sensors

MDPI

Article

A Novel Zernike Moment-Based Real-Time Head Pose and Gaze Estimation Framework for Accuracy-Sensitive Applications

Hima Deepthi Vankayalapati [1,*], Swarna Kuchibhotla [2], Mohan Sai Kumar Chadalavada [3], Shashi Kant Dargar [1], Koteswara Rao Anne [4] and Kyamakya Kyandoghere [5]

[1] Department of Electronics and Communication Engineering, Kalasalingam Academy of Research and Education, Krishnankovil 626126, India
[2] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vijayawada 522302, India
[3] Department of Electronics and Communication Engineering, VelTech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai 600062, India
[4] Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankovil 626126, India
[5] Institute for Smart Systems Technologies, University Klagenfurt, 9020 Klagenfurt am Wörthersee, Austria
* Correspondence: v.himadeepthi@klu.ac.in

Abstract: A real-time head pose and gaze estimation (HPGE) algorithm has excellent potential for technological advancements either in human–machine or human–robot interactions. For example, in high-accuracy advent applications such as Driver's Assistance System (DAS), HPGE plays a crucial role in omitting accidents and road hazards. In this paper, the authors propose a new hybrid framework for improved estimation by combining both the appearance and geometric-based conventional methods to extract local and global features. Therefore, the Zernike moments algorithm has been prominent in extracting rotation, scale, and illumination invariant features. Later, conventional discriminant algorithms were used to classify the head poses and gaze direction. Furthermore, the experiments were performed on standard datasets and real-time images to analyze the accuracy of the proposed algorithm. As a result, the proposed framework has immediately estimated the range of direction changes under different illumination conditions. We obtained an accuracy of ~85%; the average response time was 21.52 and 7.483 ms for estimating head poses and gaze, respectively, independent of illumination, background, and occlusion. The proposed method is promising for future developments of a robust system that is invariant even to blurring conditions and thus reaching much more significant performance enhancement.

Keywords: head pose and gaze estimation (HPGE); feature extraction; Zernike moments; principal component analysis; linear discriminant analysis

check for updates

Citation: Vankayalapati, H.D.; Kuchibhotla, S.; Chadalavada, M.S.K.; Dargar, S.K.; Anne, K.R.; Kyandoghere, K. A Novel Zernike Moment-Based Real-Time Head Pose and Gaze Estimation Framework for Accuracy-Sensitive Applications. *Sensors* **2022**, *22*, 8449. https://doi.org/10.3390/s22218449

Academic Editor: Maysam Abbod

Received: 29 August 2022
Accepted: 28 October 2022
Published: 3 November 2022

1. Introduction

The head pose and gaze estimation (HPGE) algorithm plays a vital role in identifying a person's direction of observation and attention while driving. Safe driving is a complex task that needs a wide range of abilities and skills in cognitive, physics, and sensory data [1]. Driving does not always occur in ideal conditions, such as being well rested, well trained, and not engaged with non-driving-related activities. According to road accident statistics, the estimation of driver distraction can help improve the country's economic status [2], and HPGE plays a very crucial step in developing Advanced Driver Assistance System (ADAS).

Distraction is considered as the diversion state of the person concentrating on a situation and object unrelated to the primary task, which leads to risky decisions in the real-time environment. The head pose estimation calculates the spatial location of the human head along with the position or rotation angle (front or side poses) in an image.

Gaze estimation measures eyeball direction to predict the human attention level and better understand human activities. Head pose always coincides with gaze directions to know the focus of attention. It also successfully provides robust and accurate perception to interact with humans [1]. It is mainly used for analyzing complex meaningful gestures such as head nodding or pointing gestures. Gaze direction is achieved by capturing and analyzing the appearance of human eye images. The eye is one of the major sensory organs that conveys reliable information to the brain to perform tasks.

The main objective of this work is to develop a framework for monitoring the attentiveness and behavior of the person by estimating the head pose and gaze directions. This information is used to analyze and understand the attitude (behavior) of the person [3]. Therefore, we need to adopt a technique that is not intrusive and should not cause inconvenience to the person for developing an inexpensive automatic Head Pose and Gaze Estimation system, which is essential to understand the attention and behavior of the person. The automatic behavior identification system is crucial in real-time applications such as psychology, biometrics, surveillance, and safety surveillance systems. The HPGE system's performance is analyzed using the elapsed time and recognition accuracy.

This paper involves and analyzes the various head pose and gaze estimation techniques that exist in the literature. We have analyzed the parameters and constraints in the existing HPGE systems and propose a hybrid Zernike moment-based framework to overcome the drawbacks. It also includes a detailed methodology with the experimental results on standard databases and images acquired in real time (which consist of facial images with different sizes and illumination conditions).

2. Related Work

This section analyzes the state of the art in estimating attentiveness by focusing on conventional appearance-based methods to recent advanced deep learning algorithms for: (1) head pose estimation (2) gaze estimation, and (3) an integrated HPGE system on images in constrained and unconstrained environments.

