OMEG: Oulu Multi-Pose Eye Gaze Dataset

Qiuhai He[1(✉)], Xiaopeng Hong[1], Xiujuan Chai[2], Jukka Holappa[1], Guoying Zhao[1], Xilin Chen[1,2], and Matti Pietikäinen[1]

[1] Department of Computer Science and Engineering, University of Oulu, Oulu, Finland
{qhe,xhong,jukkaho,gyzhao,mkp}@ee.oulu.fi
[2] Key Lab of Intelligent Information Processing of the Chinese Academy of Sciences, The Institute of Computing Technology of the Chinese Academy of Sciences, Beijing, People's Republic of China
{chaixiujuan,xlchen}@ict.ac.cn

Abstract. Data is in a very important position for pattern recognition tasks including eye gaze estimation. In the literature, most researchers used normal face datasets, which are not specifically designed for eye gaze estimation. As a result, it is difficult to obtain fine labeled eye gaze direction. Therefore large datasets with well-defined gaze directions are desired.

To facilitate related researches, we collect and establish the Oulu Multi-pose Eye Gaze Dataset. Inspired by the psychological observation that gaze direction is intrinsically linked with the head orientation, we are devoted to a new data set of eye gaze images captured under multiple head poses. It finally results in a dataset containing over 40K images from 50 subjects, who were asked to fixate on 10 special points on screen under different poses respectively. We investigate a new eye gaze estimation approach by using the IGO based description, and compare it with other popular eye gaze estimation approaches to provide the baseline results on our dataset.

Keywords: Eye gaze · Head pose · Dataset

1 Introduction

Data is in a significant position for computer vision tasks including eye gaze estimation [4, 5, 6, 7, 15, 17, 18, 27]. In the literature, most researchers used normal face datasets, which are not specifically designed for eye gaze estimation. Several datasets such as the Ulm [30] and the HPEG [26] datasets are presented more recently. However, in HPEG datasets [26], only three coarse gaze directions ('straight forward', 'extreme left' and 'extreme right') are available. Although the Ulm dataset [30] owns multiple gaze and head directions, it only contains 20 subjects. As a result, large datasets with well-defined gaze directions are desired.

Many psychological studies [18, 23, 28] indicate that gaze direction and head orientation are intrinsically linked with each other. Head pose is a coarse indication of gaze, especially when the eyes of human beings are not visible, such as under remote CCTV camera or with sunglasses. Links of head pose and eye gaze also indicate the social attention transfer. More importantly, as indicated by [19, 20], judging where

R.R. Paulsen and K.S. Pedersen (Eds.): SCIA 2015, LNCS 9127, pp. 418–427, 2015.
DOI: 10.1007/978-3-319-19665-7_35

other people are looking comes from a combination of both eye gaze directions and head orientations [9, 24]. A graphic example of the famous 'Wollaston illusion' [14] shows that though the eyes are identical between the two face images, because of the difference of head pose, the appearances of gaze direction are different. To our best knowledge, only the newly published Columbia Gaze Data Set [2] takes the intrinsic connection between eye gaze and head poses into account. The datasets for multiple-pose eye gaze estimation are still intensely expected.

To facilitate the researches in eye gaze estimation, we designed and collected the Oulu Multi-pose Eye Gaze (OMEG) dataset. It includes 200 image sequences from 50 subjects (For each subject it includes four image sequences). Each sequence consists of 225 frames captured when people are fixating on 10 targeting points on the screen. The first three sequences of each subject are captured under three fixed head poses, namely 0 (the frontal) and ±30 degree respectively. The last sequence is in a free pose style. Table 1 compares our gaze dataset with other existing ones [2].The image sequences as well as the calibrated eye gazes, 'ground truth', will be made publicly available.

Moreover, we provide baseline results on our dataset by evaluating the popular approaches on eye gaze estimation [2]. Especially, we investigate a new eye gaze estimation approach by using the image gradient orientation (IGO) based description [11], which shows good robustness to illumination variations and outliers in the tasks of image registration [1,12], face recognition [11, 29], and pose estimation [31].

