

Introducing I2Head Database

Ion Martinikorena

Electrical and Electronics Engineering Department, Public
University of Navarre
Pamplona, Navarre, SPAIN
ion.martinikorena@unavarra.es

Arantxa Villanueva

Electrical and Electronics Engineering Department, Public
University of Navarre
Pamplona, Navarre, SPAIN
avilla@unavarra.es

Rafael Cabeza

Electrical and Electronics Engineering Department, Public
University of Navarre
Pamplona, Navarre, SPAIN
rcabeza@unavarra.es

Sonia Porta

Electrical and Electronics Engineering Department, Public
University of Navarre
Pamplona, Navarre, SPAIN
sporta@unavarra.es

ABSTRACT

I2Head database has been created with the aim to become an optimal reference for low cost gaze estimation. It exhibits the following outstanding characteristics: it takes into account key aspects of low resolution eye tracking technology; it combines images of users gazing at different grids of points from alternative positions with registers of user's head position and it provides calibration information of the camera and a simple 3D head model for each user. Hardware used to build the database includes a 6D magnetic sensor and a webcam. A careful calibration method between the sensor and the camera has been developed to guarantee the accuracy of the data. Different sessions have been recorded for each user including not only static head scenarios but also controlled displacements and even free head movements. The database is an outstanding framework to test both gaze estimation algorithms and head pose estimation methods.

CCS CONCEPTS

• **Information systems** → **Multimedia databases**; • **Human-centered computing** → **Human computer interaction (HCI)**;

KEYWORDS

low cost gaze estimation, database, gaze estimation evaluation, head pose estimation

ACM Reference Format:

Ion Martinikorena, Rafael Cabeza, Arantxa Villanueva, and Sonia Porta. 2018. Introducing I2Head Database. In *PETMEI '18: 7th Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*, June 14–17, 2018, Warsaw, Poland. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3208031.3208033>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

PETMEI '18, June 14–17, 2018, Warsaw, Poland
© 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-5789-0/18/06.
<https://doi.org/10.1145/3208031.3208033>

1 INTRODUCTION

High resolution eye tracking and gaze estimation systems are a fact as the numerous companies offering nicely working systems demonstrate. However, as it is already known, the constraints required by the systems to perform in a reliable manner reduce considerably the number of applications of these systems to two main fields, i.e. the analysis of eye movements and their use as human computer interaction tool for severely disabled people [Majoranta et al. 2011]. The scientific community tries to identify new application areas for eye tracking systems. However, most of them present new challenges for the technology to become a more plug and play systems. Efforts done by researchers in the last few years are devoted to moving into low resolution hardware and unconstrained working scenarios. Trying to do eye tracking using a mobile gadget, in a car while driving or in wild lighting conditions requires to create a completely new paradigm for eye tracking and gaze estimation. The absence of infrared lighting and the possibility for the user to move freely entails to develop new eye tracking methods not only in terms of image processing but also in the gaze estimation part of the process.

Several works can be found in the literature focused on low cost eye tracking systems, where techniques proposed can be clearly classified into two groups. In the first group we find those works turning to methods based on high resolution systems, i.e. trying to adapt the existing knowledge to low cost scenarios [Valenti et al. 2012]. In the second group we find those proposals based on appearance in which methods based on machine and deep learning could be included [Krafka et al. 2016].

One of the key issues in the area of low resolution systems today is to find a consistent validation environment for the methods proposed. To this end, a bunch of databases have been created in the last few years. EYEDIAP [Funes Mora et al. 2014] database uses Kinect to obtain RGB and depth information from the subject. A 3D target is used and the user is able to move freely while gazing at a ball. In total 94 videos are contained in the database in which 16 individuals participated. The users were asked to remain static or to perform free head movement. Iterative Closest Points (ICP) method is employed to obtain 3D head pose using a personalized head model by means of a depth map. Thus, not only gaze data but also head pose information is provided in the database. MPII Gaze [Zhang et al. 2015] contains 213,659 images collected from

15 participants during natural everyday laptop use over more than three months. The authors created an application running in individuals' devices that asked them to gaze specific points on the screen during their everyday activities. Hence, this is one of the largest and more variable datasets in the field. TabletGaze [Huang et al. 2017] consists of 51 subjects, each with four different postures and 35 gaze locations. The images were recorded using the front camera of a tablet. Each user had to record four sessions holding the tablet differently while gazing at specific points on the screen.

The difficulty of obtaining large scale databases in the field of eye tracking is the fact that the labelling of the data is not straightforward. Eye images have to be linked with the gazed point and this information is not easily available. The most remarkable work in the field is the one developed by Krafska et al. [Krafska et al. 2016] containing 2.5 millions of images from 1450 participants. The method employed for obtaining labeled data is based on *crowdsourcing* by means of a designed application named as GazeCapture installed in subjects' tablets and phones. In this manner, subjects could activate the application any time and gaze specific points on the screen that could be registered together with the eye images captured by the camera of the gadget. An alternative solution for overcoming the problem of obtaining tagged data is to use "learning by synthesis" approaches. Employing simulation environments synthetic images are constructed in which the labels are already known as they have been used to build the image. Thus, enormous amount of tagged images can be obtained easily. Remarkable works in this area are the ones presenting Multi-view gaze dataset [Sugano et al. 2014] or the proposals made by Świrski and Dodgson [Świrski and Dodgson 2014] and Wood et al. [Wood et al. 2016]. Several of the aforementioned datasets have been constructed to be used as validation environments for deep learning techniques. Consequently, the only information required is images and ground truth gaze data. Deep learning, i.e. Convolutional Neural Networks (CNN), has demonstrated to work nicely in many computer vision problems such as detection or classification. Interesting efforts have been made in order to apply deep learning to gaze estimation [Krafska et al. 2016]. However, other works try to follow other knowledge-based approaches in which head pose is assumed to be a key factor to determine gaze direction [Zhang et al. 2015] [Deng and Zhu 2018] combined with deep learning. From our point of view, when moving to low cost environments it is essential to consider head pose data in the equation. Moreover, in order to face the objective of plug and play eye tracking systems, as many as possible labelled data are required.