2.1. Appearance and Template-Based Methods

In appearance-based methods, each train and test image is projected into the high dimensional feature space, and the distance between the features is calculated to find the best match. These methods extract the local features from the input image, invariant to varying illuminations [4]. An extension to these methods is template matching, which can be adapted and trained based on our application at varying environmental conditions. The performance of these methods depends on the size of the training data, occlusions, and the redundancy of input data [5,6]. Most of these methods calculate the best match using a minimum distance metric to find head pose and gaze orientations or movements. The minimum distance metric classifier is based on either Euclidean distance or Mahalanobis distances. Euclidean distance is the distance between two points x and y, represented as Equation (1)

$$\mathrm{dist}(\mathrm{x},\mathrm{y}) = \mathrm{sqrt}\left((\mathrm{x}-\mathrm{y})(\mathrm{x}-\mathrm{y})^{\mathrm{T}}\right)$$

(or)

$$\mathrm{dist}(\mathrm{x},\mathrm{y}) = \mathrm{sqrt}(\sum_{\mathrm{i}=1}^{2}(|\mathrm{x}_{\mathrm{i}}|-|\mathrm{y}_{\mathrm{i}}|)^{2}) \tag{1}$$

Mahalanobis distance is calculated as Equation (2)

$$||\mathrm{x}-\mathrm{y}||^{2}_{(\sum \mathrm{i})} = (\mathrm{x}-\mathrm{y})^{\mathrm{T}}\left(\sum \mathrm{i}\right)^{-1}(\mathrm{x}-\mathrm{y}) \tag{2}$$

To overcome the training data size and redundancy drawback, dimensionality reduction techniques are performed on the extracted feature vector. Dimensionality reduction techniques are mainly classified into (i) principal component analysis (PCA), (ii) Kernel

PCA, (iii) local discriminant analysis (LDA), (iv) canonical correlation Analysis (CCA), (iv) independent component analysis (ICA), (v) Isomaps, (vi) diffusion maps, etc.

According to the literature, appearance-based methods use an entire eye image as a high dimensional input feature vector and map them to a low dimensional gaze directional space vector. Therefore, these methods do not require extracting the eye features and giving accurate results, and the performance depends on the number of training samples [4,7].

2.2. Geometric and Model-Based Methods

These methods extract facial key points as the global features such as eyes, nose, mouth, etc. The intra-ocular distance is used to find the head pose and gaze of the person. It is desirable to consider the trade-off between the number of key points and accuracy [8]. Active Shape model and Active Appearance model are used to find the landmarks and create a deformable face shape model. These techniques are used to generate 3D face models [9]. High-accuracy applications require the extraction of a more significant number of landmarks, which increases the algorithm's complexity at changing illumination conditions, and occlusions degrade the system's performance further [3]. These methods determine the gaze direction by obtaining mapping functions from the features of 2D eye images [10]. These methods are quite simple but are unsuitable for many real-world applications because they have difficulty finding head movements [11]. Model-based approaches use three-dimensional geometrical relationships among the eyes, infrared light sources, and camera positions. The optical axis is estimated by connecting the line between the cornea and pupil center from which the visual axis is obtained from user-dependent offset angles [12,13]. The visual axis is more sensitive to focal length, relative positions among the camera, display screen, and offset angles, and it is still a challenging task to obtain gaze direction estimation accurately by using custom-made devices with one camera and a small display screen. Most existing gaze estimation techniques include model-based or feature-based techniques [14–16]. The geometric models of the eyeball and its environment extract eye features (corneal infrared reflections, pupil center and iris contours) to fit the model. However, these methods are challenging to build and calibrate. Researchers classify the existing geometric characteristics-based gaze estimation techniques into shape-based (centroid points and primitive geometric shapes such as ellipse, rectangle, circle) and motion-based (tracked features of objects in videos) techniques [3].

2.3. Deep Learning-Based Methods

Researchers used deep learning techniques for both feature extraction and classification. One of the most popular deep learning techniques is Convolutional Neural Network (CNN). CNN is used for pose estimation, but the accuracy is not up to the mark in real time. The technique was improved by extracting the local features from the image, and then, CNN was applied for classification [9,17,18].

Table 1 shows the general description of the existing HPGE techniques. These non-intrusive head and eyelid movements observation systems fit very well to the requirements of the objectives of this research work. Novel extensions are related to a better analysis of the facial dynamics and to their correlation with different mood and emotion states.

This work focuses on the hybrid approach, which combines appearance-based and geometric-based pose and gaze estimation methods. First, the Viola–Jones algorithm extracts the global features such as eyes, and then Zernike moments are used to extract the local features from the face image and eyes for HPGE.

Table 1. Description of existing techniques.

