

Horizontal Human Face Pose Determination Using Pupils and Skin Region Positions

Shahrel A. Suandi^{1,2}, Tie Sing Tai¹, Shuichi Enokida², and Toshiaki Ejima²

¹ School of Electrical & Electronic Engineering, Universiti Sains Malaysia,
Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia
shahrel@eng.usm.my, tiesingtai@yahoo.com
<http://ee.eng.usm.my/eeacad/shahrel/index.html>

² Intelligence Media Laboratory, Department of Artificial Intelligence,
Kyushu Institute of Technology, Kawazu 680-4, Iizuka City,
Fukuoka Pref., 820-8502 Japan
{shahrel,enokida,toshi}@mickey.ai.kyutech.ac.jp
<http://www.mickey.ai.kyutech.ac.jp/>

Abstract. This paper describes a novel real-time technique to determine horizontal human face pose from a video color sequence. The idea underlying this technique is that when head is at an arbitrary pose to the right or left, there are significant relationships between the distance from center of both pupils to head center, and the distance between both pupils. From these distances, we compute a ratio known as "horizontal ratio". This ratio, besides being advantageous in the sense that it reduces the dependency on facial features tracking accuracy and robust to noise, is actually the quantity that is used to determine the horizontal human face pose. The technique is simple, computational cheap and requires only information that is usually retrievable from a face and facial feature tracker.

Keywords: Multiple view image and processing, tracking and motion, face pose, horizontal ratio, eyes and skin region.

1 Introduction

Face pose determination task has become one of the challenging tasks in face related researches. While being able to detect, locate and track a face and its facial features, one would expect some other factors such as face pose, gender, facial expressions and so on. These are some additional factors that create current computer vision systems to interact more intelligently with environments. As for face pose, it is an important task for some computer vision applications such as gaze tracking [1,2], human-computer interaction [3,4], monitoring driver alertness [5,6], best shot for face recognition [7] and multimedia retrieval [8]. To date, pose determination task can be categorized into active and passive methods. The former requires some special devices like sensors to be equipped on users face while the latter is usually vision-based method, non-intrusive and therefore, more preferable for human-computer interaction applications due to its convenience.

Apart from this, while passive method is gaining more attention from computer vision researchers, it can be generally classified into two primer approaches like being described in [9] – model-based approach and appearance-based approach. Model-based approach usually employs the 3D positions of facial features and recovers the pose by first, making assumptions on the projection, e.g., perspective projection, weak projection; second, deriving a few equations corresponding to the projection, and finally, computing the poses with respect to x , y and z axes by solving the equations derived. These poses are referred to as yaw, pitch and roll, respectively. Refs. [1],[10],[9],[2] are some of the examples employing this approach. On the other hand, appearance-based approach assumes that there is a unique causal-effect relationship between 3D face pose and certain facial image properties, for instance, appearances of facial features from frontal straight forward pose are different from frontal right pose. Refs. [11],[12],[13] are some of the examples from this appearance-based approach using Support Vector Machine(SVM), Boosting and Modular Eigenspaces, respectively.

In our work, we employ model-based approach. In contrast to some existing methods [1,10,14] that employ inner eyes corners and/or outer eyes corners for horizontal human face pose determination, we use pupils and face region for this purpose. Pupils are used due to the reason that it may also reveal the direction of gaze which can also be used for monitoring one's vigilance. These features are acquired from EMOTracker (Eyes and Mouth Tracker), which is developed by Suandi et al. [15]. The ideas underlying our proposed method is that when the head is at an arbitrary pose to the right or left, there are significant relationships between the distance from center of both pupils to head center and the distance between both eyes. Additionally, in order to achieve real-time system, no additional image processing or complex tasks are required even though only limited information (pupils, mouth corners and face region) are retrievable from the image. As a result, we have achieved a simple, computational cheap and real-time system.

From the detected pupils and face region, our system determines the horizontal face pose by executing three main tasks shown in Fig. 1.

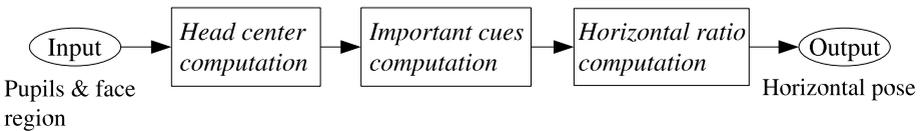


Fig. 1. Pose determination algorithms flow

Firstly, the head center is computed using nonlinear regression method by taking into account the ratio of pupils distance to pupil to skin edge (from the side where face is facing) distance. Secondly, the distance between both pupils (D_{pp}) and distance from pupils center to head center (D_{ch}) is computed. Subsequently,

“horizontal ratio (\mathcal{H})” is determined using D_{ch} and D_{pp} . Finally, the horizontal pose is determined from this ratio.