The remaining parts of this paper are organized as follows. Section 2 describes the hardware system setup used during the collection and recording procedure. We provide the calculation of ground truth in Section 3 and the baseline results in Section 4. Section 5 discusses the influence of different head poses in gaze estimation.

Table 1. Gaze dataset comparison

	HPEG	Ulm	Columbia	Ours (OMEG)
#Subjects:	10	20	56	**50**
#Gaze:	3	2-9	21	**10**
#Fixed Head Poses:	2	19	5	**3**
#Free Head Poses:	w/	w/o	w/o	**w/**
#Total image:	~5,500	2,220	5,880	**44,827**

2 Data Collection

This section describes the design of our dataset (Section 2.1), the sensor used to provide accurate location data in real world (Section 2.2), the environment setup (Section 2.3), and the data collection procedure (Section 2.4).

2.1 Basic Design

The directions of eye gaze are defined as the angles between the visual axis and the optical axis of the camera, as shown in Figure 1(a). Moreover, we assume that the

human's face is a virtual plane and use the normal direction of the facial plane to represent the head pose. To fix the visual focus, a large screen is used to display 10 red crosses (Figure 1(c)) as the viewpoints. We employ an industry tracker with high accuracy to acquire the positions of eye's center, viewpoints and the direction of facial plane.

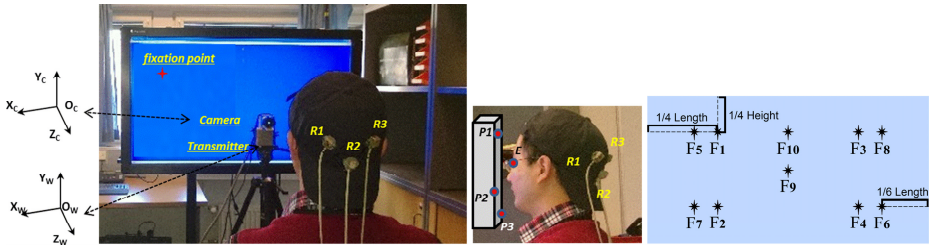


Fig. 1. Configuration for data collection. (a) Environment setting. The coordinates according to the transmitter and the camera are considered as the World Coordinate System and the Camera Coordinate System respectively; (b) $R1$ to $R3$ are three receivers fixed on a hat; $P1$ to $P3$ are three points sampled from the rectangle plane which is an approximation of the face direction; 'E' is the center of the human eyes; (c) The distribution of the fixation points.

2.2 High-accuracy Tracker System

We use the FASTRAK 3D tracker [8] to acquire the 3D position data. It consists of three units, including a system electronic unit, a transmitter, and several receivers. In the workflow, the transmitter is set as the origin of the World Coordinate System and the reference for the position and orientation measurements of the receivers. The receivers detect the magnetic fields emitted by the transmitter. The static accuracy of the receivers is claimed to be 0.08 cm root mean square (RMS) for positions, and 0.15° RMS for orientations.

2.3 Environment Setup

The environment for data collection is shown in Figure 1. A sequence of red crosses is displayed on large screen as the fixation point. A uEye 1540-c industry camera with a 1/2" CMOS image sensor and a resolution of 1280 (H) x 1024 (V) pixel is mounted in near front of the subject to capture the image sequences. The transmitter is mounted in a fixed position just under the camera. Both the camera and the screen with all the fixation points are calibrated before data collection as shown in Figure 1(a).

As can be seen from Figure 1, the subjects will wear a hat with three receivers during data collection procedure. For each subject, we assume the angle between the facial plane and the virtual plane constituted by these three receivers is constant. As a result, the virtual facial plane at any time point can be recovered by using the corresponding 3D positions of these three receivers.