Technique	Description
Head and facial movement analysis	Example 1: An infrared active sensor is used to detect both pupil and head motions in variable light conditions. By using the Kalman filter, facial features are tracked, and a smoothening of the motion of the features is ensured. The Gabor wavelets are used for fast feature detection [19]. Example 2: First, the face and eyes are detected using the Adaboost classifier and then passed through a Gabor filter. The output is normalized and passed to a data-driven classifier, a support vector machine (SVM) [20,21]. Example 3: Human face and facial features are detected using Haar Wavelets with Adaboost cascade algorithm. Then, eye closing is measured by analyzing the optical flows in the particular region [13,22].
Eye closure, blink rates	Example 1: The PERCLOS video-based system calculates the number of eyelid closures. It measures this within 1 to 3 min intervals. A special algorithm uses this number to estimate the person's drowsiness. A person under drowsiness generally has a longer eye closure than an alert person. This system has been validated in real on-road driving and with the Psychomotor Vigilance test (PVT) [23–25]. Example 2: Kinect cameras (passive stereo pair) are used to capture video of the human's head to generate the 3D pose in real time. The persons face (±1 mm, ±1 deg) as well as the eye gaze direction (±3 deg), blink rates and eye closure are measured [23,26].
Eye blink detection approach	S multi-sensor system was developed to integrate eye lid camera (measure eye blinks) and other monitoring parameters (a steering grip sensor, a lane tracker). All signals are integrated to perform the task [27–30]
Multiple measures approach	An example of alertness-monitoring technologies comprises the MINDS system and eye-gaze system to estimate the head position and gaze, two potential fitness-for-duty systems (Safety Scope and Mayo Pupillometry system). Using this system, one can measure various parameters such as physiological, behavioral and subjective sleepiness measures. Moreover, all these parameters were integrated using a neural-fuzzy hybrid scheme [31,32].

3. Methodology

The overall flowchart of the proposed automatic head poses and gaze direction estimation system is shown in Figure 1. The proposed approach consists of the training phase and the testing phase. The preprocessing, feature extraction using Zernike moments, and dimensionality reduction steps are typical for the train and test phases. The minimum distance metric classifier is used to find the head pose and gaze direction estimation.

3.1. Preprocessing and Eye Pair Detection

The input images considered for this work are color (RGB) images. RGB color-based image has high-intensity value pixels versus the gray color image. To improve the computation speed and reduce the processing time, the color images are converted to grayscale images. For head pose estimation, there is no need to perform the face detection algorithm and pass the preprocessed image directly as the input to the feature extraction step because the Viola–Jones algorithm cannot detect faces in extreme poses. The image is resized to 256 × 256. In gaze direction estimation, the Viola–Jones algorithm detects the eye pairs of the face image [33,34]. The Viola–Jones algorithm detects the eye pairs very fast and accurately. The detected eye pairs are resized to 100 × 100.

3.2. Feature Extraction

For any successful recognition system, the representation of facial features is essential. The facial features are extracted using Zernike Moments (ZM) computation, and implementation is straightforward. First, the detected face features are extracted using the ZM feature extraction technique. It is defined as a unit disk space, determining the disk center by calculating the centroid of an image. The main advantage of using these moments is simple translational and scale-invariant techniques, which give high accuracy for the detailed shapes. From this, the system obtains a high dimensional feature vector.

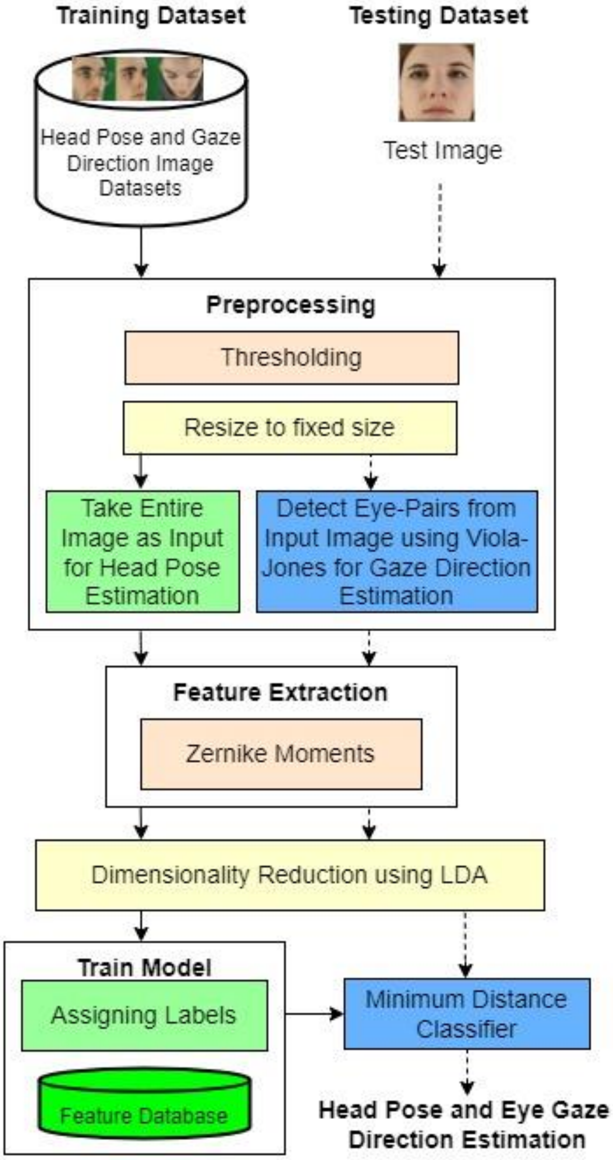

Training Dataset
Testing Dataset
Head Pose and Gaze Direction Image Datasets
Test Image
Preprocessing
Thresholding
Resize to fixed size
Take Entire Image as Input for Head Pose Estimation
Detect Eye-Pairs from Input Image using Viola-Jones for Gaze Direction Estimation
Feature Extraction
Zernike Moments
Dimensionality Reduction using LDA
Train Model
Assigning Labels
Feature Database
Minimum Distance Classifier
Head Pose and Eye Gaze Direction Estimation

Figure 1. Overall head pose and gaze direction estimation system.