For the rest of this paper, we first describe the analysis results of anthropometric statistics in Section 2. Detail explanations on horizontal face pose determination are presented in Section 3. Experimental results are presented in Section 4 and the discussion on the observations are given in Section 5. Finally, the conclusion is described in Section 6.

2 Anthropometric Statistics

Our pose determination task is tailored using a monocular camera with weak projection assumption. It is well known that this approach, although simple, lacks the capability to estimate the depth of an object. Many efforts have been done to estimate the object pose by using monocular camera, for instance, by making assumptions on the 3D object shape and dimensions prior to estimating the pose, or some may require a sophisticated calibration technique. Similarly, we make assumptions on human 3D head shape by analyzing human anthropometrics statistics. Compared to others, ours only considers minimum features, pupils and face region. According to the work reported by Davis and Vaks [3], there are actually significant differences in the positions of human facial features and head geometry. This is supported by the report on anthropometric statistics which is written by Joseph W. Young [16]. In our work, we refer to this report and analyze the anthropometric statistics for United States of America (US) male. Results from the analysis are used to derive a “*mean model*” – a model made from the mean value of meaningful items; head circumference, head breadth and bipupils breadth. Using this model, further investigation on the relationship between head center and face pose is performed.

2.1 Analysis Results of US Male Citizens Anthropometric Data

Our analysis results show that merely three items are important in our work. These are head circumference, head breadth and bipupils breadth. The summary of these data is presented in Table 1. Statistics that have been considered are mean (μ), standard deviation (σ), minimum and maximum. Due to there is no head radius statistics data available in the report, we determine the head radius manually by computing this value from the mean head circumference. Let P denotes the head circumference and r denotes the head radius, P is given as $P = 2\pi r$. Therefore, the head diameter, d is $d = 2r = 182.88\text{mm}$. Our *mean model* is made from these data, in which head breadth is considered as the face region in image plane (in a forward straight frontal pose). This mean model is referred to as *head cylindrical model*, which is discussed further in Section 3.1.

Making assumptions that a head horizontal motion when viewed from top is a circular motion and eyes are located on this circle where the radius is equal to head radius, and when looking from forward straight ahead, the head diameter d (attributed from head radius, r) is always bigger than frontal face region width,

Table 1. Summary of anthropometric statistic data from 110 US adult citizens male

	μ ,mm	σ ,mm	min,mm	max,mm
Head Breadth, H	152.39	5.34	138.94	166.12
Head Circumference, P	574.54	16.22	541.02	620.01
Bipupil Breadth, D_{pp}	61.39	3.63	54.99	70.99

significant relationship between bipupils breadth (D_{pp}) and head diameter (d) can be given with the ratio of approximately 1:3 (61.39:182.88). Please refer to Fig. 3 for graphical representation of these values. r and d are invariant quantities since they are observed from top on the same person, whereas, D_{pp} is a variant quantity which relies on the face pose. From this observations, we yield geometrical relationship between D_{pp} and d , which can be given as $d = \gamma D_{pp}$, where, $\gamma = 3.0$. Using the mean model, we study how to compute the head center from head cylindrical model. This is explained in Section 3.1.

3 Horizontal Face Pose Determination

3.1 Head Center Computation

As the information that we have are too limited to compute the head depth, it is impossible to compute the head center. Therefore, we introduce “*head cylindrical model*” as the solution.

Head Cylindrical Model. Head cylindrical model (HCM) contributes in providing a reliable model to compute head center. Our ideal model of HCM is the mean model. It carries two main properties; invariant to head motions such as rotations and side-to-side motion – regardless what the face pose is, the head center shall remain at the same position with respect to face position; invariant to scale – when the face moves near or far from the camera, the head center shall remain at the same position with respect to face size.

Considering these two properties, head center is determined as follows:

1. As the only observable quantities are pupils and skin-like region (face region), we first compute these two quantities, X_0 and X_1 . X_0 is the distance between both pupils on image plane, which equals to D_{pp} , while X_1 is the distance from the pupil (on which side the face is facing) to the edge of face region on the same side. X_2 and X_3 , which are the distances from the face region edge to the HCM edge, and from the other side pupil to its side HCM edge, respectively, are determined indirectly from X_0 and X_1 . This is depicted in Fig. 2.
2. To handle scaling problems, next, we compute R_1, R_2 and R_3 , where each of these is defined as follows: $R_1 = \frac{X_1}{X_0}, R_2 = \frac{X_2}{X_0}$ and $R_3 = \frac{X_3}{X_0}$.