2.4 Data Collection Procedure

Under our data collection design, the subjects need to fixate on the points displayed on screen. Ten fixation points are distributed as Figure 1(c) illustrates. These points are displayed continuously by every two seconds. The sequence of fixation points is displayed for five times. Therefore, the time to capture a whole loop in one pose is 100 seconds. For the fixed pose situation, the subject was asked to focus on the points only by eyes, but keep his/her head stable. During the data collection procedure, the three receivers on hat will record the transformation data of position all the time.

To represent the variations from different poses, a four-step collection is carried out. In the first three steps, the subject is asked to position in three fixed poses, namely the frontal pose, 30º to the left, and 30º to the right, respectively. The subject is asked to keep his/her head stable. In the final step, the subject is allowed to move his/her head freely. We collected samples from 50 subjects, in which 10 out of 50 are female. Some example images are illustrated in Figure 2.

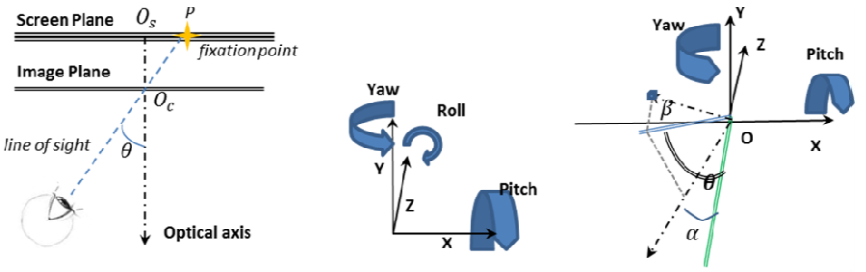
Fig. 2. Example image of the data collection: (a) Subjects in the frontal poses; (b) one subject focusing on the same fixation points in the three different poses.

3 Ground Truth Annotation

As introduced in Section 2.3, according to the spatial relationship between the facial plane and the hat plane, the direction of the facial plane can be recovered by using the 3D position of the three receivers on hat, as well as the position of the eyes' center. This section describes the details of acquiring the eye gaze direction.

3.1 Gaze Definition

A geometrical model for gaze estimation is illustrated by Figure 3. In HCI scenario, eye gaze refers to the pointer from the viewer's two eyes to an object. Note that the gaze orientation $\boldsymbol{\theta}$ in 3D space has three rotational degrees of freedom, namely the azimuth (or yaw), elevation (or pitch), and roll, as illustrated in Figure 3(b). In our research, we take the 'yaw' and 'pitch' into consideration. Thus we define the direction of eye gaze as $\boldsymbol{\theta} = \langle \alpha, \beta \rangle$, where α and β are rotations around the vertical and horizontal axis respectively as Figure 3(c).



(a) (b) (c)

Fig. 3. Geometrical model of eye gaze. (a) the 2D projection of geometrical model; (b) three rotation degrees of freedom (DOF) of a 3D angular; (c) an example to represent the gaze angle into two of the three DOFs, namely Yaw and Pitch.

3.2 Eye Gaze Direction

Before the data collection procedure, we record the initial positions of the three points ($P1^{(0)}$, $P2^{(0)}$, $P3^{(0)}$) on the facial plane and the three points ($R1^{(0)}$, $R2^{(0)}$, $R3^{(0)}$) on hat, through which we can compute the initial normal vector $V_P^{(0)}$ of facial plane by Eq. (1) , as well as the normal vector of hat plane $V_R^{(0)}$. More specifically,

$$V_P^{(0)} = \left(P2^{(0)} - P1^{(0)}\right) \times \left(P3^{(0)} - P1^{(0)}\right) . \tag{1}$$

Therefore, the angle θ between $V_P^{(0)}$ and $V_R^{(0)}$ can be computed by Eq. (2).