3.2.1. Zernike Moments Feature Computation

Zernike moments was initially introduced in the 1930s by Fritz Zernike. Later, it was adapted to images in the 1980s by Teague. Previously, hue moments were used as shape descriptors. These Zernike moments are based on orthogonality functions. ZM gives the shape information along with the pixel intensities, which improves the performance.

The Zernike moments are obtained from the transformed unit disk space for the extraction of shape descriptors invariant to translation, rotation, and scale, along with skew and stretch used for the feature extraction process to preserve more shape information [35,36]. These moments are the set of complex polynomials that form a complex orthogonal set over the interior of the unit circle, i.e., $p^2 + q^2 \leq 1$. ZM are the projection of the image function on some orthogonal basis function introduced [37].

Let *I*(*p*,*q*) be the input image of size P X Q. By decomposing *I*(*p*,*q*) into the ZM basis function, the set of polynomials is denoted by {*Vnm*(*p*,*q*)}. These polynomials can be represented in Equations (3) and (4) [24].

$$Vnm(p,q) = Vnm(rho, theta) = Rnm(rho) * e^{(j*m*theta)}, theta \leq 1 \tag{3}$$

$$rho(p,q) = \sqrt{(p^2+q^2)},\ \ theta(p,q) = tan^{-1}\left(\frac{q}{p}\right) \tag{4}$$

where *n* specifies the polynomial order used to control the number of coefficients. The value of *n* should be a positive integer or zero represents the number of iterations. The value of *m* should be taken as positive or negative integers. *n* and *m* values should satisfy the conditions in Equation (5).

$$\begin{aligned} &(a)\ n\ \epsilon\ Z^+ \\ &(b)\ n-|m|\ is\ even, and \\ &(c)|m| \leq n \end{aligned} \tag{5}$$

The *Rnm*(*rho*) is the Zernike/Radial basis polynomial, which is defined as Equation (6)

$$Rnm(rho) = \left(\sum_{s=0}^{\frac{(n-|m|)}{2}} \frac{(-1^s) * rho^{n-2s} * (n-s)!}{s! * \left(\frac{n+|m|}{2}-s\right)! * \left(\frac{n-|m|}{2}-s\right)!}\right) \tag{6}$$

where *rho* represents the length of the vector from the origin to the pixel (*p*,*q*) (image pixel radial vector). Theta represents the angle between the X-axis and the vector rho in the counter-clockwise direction.

3.2.2. Zernike Moments Calculation

1. Take an entire image *I*(*p*,*q*) of size 256 × 256 as input.
2. Determine the center of the image.
3. Create a square window of minimum size (100 × 100 for gaze estimation, 64 × 64 for pose estimation) to focus the head region without distortion.
4. Place a square window in the original input image so that the image center should be the center of the square window.
5. Determine the order (*n*, *m* values) by satisfying the condition in Equation (5).
6. Calculate ZM basis function to generate the polynomials by using Equations (4) and (6).
7. Calculate complex Zernike moments from the image as Equation (7) using the basis function.

$$Vnm(p,q) = Vnm(rho, theta) = Rnm(rho) * e^{(j*m*theta)}, theta \leq 1 \tag{7}$$

8. Reconstruct the shape using the basis function and complex Zernike moments.

Here, input image is resized to 256 × 256. For pose estimation, divide the input image into 16 sub-images, each of size 64 × 64. Create a disk with radius as 64 and center at (32,32). Features are extracted using the 8th order Zernike moments (orthogonal polynomial), and they produce 25 descriptors/features from each sub-image. Higher-order moments are numerically unstable, sensitive to noise and will obtain a ringing effect at the edges. Therefore, there are a total of 16 × 25 = 400 descriptors that are uncorrelated, scale, translation, and rotation-invariant features in the image. For gaze estimation, after the Viola–Jones algorithm, eye images are resized to 100 × 100. In this case, a disk of radius 100 center at (50,50) is created. Total features extracted from the 8th order ZM are 25 features from each eye image.

3.3. Dimensionality Reduction

After the feature extraction step, 400 features for pose estimation and 50 features for gaze estimation are extracted using 8th order Zernike moments. To improve the system performance, principal component analysis (PCA) and linear discriminant analysis (LDA) are used for dimensionality reduction. These two methods are appearance-based methods that use minimum distance classifier (Euclidean distance) for classification. In this work, Mahalanobis distance is used to improve the accuracy for multivariant data (different variations in the data) [25].