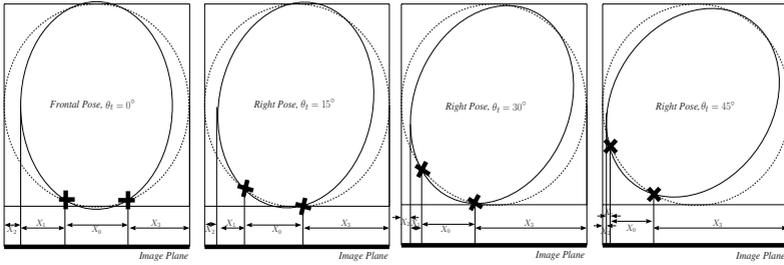


Fig. 2. Examples of HCM (mean model) viewed from top in four different horizontal poses, 0° , 15° , 30° and 45° from left to right. The observable quantities are only X_0 and X_1 . X_2 and X_3 are determined using nonlinear regression method. Notice that $X_0 = D_{pp}$.

3. Since only R_1 are determinable, we determine R_2 and R_3 by deriving each of them out using *nonlinear least square regression (NLLSR)* after R_1 is known. To establish the relationship between R_1 , R_2 and R_3 , the observations made from mean model which have been rotated to 15° , 30° and 45° are utilized.
4. Then, R_1 , R_2 and R_3 are computed. NLLSR provides the relationship between R_1 to R_2 , and R_1 to R_3 in terms of ratio, which are given in Eq. (1) and (2), respectively.

$$R_2 = 0.334552 - 0.0502925R_1 - 0.0253305R_1^{-1} \quad (1)$$

$$R_3 = 7.23527 - 25.4088R_1 + 33.7887R_1^2 - 14.6841R_1^3 \quad (2)$$

These equations show that when R_1 is known, then R_2 and R_3 may also be determined, which will consequently provide the values of X_2 and X_3 because $X_2 = R_2X_0$ and $X_3 = R_3X_0$.

5. When X_2 and X_3 are known, both edges of HCM are determined and finally, D_{ch} is computed.

Besides providing the head center, HCM is actually has the advantage to distinguish motions of rotations or side-to-side motion, that is, when changes in D_{ch} is observed, a rotation is happening, whereas, when changes is observed in x - or y -axes while at the same time there is no changes in D_{ch} , then it is a side-to-side motion. Such capabilities might be useful for monitoring one's vigilance systems.

3.2 Horizontal Ratio Computation

After computing the two important cues shown in Fig. 3, horizontal ratio, defined as \mathcal{H} in Eq. (3), is computed. It is the ratio of two cues, D_{ch} and D_{pp} , that are observable from image plane. It actually defines that profile face will provide an

infinite value of \mathcal{H} , whereas, frontal forward face will provide value of 0. Details on \mathcal{H} is presented below.

$$\mathcal{H} = \frac{D_{ch}}{D_{pp}} \tag{3}$$

Models to Define Horizontal Ratio. The model shown in Fig. 3 and 4 are actually derived from the HCM that has been introduced in Section 3.1. Making assumptions that a head is an ellipse when viewed from top and its rotation radius and distance between both eyes are r and d , respectively, two right triangles can be observed from top of the model when it rotates horizontally with respect to its origin O . The angle at time t is defined as θ_t and θ_{t0} is defined as the initial pose (frontal pose) where $\theta_{t0} = 0$. This is a requirement in the system in order for the system to compute and initialize individual parameters for the person it tracks. The two triangles are given as Triangle I and Triangle II. Values of r and d are the values that are always constant due to tracking is performed on the same person, whereas, the values of D_{ch} and D_{pp} are the observable values yielded from the image plane. Using the relationships between these two triangles, \mathcal{H} is defined as follows.