$$\theta = \cos^{-1}(V_P^{(0)} \cdot V_R^{(0)}) \tag{2}$$

During data collection, we record the position data of $R1^{(t)}$, $R2^{(t)}$ $R3^{(t)}$ at time *t*, *t* = 1,…,T. Then we compute the unit normal vector $\mathcal{U}_R^{(t)}(x_R^{(t)}, y_R^{(t)}, z_R^{(t)})$ of $V_R^{(0)}$ and $V_R^{(t)}$. As the angle θ is constant, according to [13], the rotation matrix $\mathcal{R}$ is

$$\mathcal{R} = \cos\theta \cdot \mathbf{I} + \sin\theta \cdot \left[\mathcal{U}_R^{(t)}\right]_{\mathrm{X}} + (1 - \cos\theta) \cdot \mathcal{U}_R^{(t)} \otimes \mathcal{U}_R^{(t)}, \tag{3}$$

where **I** is the identity matrix, $\left[\mathcal{U}_R^{(t)}\right]_{\mathrm{X}}$ is the cross product matrix of $\mathcal{U}_R^{(t)}$, and $\mathcal{U}_R^{(t)} \otimes \mathcal{U}_R^{(t)}$ is the tensor product matrix, which are presented by Eqs. (4) and (5), respectively.

$$\left[\mathcal{U}_R^{(t)}\right]_{\mathrm{X}} = \begin{bmatrix} 0 & -z_R^{(t)} & y_R^{(t)} \\ z_R^{(t)} & 0 & -x_R^{(t)} \\ -y_R^{(t)} & x_R^{(t)} & 0 \end{bmatrix} \tag{4}$$

$$\mathcal{U}_R^{(t)} \otimes \mathcal{U}_R^{(t)} = \begin{bmatrix} {x_R^{(t)}}^2 & x_R^{(t)} y_R^{(t)} & x_R^{(t)} z_R^{(t)} \\ z_R^{(t)} x_R^{(t)} & {y_R^{(t)}}^2 & y_R^{(t)} z_R^{(t)} \\ x_R^{(t)} z_R^{(t)} & y_R^{(t)} z_R^{(t)} & {z_R^{(t)}}^2 \end{bmatrix} \tag{5}$$

In light of the initial positions of eyes' center ($E^{(0)}$) and those three points ($R1^{(0)}$, $R2^{(0)}$, $R3^{(0)}$) on hat collected, their pairwise relation is determined. Given any of the three points on hat at time *t*, we are able to deduce the corresponding position of the eyes' center ($E^{(t)}$). To minimize the mean square error, we average the predicted positions deduced via the following equation:

$$E^{(t)} = \frac{1}{3}[\sum_{i=1}^{3}\left(R_i^{(t)} - \mathcal{R} \cdot (R_i^{(0)} - E^{(0)})\right)] \tag{6}$$

Therefore, according to the definition above, the eye gaze direction is the visual axis relative to the Camera Coordinate System (C_u, C_l), and the visual axis is the vector from eyes center to the fixation points (F_1, F_2, ⋯, F_j, ⋯, F_{10}). We then obtain the visual axis $V_E^{(j)}$ by Eq. (7).

$$V_E^{(j)} = P_j - E^{(t)} \tag{7}$$

As a consequence, we compute the pitch angle α and the yaw angle β of the eye gaze by Eq. (8).

$$\alpha = \cos^{-1}(V_E^{(j)} \cdot C_u),\ \beta = \cos^{-1}(V_E^{(j)} \cdot C_l) \tag{8}$$

4 Baseline Results

In this section, we show the landmark labelling and pre-processing procedure for eyes images (Section 4.1), and describe the methods of feature extraction (Section 4.2). Section 4.3 presents the experimental results of eye gaze estimation on OMEG.

4.1 Landmark Labelling and Pre-processing

Landmarks localization is a key step in nearly all face analysis tasks. We employed the state-of-the-art Bayesian based landmark detector [32] to obtain five landmarks including two eye centers, the nose tip and two mouth corners. It takes into account both the local and global constraints. Some example of landmark detection results on the images from the collected datasets are illustrated in Figure 4(a). It can be observed that the detector successfully detects the landmarks even with non-frontal views, non-neutral expressions or occlusions caused by glasses or beards. Despite this, to ensure the accuracy of the locations, we check the landmarks image by image and revise the ones with large errors manually.

(a) (b)

Fig. 4. Examples of normalization. (a) Landmarks detected by the approach [32]; (b) Normalized average face of all subjects in each pose.

