Exploiting the Golden Ratio on Human Faces for Head-Pose Estimation

Gianluca Fadda, Gian Luca Marcialis, Fabio Roli, and Luca Ghiani

Department of Electrical and Electronic Engineering,
University of Cagliari
Piazza d'Armi - 09123 Cagliari, Italy
{marcialis,roli,luca.ghiani}@diee.unica.it

Abstract. In this paper, a novel method for automatic head pose estimation is presented. This is based on a geometrical model of the head, in which basic features for estimating the pose consist in eyes and nose coordinates only. Worth noting, the majority of state-of-the-art approaches requires at least five features. The novelty of our work is the exploitation of the Vitruvian man's proportions and the related "Golden Ratio". The "Vitruvian man" is the well-known master-work by Leonardo Da Vinci, never used for automatic head pose estimation. Proposed method is compared by experiments with state-of-the-art ones, and shows a competitive performance although its simplicity and its low computational complexity.

1 Introduction

Head-pose estimation is based on the computation of three angles related to three reference axis, named yaw, pitch, and roll angles, which describe the rotation amount of the head along the three-dimensional space, with respect to an ideal frontal view.

In this case, brute-force approaches as machine learning-based ones, where a classifier is trained on several features from face images, is simply unsuitable. In fact, a lot of samples, covering different poses, are necessary. Moreover, head pose captured during system's operations can be significantly far from the ones used during training. This is due to the simple fact that, if no 3D information is available, estimating the direction according to 3D axes from 2D face images is a typical bad posed problem.

According to the taxonomy proposed in [1], head pose estimation approaches can be subdivided in: (1) appearance-template, where each face image is horizontally reverted and the symmetry degree with respect to the vertical axis along the xy reference plan must be assessed; (2) feature-tracking, where spatial features are extracted and tracked in a video-sequence. Experiments have shown that these methods are highly unreliable [2]; (3) moment-based, where a tessellation is projected on the face image. Among parts of the tessellation, three are selected which contain eyes and nose. From each part, a moment-based feature is extracted and used for detecting a discrete set of head poses [2]; (4)

A. Petrosino (Ed.): ICIAP 2013, Part I, LNCS 8156, pp. 280–289, 2013.

geometrical, which are based on the relative position of several features as eyes, nose and mouth.

The majority of head pose estimation methods relies on the computation of three angles named, respectively, Pitch, Yaw, Roll. Each one is related to the head rotation with respect of a reference axis in the 3D space. Thanks to these angles, it is possible to evaluate how much a certain estimated head pose is far from another one, by using appropriate distance functions [1].

Unfortunately, estimation of roll, yaw and pitch angles requires at least five features [1,3,6,7], for geometrical approaches, and this number strongly increases when considering other methods, as the ones based on neural networks or feature tracking [1]. To summarize current limits of state-of-the-art approaches [1]:

- Head pose is estimated by assuming an incremental variation of the facial position during video-sequence.
- Initial head pose in video-sequence is known.
- Scene calibration is required and several information must be extracted about the adopted camera, especially for the computation of the Pitch angle.
- All facial features are manually computed (e.g. eyes and nose position).
- The head is supposed to rotate around one axis at a time.

In this paper, we propose a novel method to automatic head pose estimation, where only three features, namely, eyes and nose locations are required. This is based on a novel model of the human head where geometrical relationships among above features are ruled by the concept of "Golden Ratio".

The "Golden Ratio" is the proportionality constant adopted by Leonardo Da Vinci in his master-work called "The Vitruvian Man". To the best of our knowledge, no work adopted the "Golden Ratio" for estimating the head pose. This method allows overcoming three limitations: (1) the use of a set of features smaller and easier to find, whose location cannot be necessarily exact; (2) no calibration of the scene is required, as well as camera parameters or characteristics; (3) it can be used for real-time applications, due to its low computational complexity.

Proposed method is tested on three benchmark data sets publicly available and adopted for comparing head pose estimation algorithms. All features are automatically computed, without human intervention. Reported results, in terms of well-known evaluation parameters, confirm that, although simpler, our approach is competitive with other state-of-the-art ones.