The main difference between PCA and LDA is that PCA dimensionality reduction will take place based on the selection of eigen values. We must choose the top eigen values/eigen vectors. The data compression and performance depend on the number of eigenvectors chosen. LDA orders the dimensions according to class separability. PCA takes more time to recognize the expressions, and the classification rate is also less than LDA.

3.4. Head Pose and Gaze Direction Estimation

After assigning labels in the training phase, the low dimensional feature vectors of the trained images are stored in the database. Then, the system estimates the head pose using the minimum distance metric (Euclidean distance) based on these features. Minimum distance class is considered as a result of reducing the false-positive rate.

4. Results and Discussion

The experiment was conducted on collected images with the proposed Zernike moment-based framework applied to different sizes, invariant illuminations, and occlusions [38]. The images were taken from Aberdeen, MIT-CBCL (Massachusetts Institute of Technology and to the Center for Biological and Computational Learning), Iranian, and IMM (Informatics and Mathematical Modelling) databases and real-time images (eight subject images are captured randomly with a low-cost and low-resolution camera with different environmental conditions in real time). First, the head pose is estimated by pan angle ranging from −90 to +90 degrees, and the values were discretized with 45 degrees step.

To find the efficiency of the proposed pose estimation technique, experiments were conducted on the listed database and real-time images. The details of the images and class are shown in Table 2. Real-time images are used to experiment on the gaze estimation technique.

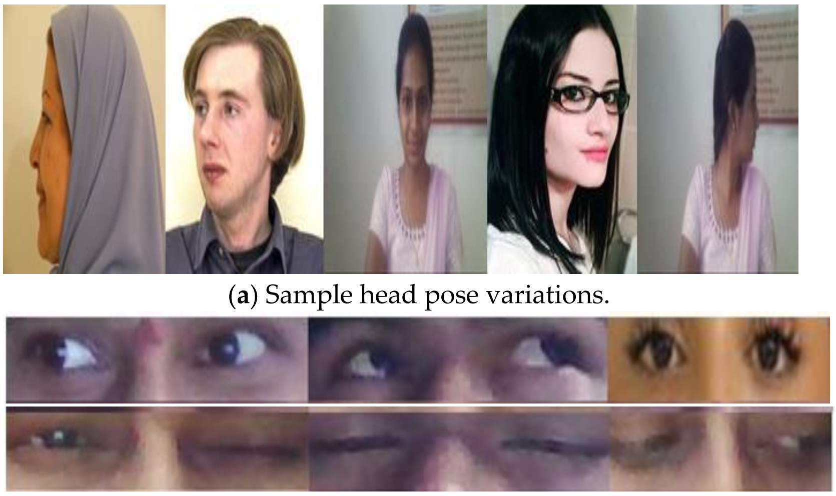
Table 2. Pose information on each dataset.

Database Name	Total No of Images	Left_90	Left_45	Front_0	Right_45	Right_90
Aberdeen	687	16	15	656	-	-
MIT-CBCL	59	10	12	16	10	11
Iranian	369	106	41	78	72	72
IMM-Data	240	-	47	151	42	-
Real-Time	120	12	40	20	35	13

A total of 1475 images were selected for pose estimation, out of which 740 images were considered for the training phase, and 735 images were considered for the testing phase randomly without any repetition of images. Experiments were conducted by considering all the database images together. Images were categorized into five classes (Left_90, Left _45, Front, Right_45, Right_90). The proposed system was evaluated based on performance accuracy and computation time.

4.1. Data Collection

In the Aberdeen database, each image varies in illumination, pose, facial expression and occlusion [39]. The MIT-CBCL database contains 10 subjects and 6 images per subject, one for each representing pose variation [38]. In contrast, the Iranian Face database contains 369 images of 34 human faces [40]. Each human face has 10 images representing different pose variations. Each image in the IMM face database represents poses or illumination [38,41]. The input images of different sizes and invariant illumination conditions were considered in real time, as shown in Figure 2. The details of the different pose images in the database and real time are mentioned in Table 2.


(**a**) Sample head pose variations.

(**b**) Sample gaze variations.

Figure 2. Sample head pose and gaze images.

For gaze estimation, images are either taken from the frontal face databases, which differ in gaze or real-time images taken from the real-time video (divide the images into frames) captured from camera. The details of the different eye gaze directions estimated in the proposed technique are shown in Table 3.

Table 3. Eye gaze directions estimated in proposed system.

	Different Gaze Directions	
Lower left	Lower middle	Lower right
Middle left	Middle (Front)	Middle right
Upper left	Upper middle	Upper right
Eyes Not Open	Half eyes open	One eye open

4.2. Performance Evaluation for Zernike Moments Feature Extraction

The performance evaluation was performed on the listed database and real-time images. Figure 3 shows the correctly estimated head poses for unknown real-time images.

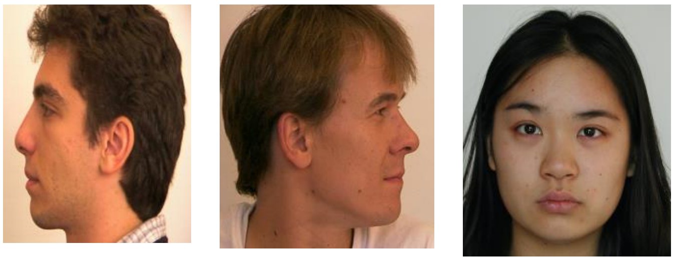


Figure 3. Estimation of the head pose for unknown images for different illumination conditions.