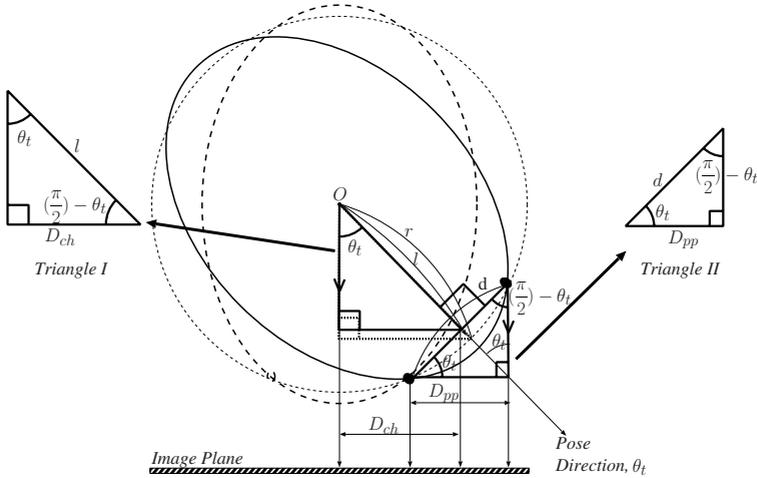


Fig. 3. HCM shown from top view. In the model are shown two triangles that can be drawn when the head moves horizontally to an arbitrary pose on the left. These relations are derived to compute the interrelation of face horizontal pose.

Triangle I. Triangle I is yielded only when variant in D_{ch} is observed. D_{ch} is defined as the distance from center of both eyes to head center on image plane, therefore, the range for this value is $0 \sim l$, which ranging from a frontal to a profile pose. Notice that $l < r$, where l can be computed indirectly from a frontal

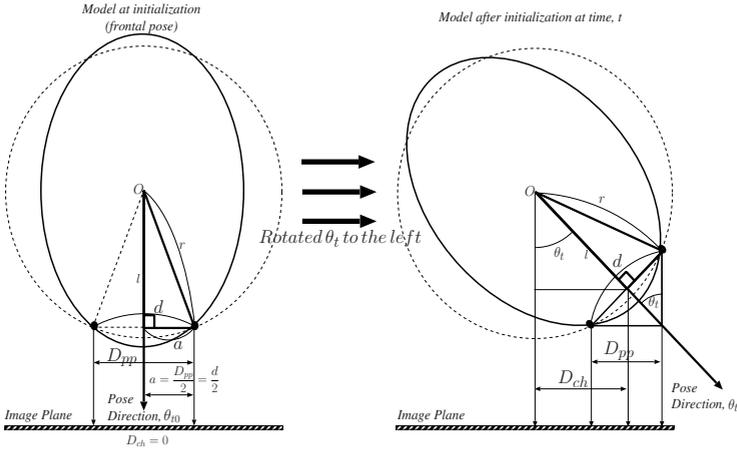


Fig. 4. Top view of HCM during initialization (left) and after rotated θ_t to the left (right). The right triangle shown in the above figure is pose invariant when observed from top view. From this model, r can be approximated from the results of analyzing anthropometrics statistic of human head and facial features. These information are then used to derive \mathcal{H} .

pose during initialization. For reading and reference convenient, we will refer to the quantities presented in Fig. 3 and 4 in the following mathematics definition.

The properties for the triangle are given as follows:

- l – this value is computed indirectly from the triangle observed during initialization. This triangle is shown in the left model from Fig. 4. r and a are given in Eq. (4). Since l and r are constant, once computed they are usable through out the tracking process until another initialization. γ in Eq. (4) is a coefficient that has been discussed in Section 2.1.

$$r = \frac{\gamma D_{pp}}{2} = \frac{\gamma d}{2} \quad \text{and} \quad a = \frac{D_{pp}}{2} = \frac{d}{2}. \quad (4)$$

During initialization, $d = D_{pp}$. From these two equations, we use theorem of Pythagoras to compute l .

$$r^2 = a^2 + l^2 = \left(\frac{D_{pp}}{2}\right)^2 + l^2 \quad \text{where,} \quad l = \sqrt{r^2 - \left(\frac{D_{pp}}{2}\right)^2} = \frac{D_{pp}}{2} \sqrt{\gamma^2 - 1}. \quad (5)$$

As l is a constant, we replace $D_{pp} = d$ in the Eq. (5) so that it is more appropriate. Therefore, we yield

$$l = \frac{d}{2} \sqrt{\gamma^2 - 1}. \quad (6)$$

- θ_t - This value defines the pose angle at time t . Knowing the value l as given in Eq. (6), the mathematical relationship yielded from Triangle I are given as follows:

$$\sin \theta_t = \frac{D_{pp}}{l} = \frac{2D_{pp}}{d\sqrt{\gamma^2 - 1}} . \tag{7}$$

Therefore, at initialization stage where $m = 0$, $\sin \theta_{t0} = 0$. Furthermore, computing θ_t from Eq. (7) yields,

$$\theta_t = \sin^{-1} \frac{2D_{pp}}{d\sqrt{\gamma^2 - 1}} . \tag{8}$$

To indicate left or right pose, positive (left) and negative (right) value are used.

