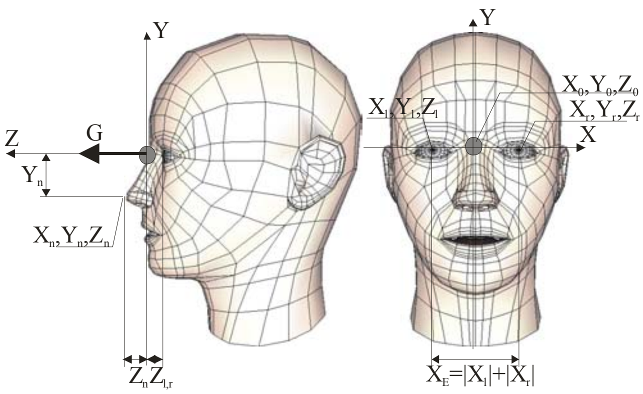

Fig. 2. Location of characteristics points and centre of rotation

Face detection and tracking

The quality of image is important for good detection and tracking. Our aim is create method which works with usual web camera. We use PC web camera (possible resolutions 240x120, 320x240, 640x480) and 24 bit color depth. Next important criterion is good and smooth lighting. Face should fill 60-90% of the view.

We used several methods to extract face and characteristic points (eyes, nose, lips): extracting characteristic face points from 2D image created by author [6], detection by skin color [7] and OpenCV's rapid object detection [8,9]. Fastest and best results showed OpenCV. For object detection OpenCV require training with positive and negative objects (eyes, nose) samples. OpenCV was training with 1500 positive and 5000 negative sample. Samples where collected from several database: positive from "Color FERET"[10] and "Cohn–Kanade AU–Coded Facial Expression Database"[11], negative - CorelDraw PhotoCD. For better detection OpenCV was improved by additional facial features [6]. Method's summary accuracy of eyes detection is about 94-96% (detecting pair of eyes), nose about 94%. OpenCV is used as primary eyes and nose coordinates detector. For continuously real time tracking the normalized correlation method was used. If error occurred during tracking by correlation the points detected by OpenCV is repeated (Fig. 3).

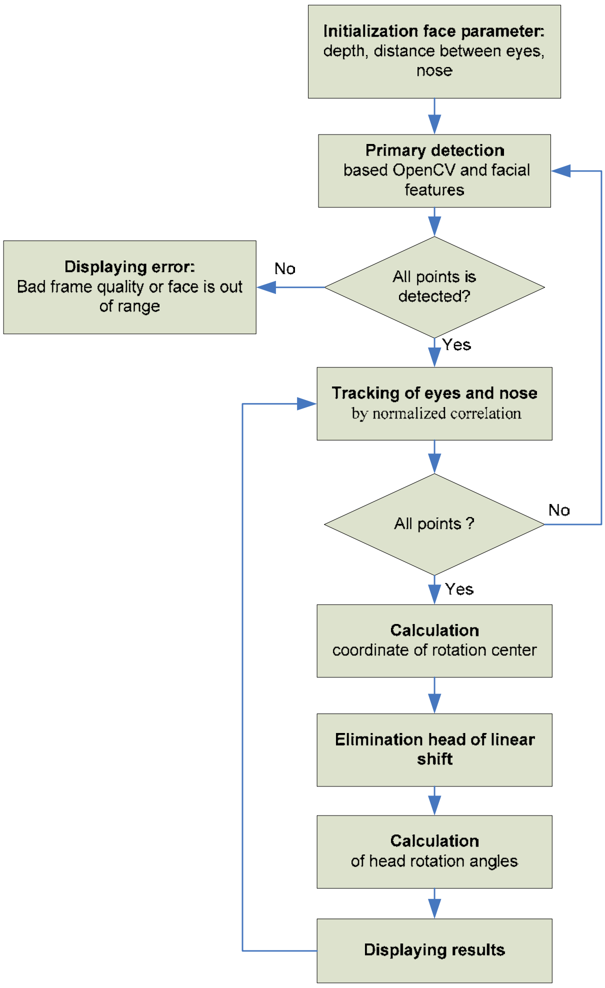

Fig. 3. Proposed head pose estimation algorithm using characteristic points of face

Results

To evaluate our algorithm, we used characteristic points coordinates generate by 3D head model. Finally coordinates are corrupted by Gausian noise with standard deviation 0.1, 0.5, 1.0 and 2.0. All rotation angles were equal to 10 degrees, and camera resolution was 240x120, 480x240 and 640x480. The estimated orientations were compared with calculated "ground-truth" orientation. Experimental results are shown in Fig. 4