Paper is organized as follows. Section II describes the proposed model and algorithm. Section III reports some experimental results. Section IV concludes the paper.

2 The Proposed Model

The mathematical concept behind this work is the so-called "Golden Ratio". The "Golden Ratio" is the proportionality constant adopted by Leonardo Da Vinci in his master-work called "The Vitruvian Man" [4]. This is largely used in dentistry

Fig. 1. Face and eyes proportions based on the "Golden Ratio", from a picture by Leonardo da Vinci, 1488–9

and plastic surgery as it represents the ideal "perfection" and harmony of human proportions [5]. It easy to find several approaches to assess the human face beauty, based on the Vitruvian man's proportions. A detail on such proportions is reported in Fig. 1.

In few words, the concept consists in recursively subdividing a certain line, which may be represented by the duration/pitch of musical notes, or the size of physical objects.

The "Golden ratio" (*phi*) is a value which represents the proportionality constant of a line divided into two segments a and b, such that the whole line is to the longer a segment as the a segment is to the shorter b segment: $\phi = (a + b) : a = a : b$. By solving the related harmonic equation we obtain:

$$\phi = \frac{1 + \sqrt{5}}{2} = 1.61803399\ldots \tag{1}$$

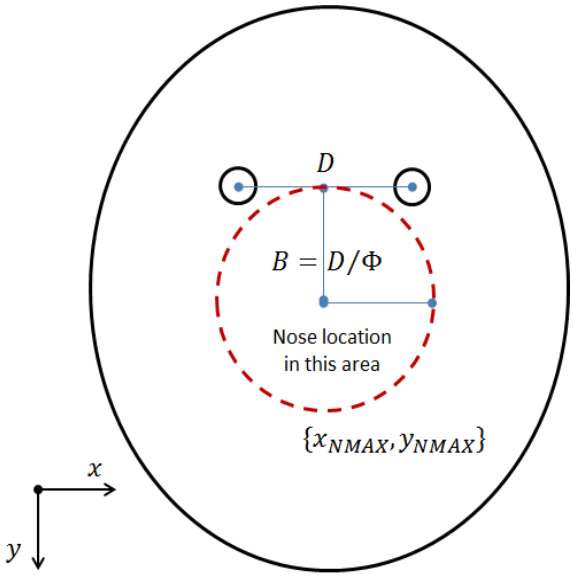

Fig. 2. Proposed model of a frontal head

This number, ϕ (*phi*), re-calling the initial letters of the sculptor Fidia, who used the "Golden Ratio" to create the Parthenon sculptures, is called "Golden Ratio".

Fig. 2 shows the proposed geometrical model of head. It is easy to see that this requires the eyes position (x-y coordinates). The "Golden Ratio" is exploited in locating the "ideal" nose location (x_{FN}, y_{FN}), according to the Vitruvian man's proportion. This point can be positioned along the line orthogonal with respect to the one jointing eyes, and passing on the middle point of that line (x', y'). The distance from this point is given by B, as shown in Fig. 2. Therefore, given the interocular distance D, we obtain $B = |y' - y_{FN}|$ that can be approximated by following Eq. (1):

$$\frac{D}{B} = \phi \Rightarrow B = \frac{D}{\phi} \approx 0.618D \tag{2}$$

Where: $\phi = 1.61803399\ldots$ is the "Golden Ratio".

By following the above assumption, three angles may be easily computed as follows.

Roll Angle Computation. According to Fig. 3, Roll angle is computed as:

$$Roll = arctg\left(\frac{dy}{dx}\right) \tag{3}$$

On the basis of dy sign, detectable roll angles are: $(-90° \div +90°)$.