The proposed approach gives better results in terms of accuracy and response time even at different lighting conditions, as shown in Figure 3.

It also works well even in occlusion conditions. Figure 4 shows that the proposed system correctly estimates the head poses with occlusions.

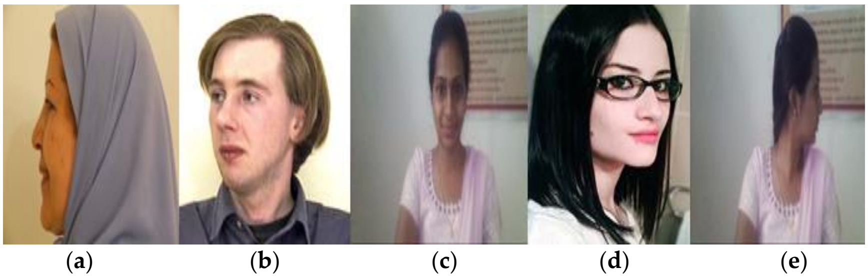


Figure 4. Estimation of head pose for occluded images. (**a**) Right pose; (**b**) Right pose with 45°; (**c**) Front pose; (**d**) Left pose with glasses; (**e**) Left pose.

The gaze direction estimation is also performed with good accuracy in different illumination and occlusion conditions as shown in Figure 5.

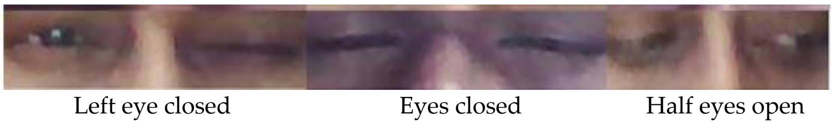


Figure 5. Estimation of gaze direction for varying illumination and occlusion conditions.

The system's response time is calculated from the images of the database, as shown in Table 4.

Table 4. Elapsed time to estimate each head pose using PCA and LDA.

Elapsed Time (ms)	PCA	LDA
Left_90	22.5	21.6
Left_45	22	21.5
Front_0	21.9	21.5
Right_45	21.4	21.4
Right_90	21.8	21.6

The dimensionality reduction techniques (PCA and LDA) are performed on the feature vector to obtain accurate information and increase the operational speed. LDA takes less time to estimate when compared to PCA. The time required for estimating each gaze direction using PCA and LDA is shown in Table 5.

Table 5. Elapsed time to estimate the gaze direction using PCA and LDA.

Elapsed Time (ms)	PCA	LDA
Eyes Open	7.7	7.5
Eyes Closed	7.61	7.45
Half Eyes Open	7.43	7.37
One Eye Open	7.66	7.48
Right	7.7	7.7
Left	7.8	7.4

Tables 4 and 5 show the performance of both PCA and LDA techniques. LDA gives better classification results compared to PCA. The accuracy is performed on images of the database or unknown images using PCA and LDA, as shown in Table 6.

Table 6. Performance analysis of proposed system using PCA and LDA.

Accuracy	PCA	LDA
Proposed Head Pose Estimation	79%	87%
Proposed Gaze Direction Estimation	77%	85%

Real-time head poses and gaze direction estimation plays a key role in many real-time applications. In this approach, the eye pair is detected using the Viola–Jones algorithms. Here, the Viola–Jones technique is not used for face detection because it cannot detect extreme pose conditions. Thus, the entire face image is passed as an input to the feature extraction technique. The Zernike moment technique is used for feature extraction, scale, translation-invariant and illumination-invariant features. For dimensionality reduction, both LDA and PCA are used. Compared to both, LDA gives better results in less response time. The minimum distance metric method estimates the pose and gaze direction. However, it will not estimate the correct results when blurring the images. The performance is evaluated by taking the images from the public databases or real-time images.

The proposed approach attains an accuracy of 87% using LDA and 79% using PCA for estimating the head pose and 85% using LDA and 77% using PCA for gaze direction estimation, as shown in Table 6. These experimental results show that pose and gaze direction is estimated with less response time and more accuracy on real-time images even if there is a change in illumination conditions, background, and occlusion.

5. Conclusions

We successfully estimated the head pose and gaze direction from the proposed Zernike moment-based approach with an accuracy of 85% after using LDA. While the existing system cannot estimate the results correctly in low-resolution blurred camera conditions, our proposed method still seems promising. Furthermore, even the most accuracy-significant real-time applications, such as driver assistance or interactive system, are operable at an accuracy of 90%. In this situation, our findings closely approach those benchmarks. Furthermore, it reveals that the proposed method can be suitable for implementing the system invariant to scale, translation, illumination and blurring conditions for performance enhancement.