Triangle II. In contrast to Triangle I, relationship observed in Triangle II is simpler to derive. This is due to all its quantities are directly computable from the model. d is a constant observed during initialization and n is the observable value on the image plane. From these quantities, we know that $\cos \theta_t = \frac{D_{pp}}{d}$. Therefore, θ_t can be computed from Eq. (9),

$$\theta_t = \cos^{-1} \frac{D_{pp}}{d} . \tag{9}$$

where, the given angle, θ_t is actually equal to the angle derived by Eq. (8).

Although it is shown that either using Eq. (8) or (9) can give the face pose, considering both of them simultaneously has been shown empirically to be more robust. Since we have sin and cos relation, we compute tan using these quantities.

$$\tan \theta_t = \frac{\sin \theta_t}{\cos \theta_t} = \frac{2D_{ch}}{d\sqrt{\gamma^2 - 1}} \div \frac{D_{pp}}{d} = \frac{2D_{ch}}{D_{pp}\sqrt{\gamma^2 - 1}} . \tag{10}$$

Rearranging Eq. (10), we yield

$$\frac{D_{ch}}{D_{pp}} = \frac{\sqrt{\gamma^2 - 1}}{2} \tan \theta_t . \tag{11}$$

As $\gamma = 3.0$, Eq. (11) becomes $\frac{D_{ch}}{D_{pp}} = \sqrt{2} \tan \theta_t$. The horizontal ratio, \mathcal{H} is defined as:

$$\mathcal{H} = \frac{D_{ch}}{D_{pp}} = \sqrt{2} \tan \theta_t . \tag{12}$$

Therefore, considering Eq. (12), yaw angle at time t , given as θ_t can be straight forward computed as follows:

$$\theta_t = \tan^{-1} \left(\frac{D_{ch}}{D_{pp}\sqrt{2}} \right) , \tag{13}$$

which means yaw angle (horizontal pose) can be computed by only using \tan^{-1} relation as shown in the Eq. (13).

Apart from this, using sin and cos relation from Eq. (8) and (9) may also reveal θ_t . However, unlike using \tan^{-1} relation, using neither sin nor cos relations is appropriate due to influence of noise during tracking. For example, when D_{ch} equals to a value that is greater than l due to noise, $\sin \theta_t$ will become $\sin \theta_t > 1$ which is not true. This reason can be applied to D_{pp} in Eq. (9) as well.

4 Experimental Results

Three main experiments have been carried out in this work. The first experiment is to evaluate the validity of horizontal ratio as the cue to represent horizontal face pose, the second and third experiments are to evaluate horizontal ratio in determining horizontal face pose manually and automatically, respectively, from video sequences. For automatic detection purpose, we use EMoTracker. All experiments are performed on a 2.2GHz Celeron CPU machine equipped with Linux OS. Figure 5 shows the database that have been used in our experiments.

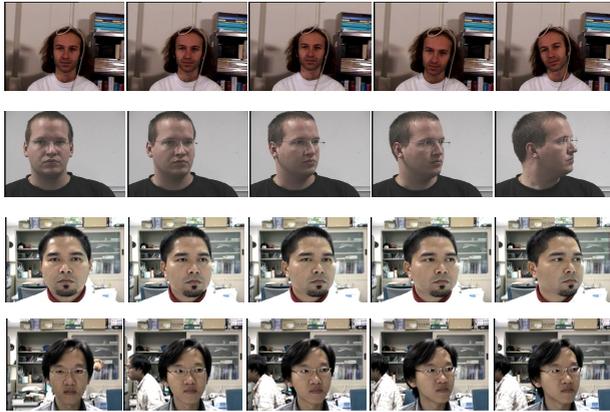


Fig. 5. Samples of database used in the experiments. Top to down row: Boston University Database, Pointing ICPR'04 Workshop Database and two video sequences taken at our lab.

4.1 \mathcal{H} Validity Check Experiment

In this experiment, database provided by Pointing'04 ICPR Workshop [17] is used. This database consists of 15 sets of images. Each set has another 2 different sets with 93 images. We only consider images within vertical 0° and horizontal $\pm 45^\circ$ from this database. From this data set, pupils and face region (left, top, right and bottom) positions are recorded manually. Then D_{ch} , D_{pp} , \mathcal{H} and θ_t are computed from the recorded positions. The results are summarized in Table 2.