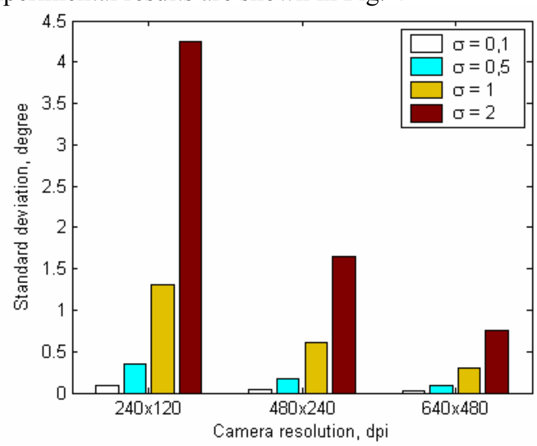

Fig. 4. Accuracy of estimation head angle under different camera resolution and noise level (σ), where angles $\varphi = 10^o$, $\Theta = 10^o$, $\psi = 10^o$

In Fig. 5 angles estimation absolute errors versus different rotations (Yaw, Pitch and Roll) are shown. If one of angles is changed, others are set on 10 degrees. Noise level σ=1, camera resolution 320x240. Remark that yaw angle less 0 degree is more unstable – absolute error approximately 1 degree, then above 0 degree absolute error is approximately 0.4 degree. Pitch and roll is stable in all range approximately 0.2 – 0.4 degree.

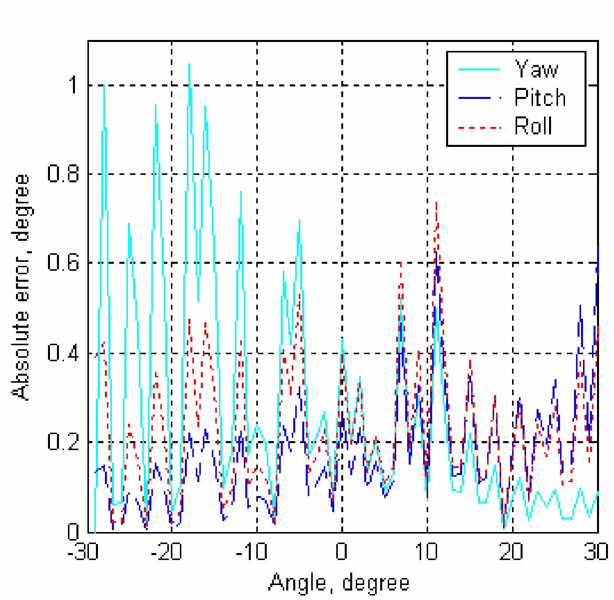

Fig. 5. Angles estimation absolute error versus different simultaneous rotations (other angle equal 10 degree) under noise level σ=1, camera resolution 320x240 dpi

Often in HCI a roll of head is secondary angle. Only two rotation angles are used: pitch and yaw. In Fig. 6 head gaze direction angle error versus yaw and pitch rotations under roll $\varphi=0^o$ are shown.

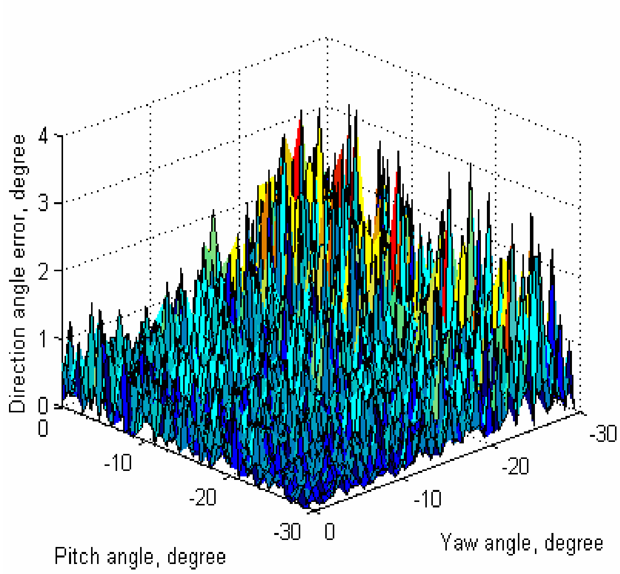

Fig. 6. 3D plot of the direction angle errors versus yaw and pitch rotations, under roll = 0^o, noise level σ = 1, camera resolution 320x240 dpi

Conclusions

We described 3D head pose estimation from a single monocular camera. Simulation of errors shows, that our method's sensitivity is high enough. Without noise under normal resolution (320x240) the estimation errors are less 0.3 degree, with noise level $\sigma=1$, error is less than 1 degree.

High accuracy and robustness of characteristic point detection is very important. We neglected this to face expression and it should be the further research direction.

Acknowledgement

This research is supported by the Lithuanian Science and Studies Foundation (Project title: Environment Control by Gaze System for the Disabled; Registration Number: V-06003).