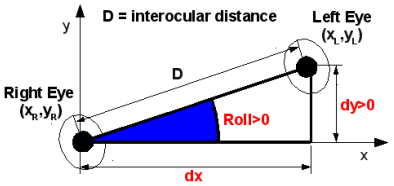

Fig. 3. Roll angle computation

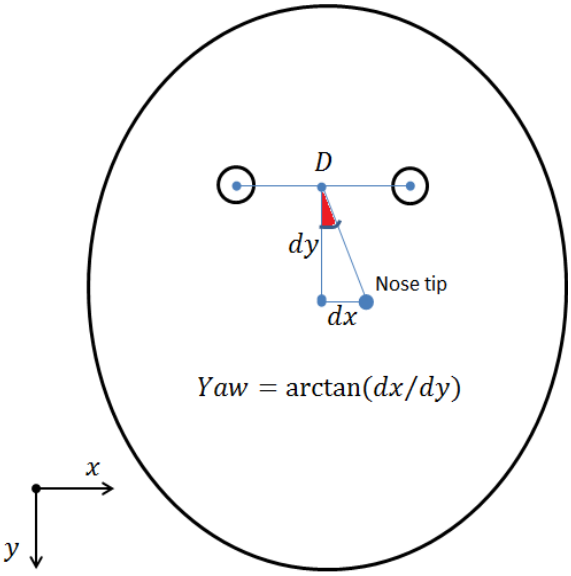

Fig. 4. Yaw angle computation

Yaw Angle Computation. Let (x_N, y_N) be the coordinates related to the nose location. According to Fig. 4, Yaw angle is computed as:

$$Yaw = arctg\left(\frac{dx}{dy}\right) \tag{4}$$

Pitch Angle Computation. Main innovation of this paper is the computation of Pitch angle, which is the most difficult to compute in the majority of approaches to head pose estimation, since it requires knowledge about the "depth" of the scene. In other words, 3D information, along the axis orthogonal to the image plan (xy).

State-of-the-art methods solve the problem of Pitch angle estimation by adding information about the used camera. Since it is necessary to know a reference distance between xy plan and another point along z axis, the usual solution is to consider the focal point of the camera. However, this approach makes the overall head pose estimation algorithm strongly dependent on the adopted hardware.

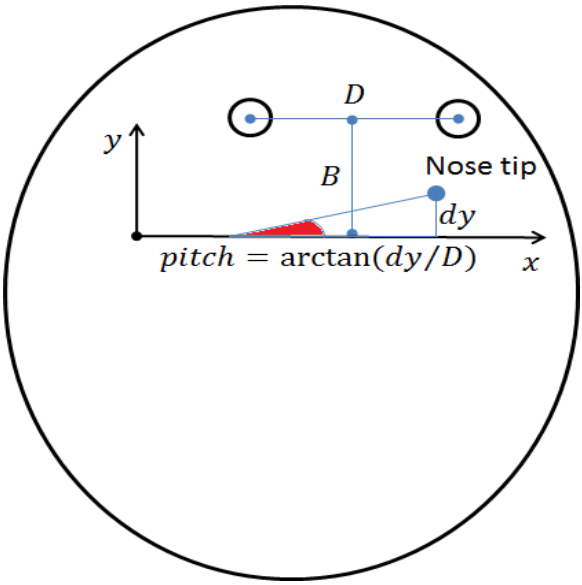

Fig. 5. Pitch angle computation

Solution proposed here overcome above limitations. Fig. 5 shows our face model, where the Pitch angle is given by:

$$Pitch = arctan\left(\frac{dy}{D}\right) \tag{5}$$

Maximum Angle Values. From equations above, it is possible to derive the maximum values of roll, pitch, and yaw angles which are tolerable by this approach.

- Roll angle. Depending on the sign of dy, detectable values are: $(-90° \div +90°)$.
- Yaw angle. Since the worst case corresponds to the nose location on the circumference of the circle depicted in Fig. 2, detectable angles are: $(-45° \div +45°)$.
- Pitch angle. As in the previous case, the maximum detectable values correspond to the nose location equal to (x_{NMAX}, y_{NMAX}). Accordingly, we obtain the interval: $(-31.716° \div +31.716°)$.

Since a non-frontal pose points out the loss of "harmony" with respect to the Golden ratio-based proportions, we expect that this method can be affected by an estimation error of the pose. In the next Section, we quantify this loss and compare it with some state-of-the-art algorithms.