Table 2. Results of evaluating \mathcal{H} to represent horizontal face pose, θ_t (in degree) using database provided by Pointing ICPR'04 Workshop

Face pose, θ_t	-45°	-30°	-15°	0°	15°	30°	45°
Mean, μ	-43.92	-29.02	-15.27	0.00	15.13	29.46	46.09
Std. Dev., σ	6.00	4.72	4.65	0.00	3.58	3.91	5.62

Table 3. Total of differences (in degree) between ground truth data and actual experiment data using Boston University Database

	<i>JAM5</i>	<i>JIM1</i>	<i>LLM8</i>	<i>SSM8</i>	<i>VAM8</i>
Mean, μ	-0.09	-0.55	-0.19	0.10	0.06
Std. Dev., σ	-0.15	0.23	0.84	0.75	0.85

For each face pose category, we compute the statistics for θ_t in terms of mean and standard deviation.

Analyzing the mean for each pose, we know that \mathcal{H} defined in our proposed method is feasible to determine θ_t . In fact, while considering the standard deviation results, it promotes that θ_t determination using \mathcal{H} is the best for frontal pose. But however, for other than this pose, the standard deviations are bigger than the standard deviation given for frontal pose but as overall, they are smaller than 15. This ensures us that the results given are within the range of $\pm 3^\circ \sim \pm 6^\circ$ and therefore, suggesting the validity of our proposed method.

4.2 Determining θ_t from Video Sequence Database - Manual Features Detection

For this experiment, we use video sequences database provided by Boston University [18]. This database provides the ground truth data to benchmark our proposed method. However, since we concentrate only on horizontal pose in uniform lighting condition, only some part of the data are used in the experiment. This database consists of five different video sequences, given as JAM5, JIM1, LLM8, SSM8 and VAM8. Each of them is taken from different subjects and contains about 200 frames. Similar to the preceding experiment, pupils and face region (left, top, right and bottom) positions have been recorded manually and then D_{ch} , D_{pp} , \mathcal{H} and θ_t are computed. Results of total difference (in mean) between ground truth and experimental data are shown in Table 3. The results show that the difference is very low, i.e. within $-0.15^\circ \sim 0.85^\circ$. When plotted into graphs, each of the results can be observed as in Fig. 6. Almost accurate results have been achieved. From this experiment, we can conclude that when good facial features are detected in the video sequence, good face pose angle can be determined.

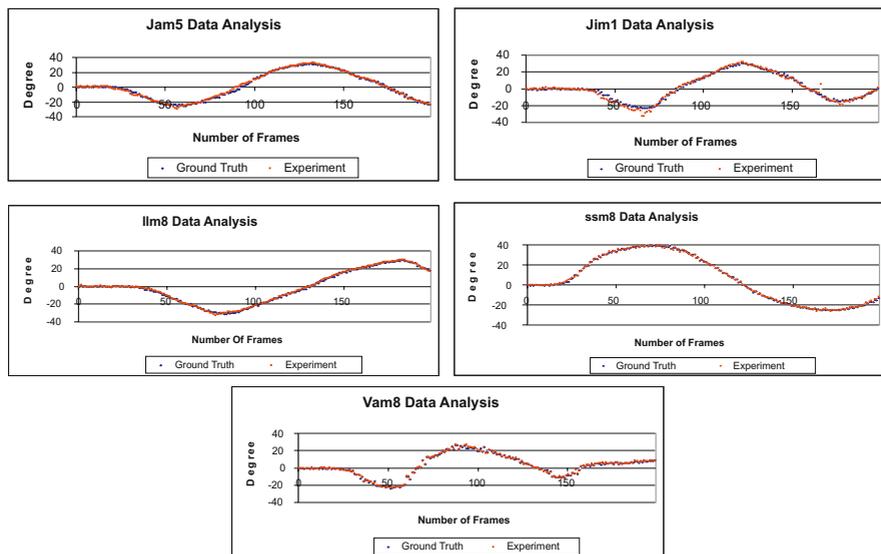


Fig. 6. Comparison between pose given in ground truth and pose computed using proposed method using Boston University Database