We align and crop all the face images by three steps. The first step is to rescale each image such that the ratio of the upper facial height to the cropped image height is 96:256. The upper facial height is the distance between the center of two eyes and the mouth center. Then the nose tip is set to be the center of each image. Finally, we crop the face images by the resolution of 256 × 256 pixels. Figure 4(b) shows the average face of all subjects in each fixed pose respectively.

For all experiments, we crop the eyes image from the normalized face images. In specific, we align the eyes center as the eyes image center, then crop the eyes image in the resolution of 30 × 150 pixels. Figure 5 shows examples for 10 eye gaze images from the 3 head poses respectively.

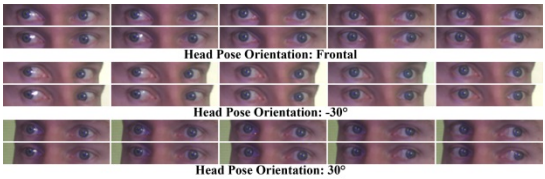


Fig. 5. Sample of eye images.

4.2 Feature Extraction

In [2], PCA [21] and multiple discriminant analysis (MDA) [22] based subspaces were used as the feature vectors and the support vector machine are used to learn the mapping from feature vector to the gaze locking labels. It is shown that this simple approach still achieves promising results. In analogue to [2], we report the results of baseline experiments with PCA and MDA subspaces.

Moreover, fueled by the success of IGO based facial descriptors [11, 31], we propose to use them for eye gaze estimation. More specifically, we characterize each image pixel by its gradient orientation. The IGO is expressed by a complex number, where both the real and imaginary parts are taken into account. Then we encode the IGO images via dimensionality reduction including 2D discrete cosine transform (DCT2), PCA and MDA. All feature vectors are of the same dimension for fair comparisons.

4.3 Recognition Procedure

Instead of treating the eye gaze estimation as a classification problem [2], we regard it as a regression problem. The reason is that the distribution of the yaw and pitch on the whole dataset are continuous in -38° to 36°and -10° to 29° respectively, as shown in Figure 6.We use off-the-shelf nonlinear regression models, support vector regression (SVR) [3], to learn the mapping from the feature space to the continuous gaze label. For those representations in angle, cosine kernel is used. For others, the RBF kernel is used unless otherwise noted. The optimal parameter settings are reached via grid search.

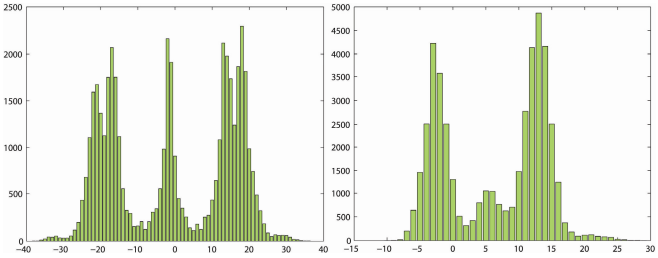


Fig. 6. The distribution of the yaw and pitch on the whole dataset.

We evaluate the eye gaze direction via leave-one-out cross validation. In specific, for each round, eyes' images in four sequences of 49 subjects are used for training while the images of the remaining subject are for testing. The average Mean Absolute Error (MAE) and average Pearson product-moment Correlation Coefficient (PCC) between the predicted gaze angle and the ground truth are used for performance measurement. Table 2 provides the performance of both the pixel intensity and IGO in DCT2, PCA and MDA subspaces, where d is the dimensionality of the extracted feature. As the experimental results suggest, the IGO eye gaze descriptor performs comparably or better than the pixel intensity. One exception is for yaw estimation. The intensity-PCA shows the best result. The reason is that the luminance distribution, which is lost in the IGO based representations, is crucial to the gaze perception for yaw changes [25]. Moreover, MDA, which is usually used for classification tasks, does not work as well as PCA and DCT2, as the eye gaze estimation is regarded as a regression problem in this paper.