3 Esperimental Results

Performance of the proposed system has been evaluated on three benchmark data sets:

- *Pointing '04 Head Pose Image Database* [10] is made up of 15 image galleries related to 15 different persons. Each gallery contains two sequences of 93 face images. Fifteen persons in the data set exhibit different characteristics in terms of age, eye glasses, skin colour. Since a quantitative performance of head pose estimation algorithms is necessary, a ground truth is given in terms of yaw and pitch angle, both discretized between $-90°$ and $+90°$ (thirteen values for yaw and nine values for pitch). The ground truth has been obtained by constraining captured subjects to look at several markers in the scene. Roll angle is not given.
- *Boston University (BU) Face tracking dataset* [3], is made up of 72 sequences taken from five subjects. First 45 sequences (9 per subject) have been captured under uniform lighting. Second 27 sequences (9 per a subset of 3 subjects) under different lighting variations. Each sequence is made up of 200 frames during which several free rotations and translations of the head have been taken. Ground truth is given in terms of Yaw, Roll and Pitch angles, evaluated through a magnetic sensor.
- *Head Pose and Eye Gaze (HPEG) Dataset* [8] is made up of 20 video-sequences subdivided in two different sessions. First one is studied for evaluate exactly the head pose; second one is aimed to the estimation of eye gaze. In each session, 10 subjects are captured. Different head rotations are allowed. Ground truth is given in terms of Yaw and Pitch angles, and computed thanks to a semi-automatic labelling process, based on three leds located around the subject's face.

All video-sequence frames are evaluated in order to assess the performance of the proposed method.

Systems we developed for head detection and facial feature localization is based on the Viola-Jones framework available on the OpenCV libraries [9]. Eyes and nose locations are evaluated using Viola-Jones classifiers esplicitly trained for the detection of these biometrics.

Starting from basic state-of-the-art algorithms, we assessed the estimation of the Yaw, Roll, and Pitch angles according to the formulas we showed in the previous Section.

Adopted evaluation parameters have been suggested in [1], and explained in the following.

Mean Absolute Error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |a_i - GT_i] \tag{6}$$

Table 1. MAE of the proposed method on images of the Pointing'04 data set using manual and automatic localization of eyes and nose. Results refer to all 35 poses and 15 individual on the Pointing'04 data set.

	Mean Absolute Error	
	Yaw	Pitch
Automatic localization	9.6°	13.6°
Manual localization	12.6°	13.7°

And Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(a_i - GT_i)^2} \tag{7}$$

Where a_i and GT_i are the considered angle (Yaw/Roll/Pitch) and the correspondent ground truth, and N is the overall number of frames.

Table 1 reports, first of all, an experiment aimed to see the difference of MAE parameter where nose and eyes are localized manually and automatically. Results are averaged on all images available in the Pointing'04 data set, that is, all 35 poses over 15 individuals are taken into account. It is possible to see that the Yaw angle is the most affected by a unaccurate estimation of eyes and nose, whilst no appreciable difference can be noticed on the Pitch angle. The Pointing'04 data set contains specific images for each range of possible poses, thus the assessment of results in simplified. Moreover, Pointing'04 data set is used as benchmark in the majority of papers on head pose estimation [1].

Tables 2-3 report the performance of our method and the one of other state-of-the-art approaches extracted from [1]. These evidences show that our simple method, based on a geometrical analysis of the face, and the Vitruvian man's proportions exploitation, is suitable for fine head pose estimation. Reported results clearly show that our method can be successfully used for this aim, and, in particular, allows an estimation degree more accurate than other state-of-the-art solutions [1]. It is worth to point out that state-of-the-art geometrical approaches perform much worse than ours, and this is the reason for which we did not report comparison with such approaches[1]. State-of-the-art methods reported here are based on much more complex approaches: in particular they exploit the video information by feature tracking, whilst out system estimates the head pose frame-by-frame. Therefore, whilst these methods can be effective only if a video-sequence is available, our method show superior or comparable performance by exploiting information extracted on still images. Since state-of-the-art algorithms are used to extract eyes positions, and that data sets used are made up of video-sequences, our method can be used in real time, as well as other reported ones.