4.3 Determining θ_t from Video Sequence Database – Automatic Facial Features Detection

In this experiment, we use EMOTracker to automatically detect and track corresponding facial features. The main purpose of this experiment is to observe how the automatic detection results influence the face pose determination results. We prepare two video sequences taken at our lab as the data set. Each of them is taken from two different subjects and contains about 400 frames. These subjects were asked to start with a frontal pose and after a while, rotate their faces horizontally about one or two cycles (for example, left-right-left-right) followed by vertical motion about the same cycles. For comparison purpose, data for pupils and face region have been taken manually beforehand and θ_t is already computed from this manual database. Results given by the data taken manually and automatically are referred as “manual” and “auto”, respectively. These are shown in Fig. 7. Analyzing results for both subjects, we have observed that there is not much different between the ground truth and experiment data. An obvious difference can be observed from Subject 2 results when the pose is nearly to 45° . This is due to inconsistent pupils tracking using EMOTracker when face is within this range. Moreover, it is also difficult to track pupils within this pose range if the subject wears spectacle. We are currently in the stage of improving EMOTracker to solve this problem using separability filter [19,20,21]. For this particular experiment, we have achieved about 25 ~ 30fps rates for tracking and face pose determination using proposed method.

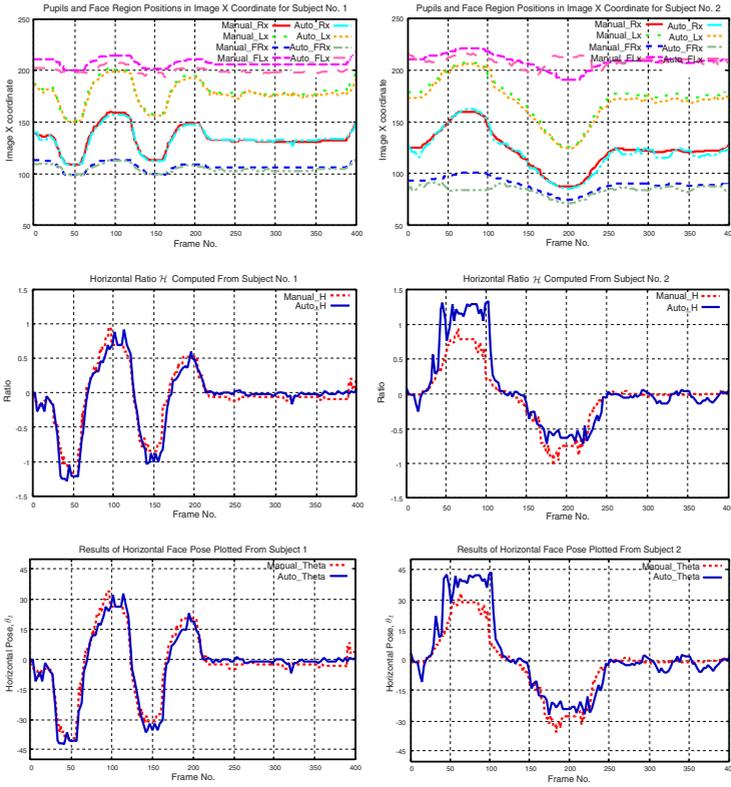


Fig. 7. From top row: manual and automatic pupils and face region data, horizontal ratio and horizontal face pose results plotted from subject 1 (left column) and 2 (right column)

5 Discussions

\mathcal{H} is generated using pupils and face region positions. As being described in Section 3.2, \mathcal{H} is defined from the values of D_{ch} and D_{pp} . Failing to detect the pupils precisely will cause false-positive D_{ch} and D_{pp} will be given to Eq. (3), which consequently affect the results. This also explains the disadvantage when D_{ch} or D_{pp} is considered independently to compute the pose using Eq. (8) or (9). Considering Eq. (8), a small observation noise, ΔD_{ch} , will contribute to large difference in θ_t due to $\frac{\partial \theta_t}{\partial D_{ch}} = \sec \theta_t$ curve characteristics. The same observation can be seen if Eq. (9) is considered as well, but with $\frac{\partial \theta_t}{\partial D_{pp}} = \csc^2 \theta_t$ curve characteristics. Whereas, while considering \mathcal{H} as the ratio of quantities, D_{ch} and D_{pp} , a small observation noise, ΔH , will contribute smaller difference in θ_t due to $\frac{\partial \theta_t}{\partial \mathcal{H}} = \cos^2 \theta_t$ curve characteristics. This confirms that using \mathcal{H} as defined in our work is robust against observation noise and therefore, appropriate for

this kind of framework. Besides, it also reduces the dependency on the tracking accuracy, which is one of the most difficult task in pose determination work.