Author Contributions: Conceptualization, H.D.V. and S.K.; Data curation, H.D.V.; Formal analysis, H.D.V. and S.K.D.; Investigation, H.D.V.; Methodology, H.D.V. and S.K.; Resources, H.D.V. and S.K.D.; Software, H.D.V. and M.S.K.C.; Supervision, K.R.A. and K.K.; Validation, H.D.V., S.K., M.S.K.C. and K.R.A.; Visualization, H.D.V.; Writing—original draft, H.D.V., S.K., M.S.K.C. and S.K.D.; Writing—review and editing, K.R.A. and K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data relevant to this work are available in the manuscript. Raw data and material are available on request.

Acknowledgments: The authors would like to thank all the authors of the research articles used in the literature study of this work and for their reports which laid the foundation of this work. Thanks to the Pattern Research Center, Iran, for providing the database of facial images.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Krinidis, M.; Nikolaidis, N.; Pitas, I. 3-D Head Pose Estimation in Monocular Video Sequences Using Deformable Surfaces And Radial Basis Functions. *IEEE Trans. Circuits Syst. Video Technol.* **2009**, *19*, 261–272. [CrossRef]
2. National Center for Statistics and Analysis. (2021, April). Distracted driving 2019 (Research Note. Report No. DOT HS 813 111). National Highway Traffic Safety Administration. Available online: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813111 (accessed on 12 August 2022).
3. University of North Carolina Highway Safety Research Center. Available online: https://www.hsrc.unc.edu/news/announcements/hsrc-to-lead-ncdot-center-of-excellence/ (accessed on 12 August 2022).
4. Lu, F.; Sugano, Y.; Okabe, T.; Sato, Y. Adaptive Linear Regression for Appearance Based Gaze Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 2033–2046. [CrossRef] [PubMed]
5. Zavan, F.H.; Nascimento, A.C.; Bellon, O.R.; Silva, L. Nose pose: A competitive, landmark-free methodology for head pose estimation in the wild. In Proceedings of the Conference on Graphics, Patterns and Images-W. Face Processing, Sao Paulo, Brazil, 4–7 October 2016.
6. Hossain, M.A.; Assiri, B. An Enhanced Eye-Tracking Approach Using Pipeline Computation. *Arab. J. Sci. Eng.* **2020**, *45*, 3191–3204. [CrossRef]
7. Svanera, M.; Muhammad, U.; Leonardi, R.; Benini, S. Figaro, hair detection and segmentation in the wild. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 933–937.
8. Wang, J.G.; Sung, E. Study on Eye Gaze Estimation. *IEEE Trans. Syst. Man Cybern.* **2002**, *32*, 332–350. [CrossRef] [PubMed]
9. Ranjan, R.; Patel, V.M.; Chellappa, R. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *41*, 121–135. [CrossRef]
10. Lee, H.; Iqbal, N.; Chang, W.; Lee, S.-Y. A Calibration Method for Eye Gaze Estimation System Based on 3D Geometrical Optics. *IEEE Sens. J.* **2013**, *13*, 3219–3225. [CrossRef]
11. Zhu, Y.; Xue, Z.; Li, C. Automatic Head Pose Estimation with Synchronized Sub Manifold Embedding and Random Regression Forests. *Int. J. Signal Process. Image Process. Pattern Recognit.* **2014**, *7*, 123–134. [CrossRef]
12. Erik Murphy-Chutorian, I.; Trivedi, M.M. Head Pose Estimation in Computer Vision: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 607–626. [CrossRef] [PubMed]
13. Fu, Y.; Huang, T.S. Graph Embedded Analysis for Head Pose Estimation. In Proceedings of the IEEE 7th International Conference on Automatic Face and Gesture Recognition (FGR'06), Southampton, UK, 10–12 April 2006; pp. 3–8.
14. Khana, K.; Khan, R.U.; Leonardi, R.; Migliorati, P.; Beninic, S. Head pose estimation: A survey of the last ten years. *Signal Process. Image Commun.* **2021**, *99*, 116479. [CrossRef]
15. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [CrossRef]
16. Vankayalapati, H.D.; Anne, K.R.; Sri Harsha, S. Estimating Driver Attentiveness Through Head Pose Using Hybrid Geometric-Based Method. *Smart Intell. Comput. Appl.* **2022**, *1*, 197–204.
17. Lal, S.K.; Craig, A.; Boord, P.; Kirkup, L.; Nguyen, H. Development of an algorithm for an EEG-based driver fatigue countermeasure. *J. Saf. Res.* **2003**, *34*, 321–328. [CrossRef]
18. Li, C.; Tong, R.; Tang, M. Modelling Human Body Pose for Action Recognition Using Deep Neural Networks. *Arab. J. Sci. Eng.* **2018**, *43*, 7777–7788. [CrossRef]
19. Kerr-Gaffney, J.; Harrison, A.; Tchanturia, K. Eye-tracking research in eating disorders: A systematic review. *Int. J. Eat. Disord.* **2019**, *52*, 3–27. [CrossRef] [PubMed]
20. Wijesoma, W.S.; Kodagoda, K.R.S.; Balasuriya, A.P. Road boundary detection and tracking using ladar sensing. *IEEE Trans. Robot. Autom.* **2004**, *20*, 456–464. [CrossRef]
21. Zhang, X.; Sugano, Y.; Bulling, A. Evaluation of appearance-based methods and implications for gaze-based applications. In *CHI Conference on Human Factors in Computing Systems, ser. CHI '19*; Association for Computing Machinery: New York, NY, USA, 2019.
22. Zhu, X.; Lei, Z.; Liu, X.; Shi, H.; Li, S.Z. Face alignment across large poses: A 3D solution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 146–155.
23. Dinges, D.F.; Grace, R. *PERCLOS: A Valid Psychophysiological Measure of Alertness as Assessed by Psychomotor Vigilance*; Highways; Operations and Traffic Management; Safety and Human Factors; I83: Accidents and the Human Factor; Federal Motor Carrier Safety Administration: Washington, DC, USA, 1999; FHWA-MCRT-98-006.
24. Hiraiwa, J.; Vargas, E.; Toral, S. An FPGA based Embedded Vision System for Real-Time Motion Segmentation. In Proceedings of the 17th International Conference on Systems, Signals and Image Processing, Rio de Janeiro, Brazil, 17–19 June 2010; pp. 360–363.
25. Videla, L.S.; Rao, M.R.; Anand, D.; Vankayalapati, H.D.; Razia, S. Deformable facial fitting using active appearance model for emotion recognition. In *Smart Intelligent Computing and Applications; Smart Innovation, Systems and Technologies*; Springer: Singapore, 2019; Volume 104, pp. 135–144.
26. Xu, Y.; Jung, C.; Chang, Y. Head pose estimation using deep neural networks and 3D point clouds. *Pattern Recognit.* **2022**, *121*, 108210. [CrossRef]
27. Yuan, H.; Li, M.; Hou, J.; Xiao, J. Single image-based head pose estimation with spherical parametrization and 3D morphing. *Pattern Recognit.* **2020**, *103*, 107316. [CrossRef]