Finally, it is worth to point out that adopted data sets exhibit a third-party ground truth. Thus, they allow the comparison and assessment of the methods

Table 2. Comparison of results on BU Face Tracking dataset

	Mean Absolute Error		
	Yaw	Pitch	Roll
Proposed Method	**5.5°**	**3.8°**	**3.4°**
La Cascia (Tracking) [3]	3.3°	6.1°	9.8°
Xiao (Tracking) [6]	3.8°	3.2°	1.4°

Table 3. Comparison of results on HPEG dataset

	Root Mean Square Error	
	Yaw	Pitch
Proposed Method	**7.45°**	**5.10°**
Asteriadis (Feat. Track. with Optical Flow) [7]	8.39°	5.51°
Asteriadis (Feat. Track. with Distance Vector Fields) [7]	6.65°	5.59°

performance. On the basis of reported results, estimated angles by our method are very near to this ground truth. With regard to the Yaw angle, error is less than 7 degrees, on average, and appears to be the most affected by a unaccurate estimation of eyes and nose. On the other hand, average error is less than for the Pitch angle, which is the most difficult to estimate and also the one requiring significant information especially for other state-of-the-art approaches. With regard to this point, it is worth remarking that our method does not require knowledge about the scene or the camera characteristics, thus reported results are noticeable.

4 Conclusions

In this paper, we presented a novel geometrical model for head pose estimation. Computation of Roll, Yaw and Pitch angles requires the location of eyes and nose, which is a number of feature much inferior than that required by other geometrical and non-geometrical methods. This low number of required features is due to the exploitation of the "Golden Ratio" among such features, which led, in particular, to a very effective estimation of the Pitch angle. In fact, estimating this angle requires 3D information which state-of-the-art approaches derive from the scene or the camera adopted. As a further advantage, no scene calibration is required.

Our method has been quantitatively evaluated on three benchmark data sets already used for fine head pose estimation. Method has been also compared with other non-geometrical approaches, resulting in a better performance on average, which is coupled with the low amount of computational complexity required.

Proposed algorithm suffers from the inexact computation of eyes and nose coordinates, as usual for all geometrical methods. However, on overall, it appears as quite stable, as shown here, where all feature have been automatically detected.

Finally, thanks to the our approach based on the "Golden Ratio", we have been able to derive the maximum estimation range allowable analytically, thus predicting the applications set for which our model can be suitable.

This preliminary set of experiments has shown that this geometrical approach is worth of further investigations. Future works will include extensive experiments on other data sets, and also possible countermeasures to correct estimation when nose and eyes can't be found reliably.

References

1. Murphy-Chutorian, E., Trivedi, M.M.: Head pose estimation in computer vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(4), 607–626 (2009)
2. Moeslund, T.B., Mortensen, B.K., Hansen, D.M.: Detecting head orientation in low resolution surveillance video. Technical report, CVMT-06-02 ISSN 1601-3646 (2006)
3. 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 22, 322–336 (2000)
4. Leonardo da Vinci (1452–1519). Uomo Vitruviano, Gallerie dell'Accademia, Venice, ITALY (ca. 1490), `http://brunelleschi.imss.fi.it/stampa_leonardo/images/uomo_vitruviano_accademia_v.jpg`
5. Baker, B.W., Woods, M.G.: The role of the divine proportion in the esthetic improvement of patients undergoing combined orthodontic/orthognathic surgical treatment. International Journal of Adult Orthodon Orthognath Surgery 16(2), 108–120 (2001)
6. Xiao, J., Moriyama, T., Kanade, T., Cohn, J.: Robust full-motion recovery of head by dynamic templates and re-registration techniques. International Journal on Imaging Systems and Technology 13(1), 85–94 (2003)
7. Asteriadis, S., Karpouzis, K., Kollias, S.: Head Pose Estimation with One Camera, in Uncalibrated Environments. In: International Workshop on Eye Gaze in Intelligent Human Machine Interaction (2010)
8. Asteriadis, S., Soufleros, D., Karpouzis, K., Kollias, S.: A natural head pose and eye gaze dataset. In: International Conference on Multimodal Interfaces (ICMI 2009), Boston, MA (November 2-6, 2009)
9. Viola, P., Jones, M.: Robust real-time object detection. International Journal of Computer Vision 57(2), 137–154 (2004)
10. Stiefelhagen, R.: Estimating Head Pose with Neural Networks - Results on the Pointing04. In: ICPR Workshop on Visual Observation of Deictic Gestures, Cambridge, UK (2004)