6 Conclusions

A novel technique to determine horizontal human face pose from pupils and face region has been introduced in this paper. Considering the results of analyzing anthropometrics statistics data, we derive a model known as head cylindrical model and use this model to compute the head center. Head center is an additional information to compute a ratio known as horizontal ratio. This ratio is used to determine the face pose. Although the desired pose can be computed straight forward without taking the ratio into account, it has been shown that using ratio is more robust and capable of reducing the dependency on tracking accuracy. Comparison between truth and experiment data has also been performed, in which a very satisfactory results have been achieved. We have also encountered two major problems, inconsistent tracking when face pose is greater than 30° and when face region is not given as symmetrical during initialization. The solutions to these problems will be addressed in our future work.

Acknowledgements

This work is partially funded by Universiti Sains Malaysia short term grant.

References

1. Gee, A., Cipolla, R.: Determining the gaze of faces in images. *Image and Vision Computing* 12(10), 639–647 (1994)
2. Park, K.R., Lee, J.J., Kim, J.: Gaze position detection by computing the three dimensional facial positions and motions. *Pattern Recognition* 35(11), 2559–2569 (2002)
3. Davis, J.W., Vaks, S.: A perceptual user interface for recognizing head gesture acknowledgements. In: *PUI 2001: Proceedings of the 2001 workshop on Perceptive user interfaces*, pp. 1–7. ACM Press, New York (2001)
4. Heinzmann, J., Zelinsky, A.: Robust real-time face tracking and gesture recognition. In: *International Joint Conference on Artificial Intelligence, IJCAI 1997*, vol. 2, pp. 1525–1530 (1997)
5. Smith, P., Shah, M., da Vitoria Lobo, N.: Monitoring head/eye motion for driver alertness with one camera. In: *IEEE International Conference on Pattern Recognition (ICPR 2000)*, pp. 4636–4642. IEEE Computer Society Press, Los Alamitos (2000)
6. Ji, Q., Yang, X.: Real-time eye, gaze and face pose tracking for monitoring driver vigilance. *Real-Time Imaging* 8(5), 357–377 (2002)
7. Yang, Z., Ai, H., Wu, B., Lao, S., Cai, L.: Face pose estimation and its application in video shot selection. In: *IEEE International Conference on Pattern Recognition (ICPR 2004)*, vol. 1., pp. 322–325 (2004)

8. Garcia, C., Tziritas, G.: Face detection using quantized skin colour regions merging and wavelet packet analysis. *IEEE Transactions on Multimedia MM-1*, 264–277 (1999)
9. Ji, Q., Hu, R.: 3d face pose estimation and tracking from a monocular camera. *Image and Vision Computing* 20(7), 499–511 (2002)
10. Horprasert, T., Yacoob, Y., Davis, L.S.: Computing 3-d head orientation from a monocular image sequence. In: *IEEE International Conference on Automatic Face and Gesture Recognition (FGR 1996)*, pp. 242–247. IEEE Computer Society Press, Los Alamitos (1996)
11. Osuna, E., Freund, R., Girosit, F.: Training support vector machines: An application to face detection. In: *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 1997)*, pp. 130–136. IEEE Computer Society Press, Los Alamitos (1997)
12. Schneiderman, H.W.: Learning statistical structure for object detection. In: *Computer Analysis of Images and Pattern (CAIP)*, pp. 434–441. Springer, Heidelberg (2003)
13. Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 1994)*, IEEE Computer Society Press, Los Alamitos (1994)
14. Ho, S.Y., Huang, H.L.: An analytic solution for the pose determination of human faces from a monocular image. *Pattern Recognition Letters* 19(11), 1045–1054 (1998)
15. Suandi, S.A., Enokida, S., Ejima, T.: Emotracker: Eyes and mouth tracker based on energy minimization criterion. In: *4th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2004)*, IAPR, pp. 269–274 (2004)
16. Young, J.W.: Head and face anthropometry of adult u.s. citizens. Technical Report R0221201, Beta Research Inc. (1993)
17. Gourier, N., Hall, D., Crowley, J.L.: Estimating Face Orientation from Robust Detection of Salient Facial Features. In: *Proceedings of Pointing 2004, ICPR International Workshop on Visual Observation of Deictic Gestures* (2004)
18. Cascia, M.L., Sclaroff, S., Athitsos, V.: Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 22(4), 322–336 (2000)
19. Loy, C.Y., Suandi, S.A.: Precise pupils detection using separability filter. In: *International Conference on Robotics, Vision, Information and Signal Processing (ROVISP)* (to be published, 2007)
20. Fukui, K., Yamaguchi, O.: Facial feature point extraction method based on combination of shape extraction and pattern matching. *Systems and Computers in Japan* 29(6), 49–58 (1998)
21. Kawaguchi, T., Rizon, M.: Iris detection using intensity and edge information. *Pattern Recognition* 36(2), 549–562 (2003)