Table 2. Baseline results ($d = 90$)

	IGO-DCT2	IGO-PCA	IGO-MDA	Intensity-DCT2	Intensity-PCA	Intensity-MDA
MAE(Yaw)	9.05	8.69	18.08	9.33	7.1	14.04
PCC(Yaw)	0.81	0.75	-0.02	0.75	0.86	0.16
MAE(Pitch)	5.96	5.77	7.02	6.08	7.32	6.91
PCC(Pitch)	0.65	0.56	0.01	0.56	0.68	-0.02

5 Discussion

To discuss the influence of different head poses in gaze estimation, we designed a set of experiments, which randomly pick fifty percent of subjects in one pose as the training set and the left subjects in other poses as the testing set. The data in three fixed discrete yaws, are chosen for these experiments. In terms of accuracy in Table 2 we employ the intensity-PCA as an example to extract features. As listed in Table 3, the performance under the same pose is much higher than the one under different poses. It means the eye gazes are distinctively different when head poses vary. Therefore, multi-pose gaze estimation is highly challenging. Fortunately, our dataset provides an opportunity of in-depth investigation in this issue.

Table 3. The pose influence in gaze estimation. Pose-(number) reflects the frontal, -30º and 30º head poses respectively.

(MAE/PCC)	Train Pose-1	Train Pose-2	Train Pose-3
Test Pose-1	**5.31/0.90**	11.54/0.57	11.66/0.64
Test Pose-2	13.85/0.25	**8.73/0.73**	15.44/0.18
Test Pose-3	13.70/0.37	18.64/0.22	**9.22/0.71**

6 Conclusions

In this paper we introduced the Oulu Multi-pose Eye Gaze Dataset. Unlike most existing gaze datasets, we collected the eye gaze data under multiple poses. We provided five landmark labels and the eye gaze angles as the ground truth. Lastly, to evaluate the effectiveness of our dataset, we reported baseline results of using IGO based description and the intensity via three dimensionality reduction techniques. The experimental results suggest that multi-pose gaze estimation is highly challenging. In spite of this, our dataset enables in-depth investigations in this topic. In future, we will evaluate other commonly used approaches such as [10, 16, 24, 25, 33], and publish the OMEG dataset with the ground truth.

Acknowledgements. This work was supported by the Academy of Finland and Infotech Oulu, and partially supported by the FiDiPro program of Tekes. Xilin Chen was partially supported by the Natural Science Foundation of China under contract No. 61390511.

Refferences

1. Fitch, A., Kadyrov, A., Christmas, W., Kittler, J.: Orientation correlation. In: Proc. BMVC. Citeseer, pp. 1–10 (2002)
2. Smith, B., Yin, Q., Nayar, S.: Gaze locking: passive eye contact detection for human-object interaction. In: Proc. UIST., pp. 271–280. ACM (2013)
3. Chang, C., Lin, C.: LIBSVM: a library for support vector machines. ACM Trans. Intelligent Systems and Technology, **2**(27) (2011)
4. Morimoto, C., Amir, A., Flickner, M.: Detecting eye position and gaze from a single camera and 2 light sources. In: Proc. ICPR., pp. 314–317. IEEE (2002)
5. Beymer, D., Flickner, M.: Eye gaze tracking using an active stereo head. In: Proc. CVPR. IEEE (2003)
6. Hansen, D., Pece, A.: Eye tracking in the wild. Comput. Vis. Image Und. **98**(1), 155–181 (2005)
7. Hansen, D., Ji, Q.: In the eye of the beholder: A survey of models for eyes and gaze. IEEE Trans. Pattern Anal. Mach. Intell. **32**(3), 478–500 (2010)
8. FASTRAK 3D tracker. http://www.cortechsolutions.com/Products/EL/EL-FP/EL-FP-FT
9. Lu, F., Okabe, T., Sugano, Y., Sato, Y.: Learning gaze biases with head motion for head pose-free gaze estimation. Image and Vision Computing **32**(3), 169–179 (2014)
10. Lu, F., Sugano, Y., Okabe, T., Sato, Y.: Adaptive Linear Regression for Appearance-Based Gaze Estimation. IEEE Trans. Pattern Anal. Mach. Intell., PrePrints (2014)

11. Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Subspace learning from image gradient orientations. IEEE Trans. Pattern Anal. Mach. Intell. **34**(12), 2454–2466 (2012)
12. Tzimiropoulos, G., Argyriou, V., Zafeiriou, S., Stathaki, T.: Robust fft-based scale-invariant image registration with image gradients. IEEE Trans Pattern Anal. Mach. Intell. **32**(10), 1899–1906 (2010)
13. Bar-Itzhack, I.: New method for extracting the quaternion from a rotation matrix. Journal of Guidance, Control, and Dynamics **23**(6), 1085–1087 (2000)
14. Illusion of face 9. http://www.psy.ritsumei.ac.jp/~akitaoka/kao9e.html
15. Wang, J., Sung, E.: Study on eye gaze estimation. IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics **32**(3), 332–350 (2002)
16. Kim, K., Ramakrishna, R.: Vision-based eye-gaze tracking for human computer interface. In: Proc. SMC, pp. 324–329. IEEE (1999)
17. Tan, K., Kriegman, D., Ahuja, N.: Appearance-based eye gaze estimation. In: Proc. WACV, pp. 191–195. IEEE (2002)
18. Symons, L., Lee, K., Cedrone, C., Nishimura, M.: What are you looking at? acuity for triadic eye gaze. J. Gen. Psychol. **131**(4), 451–469 (2004)
19. Cline, M.: The perception of where a person is looking. Am. J. Psychol. **80**(1), 41–50 (1967)
20. Gamer, M., Hecht, H.: Are you looking at me? Measuring the cone of gaze. J. Exp. Psychol. [Hum Percept.]. **33**(3), 705–715 (2007)
21. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3**(1), 71–86 (1991)
22. Duda, R., Hart, P., Stork, D.: Pattern classification. John Willey & Sons **2**, 114–124 (2001)
23. Jenkins, R.: The lighter side of gaze perception. Perception **36**, 1266–1268 (2007)
24. Valenti, R., Sebe, N., Gevers, T.: Combining Head Pose and Eye Location Information for Gaze Estimation. IEEE Trans. Image Process **21**(2), 802–815 (2012)
25. Ando, S.: Luminance-induced shift in the apparent direction of gaze. Perception **31**, 657–674 (2002)
26. Asteriadis, S., Soufleros, D., Karpouzis, K., Kollias, S.: A natural head pose and eye gaze dataset. In: Proc. AFFINE Workshop (2009)
27. Baluja, S., Pomerleau, D.: Non-intrusive gaze tracking using artificial neural networks. Tech. rep., Department of Computer Science, Carnegie Mellon University (1994)
28. Langton, S., Honeyman, H., Tessler, E.: The influence of head contour and nose angle on the perception of eye-gaze direction. Perception & Psychophysics **66**(5), 752–771 (2004)
29. Zhang, T., Tang, Y.Y., Fang, B., Shang, Z., Liu, X.: Face recognition under varying illumination using gradientfaces. IEEE Trans. Image Process. **18**(11), 2599–2606 (2009)
30. Weidenbacher, U., Layher, G., Strauss, P., Neumann, H.: A comprehensive head pose and gaze database. In: Proc. IET (2007)
31. Hong, X., Zhao, G., Pietikainen, M.: Pose Estimation via Complex-Frequency Domain Analysis of Image Gradient Orientations. In: Proc. ICPR. IEEE (2014)
32. Zhao, X., Shan, S., Chai, X., Chen, X.: Cascaded Shape Space Pruning for Robust Facial Landmark Detection. In: Proc. ICCV. IEEE (2013)
33. Sugano, Y., Matsushita, Y., Sato, Y.: Appearance-Based Gaze Estimation Using Visual Saliency. IEEE Trans. Pattern Anal. Mach. Intell. **35**(2), 329–341 (2013)