References

1. **Qiang Ji.** 3D Face pose estimation and tracking from a monocular camera // Image Vision Computing. – 2002. – No. 20. – P. 499 – 511.
2. **Rae R., Ritter H.J.** Recognition of human head orientation based on artificial neural networks // IEEE Transactions on Neural Networks. – 1998. – No. 9. – P. 257 – 265.
3. **Darrell T., Moghaddam B., Pentland A.P.** Active face tracking and pose estimation in an interactive room // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – 1996.
4. **Gee A., Cipolla R.** Fast visual tracking by temporal consensus // Image and Vision Computing. – 1996. – No. 14. – P. 105 – 114.
5. **Horprasert A.T., Yacoob Y., Davis L.S.** Computing 3D head orientation from a monocular image sequence // Proceedings of SPIE 25th, AIPR Workshop: Emerging Applications of Computer Vision. – 1996. – No. 2962. – P. 244 – 252.
6. **Dervinis D.** Extracting characteristic face points from 2D image // Electronics and Electrical Engineering. – Kaunas: Technologija, 2004.– No.5(54). – P. 16 – 20. (in Lithuanian).
7. **Yang J., Waibel A.** A real-time face tracker // In Proceedings of the Third IEEE Workshop on Applications of Computer Vision. – Sarasota, FL,1996. – P. 142–147.
8. Open source computer vision library. Web links: http://www.intel.com/technology/computing/opencv/
9. **Qingcang Yu, Harry H.** Interactive Open Architecture Computer Vision//ICTAI'03, 15th IEEE International Conference on Tools with Artificial Intelligence. – 2003. – P. 406.
10. The **Color FERET** Database. Web links: http://www.nist.gov/humanid/colorferet/colorferet.html.
11. **Cohn–Kanade AU–Coded Facial Expression Database**. Web links: http://www.cs.cmu.edu/~face.

Submitted for publication 2006 06 15

D. Dervinis. Head Orientation Estimation using Characteristic Points of Face // Electronics and Electrical Engineering. – Kaunas: Technologija, 2006. – No. 8(72). – P. 61–64.

Model-based method for estimation 3D head position is proposed. We used monocular web camera (camera maximum resolution 640x480 dpi). First, we created head model, where 3D rotation center is located between eyes. Second, by method OpenCV characteristic points of face are detected and continuous real time tracking was realized by normalized correlation. Rotation of head is calculated using Euler rotation matrixes. Equation system is solved by minimization method. In result we got three angles φ (roll), Θ (yaw) and ψ (pitch), which characterized 3D head rotation The experimental results show sensitivity to image noise. If noise level with standard deviation $\sigma=1$ is added, the estimation errors are less than 1 degree, if $\sigma=2$ – less than 2 degrees (resolution 320x240 dpi). Ill. 6, bibl. 11 (in English; summaries in English, Russian and Lithuanian).

Д. Дервинис. Оценка углов поворота головы по характерным точкам лица // Электроника и электротехника. – Каунас: Технология, 2006. – № 8(72). – С. 61–64.

Предлагается метод оценки 3D положения головы монокулярной камерой (максимальная разрешающая способность камеры до 640x480 дпи.). Сначала создаётся модель головы, где центр вращения 3D идёт между глазами. Далее определяются характерные точки лица методом OpenCV. Непрерывное отслеживание точек в реальном времени реализуется вычислением нормализированной корреляции. Вращение головы расчитано по матрицам вращения Эйлера. В результате получаются углы 3D вращения головы: φ (крен), Θ (наклон) и ψ (тангаж). В работе определена чувствительность метода к шуму: если к координатам добавлен шум со стандартным отклонением $\sigma=1$, то ошибка не превышает 1 градуса, если с $\sigma=2$, – до 2 градусов (при 320x240 разрешающей способности камеры). Ил. 6, библ 11 (на английском языке; рефераты на английском, русском и литовском яз.).

D. Dervinis. Galvos posūkių nustatymas pagal būdingus veido taškus // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2006. – Nr. 8(72). – P. 61–64.

Aprašomas modeliu pagrįstas 3D galvos posūkio kampo nustatymo viena kamera būdas (maksimali skiriamoji geba 640x480 dpi). Pirmiausia sukuriamas galvos matematinis modelis, kur galvos 3D posūkio ašys eina per tašką, esantį tarp akių. OpenCV ir veido būdingų požymių radimo būdais nustatomi reikalingi būdingi taškai ir jų koordinatės. Toliau būdingi taškai sekami normalizuota koreliacija. Galvos posūkis apskaičiuojamas naudojant Oilerio posūkio matricas. Sudaromos lygčių sistemos iš pradinių taškų koordinačių ir po galvos posūkio gautų koordinačių. Sistema sprendžiama minimizavimo būdu. Gaunami trys galvos posūkio kampai φ (kampinis nukrypimas), Θ (posvyris) ir ψ (išilginis pokrypis), kurie apibūdina galvos 3D posūkį. Atlikus eksperimentus nustatytas būdo atsparumas triukšmui: jei pridedamas triukšmas, kurio standartinis nuokrypis $\sigma=1$, gaunami nuokrypiai, mažesnį už 1 laipsnį, jei $\sigma=2$ – mažesni nei 2 laipsniai (kameros skiriamoji geba 320x240 dpi). Il. 6, bibl. 11 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).