28. Jiang, N.; Yu, W.; Tang, S.; Goto, S. Cascade Detector for Rapid Face Detection. In Proceedings of the IEEE 7th International Colloquium on Signal Processing and its Applications, Penang, Malaysia, 4–6 March 2011; pp. 155–158.
29. Khotanzad, A.; Hong, Y.H. Invariant Image Recognition by Zernike Moments. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 489–498. [CrossRef]
30. Aunsri, N.; Rattarom, S. Novel eye-based features for head pose-free gaze estimation with web camera: New model and low-cost device. *Ain Shams Eng. J.* **2022**, *13*, 101731. [CrossRef]
31. Svensson, U. *Blink Behaviour-Based Drowsiness Detection, Linkoping University*; Swedish National Road and Transport Research Institute: Linkoping, Sweden, 2004.
32. Barra, P.; Barra, S.; Bisogni, C.; de Marsico, M.; Nappi, M. Web-shaped model for head pose estimation: An approach for best exemplar selection. *IEEE Trans. Image Process.* **2020**, *29*, 5457–5468. [CrossRef]
33. Ji, Q.; Zhu, Z.; Lan, P. Real Time Non-Intrusive Monitoring and Prediction of Driver Fatigue. *IEEE Trans. Veh. Technol.* **2004**, *53*, 1052–1068. [CrossRef]
34. Vijayalakshmi, G.V.M.; Raj, A.N.J. Zernike Moments and Machine Learning Based Gender Classification Using Facial Images. In Proceedings of the Eighth International Conference on Soft Computing and Pattern Recognition, Vellore, India, 19–21 December 2016; Springer: Singapore, 2018; pp. 398–408.
35. Paul, V.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Kauai, HI, USA, 8–14 December 2003; Volume 1.
36. Hasan, S.Y. Study of Zernike Moments Using Analytical Zernike Polynomials. *Adv. Appl. Sci. Res.* **2012**, *3*, 583–590.
37. Amayeh, G.; Kasaei, S.; Bebis, G.; Tavakkoli, A.; Veropoul, K. Improvement of Zernike Moment Descriptors on Affine Transformed Shapes. In Proceedings of the 9th International Symposium on Signal Processing and Its Applications ISSPA, Sharjah, United Arab Emirates, 12–15 February 2007.
38. Fagertun, J.; Stegmann, M.B. Free Datasets for Statistical Models of Shape; Information and Mathematical Modeling, Technical University of Denmark, DTU. 2005. Available online: http://www.imm.dtu.dk/~{}aam/datasets/datasets.html (accessed on 13 February 2020).
39. Weyrauch, B.; Huang, J.; Heisele, B.; Blanz, V. Component-based Face Recognition with 3D Morphable Models. In Proceedings of the IEEE Workshop on Face Processing in Video, Washington, DC, USA, 27 June–2 July 2004.
40. Gross, R.; Li, S.Z.; Jain, A.K. Face Databases. 2020. Available online: https://www.face-rec.org/databases/ (accessed on 13 February 2020).
41. Nordstrom, M.M.; Larsen, M.; Sierakowski, J.; Stegmann, M.B. *The IMM Face Database—An Annotated Dataset of 240 Face Images*; Informatics and Mathematical Modelling, Technical University of Denmark: Lyngby, Denmark, 2004.