We would like to contribute to the challenge of low resolution eye tracking by introducing I2Head dataset in which all the fundamental aspects of a database devoted to low cost eye tracking are taken into account in an outstanding accurate fashion. The database represents a significant contribution to the state of the art in eye tracking as it will be later described, i.e. apart from gaze data, ground truth values for head pose and camera calibration parameters are provided. In addition, a simplified user's head model is given, specific head movement scenarios are considered and different grids of points are introduced in the database in order to prevent overfitting problems. I2Head aims to be part of the state-of-the-art databases regarding low cost gaze estimation together with other examples such as EYEDIAP or MPII Gaze. The main contribution of I2Head is not

only the high accuracy of the data but also the head model provided for each user which facilitates the evaluation of model based head pose estimation methods. Head pose being key for accurate gaze estimation in low resolution this database provides the perfect framework to evaluate head and gaze estimation methods. I2Head is a publicly available database.

In the next section the hardware employed to construct the database is sketched. To follow, the details of the setup calibration procedure are provided. In section 4 a description of the users recording method is given. To finish, the final structure of the database is presented before the conclusions.

2 HARDWARE

The hardware employed to construct the database consists in a magnetic sensor (The Flock of Birds by Ascension Technologies) for 3D pose estimation and a camera. The magnetic sensor consists of a transmitter to which two probes can be attached, i.e. S_1 and S_2 sensors. In other words, Trackstar system can register the position of both sensors simultaneously with an accuracy value of 1.4 mm rms and 0.5° rms as provided by the manufacturer. The system is used to register the head pose with respect to the transmitter. To this end, one of the sensors is attached to the user's head while performing head movements. The sensor registers 240 samples per second. The output of the sensor is a 6D vector containing translation and rotation, i.e. roll, yaw and pitch angles, information. The employed camera is a standard Logitech webcam with a resolution of 1280x720 pixels working at 30 fps.

Both, the camera and the transmitter are placed together in a robust structure. The objective of this structure is to assure that any movement or impact is affecting equally the camera and the transmitter keeping constant the relative position of both elements.

The gazing surface is also attached to the same structure using a panel. The panel contains two different grids of points composed by 17 and 65 points. It is decided to construct the gazing area over a plastic material in order to facilitate the calibration of the system when the gazed points are to be referred to the camera hence the relative position of the gazing surface with respect to the camera is due to be calculated. In figure 1 a photo of a recording session is represented using a mannequin to substitute a real user.

Two plastic pieces have been designed and constructed as holders to host the sensors. One of the pieces behaves as a pointing device as it will be later explained (see figure 2).

3 SYSTEM CALIBRATION

One of the most critical parts of the database is the setup calibration between the transmitter named as T and the camera, i.e. C considered to be the origin of the World Coordinate System (WCS). The system calibration has the following objectives: to obtain the intrinsic parameters of the camera, to calibrate the position of the magnetic system with respect to the camera and to calculate the positions of gazed points with respect to the camera. In summary, as the camera projection center is considered to be the origin of WCS any element of the framework has to be referenced into camera coordinate system. Accuracy is an essential requirement for a database as the one presented in which an error of millimeters in the head position can affect significantly the gaze detection error.



Figure 1: In the photograph the mannequin represents the user with the sensor attached to the head. The camera, the transmitter and the gazing surface are placed in the same wood structure in order to fix their relative poses.

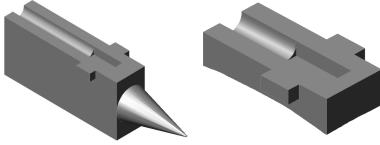


Figure 2: The designed plastic pieces are used as holders for the sensors. The one on the right is used to place the sensor when it is attached to the user's head. The one on the left is used as a pointing device

To this end, a careful calibration procedure has been designed. Calibration procedure is similar to that carried out in the work by Ariz et al. [Ariz et al. 2016], where the objective was to create a database of videos for HPE algorithms. It is decided to detail the procedure again due to two main factors: first, the objective now is to construct a database including gaze data requiring the knowledge of the position of gazed points; second, important changes have been introduced in the calibration to make it simpler and to improve its accuracy. Moreover, for the sake of clarity its worth to detail the calibration procedure of the I2Head database.

A calibration grid composed by a chess pattern is employed in which the first sensor, i.e. S_1 is attached as it is shown in figure 3.

During the camera calibration fifty images of the grid in alternative positions are recorded following an ordinary camera calibration procedure. At the same time the position of the grid with respect to the transmitter is recorded for each one of the calibration images using the matrix $M_{S_1 T}^i$ where $i = 1 \dots 50$. Regarding the notation M_{AB} represents the transformation matrix, i.e. rotation and translation, between A and B systems. Once the camera calibration procedure is finished both, intrinsic and extrinsic parameters of the camera are obtained. The extrinsic parameters contain the position of each one of the calibration grids with respect to the camera, i.e. M_{GC}^i , where G represents the grid pattern.

One of the improvements introduced in the construction of the I2Head database is the estimation of the position of the grid with respect to the sensor. In other words, S_1 is attached to the grid in a

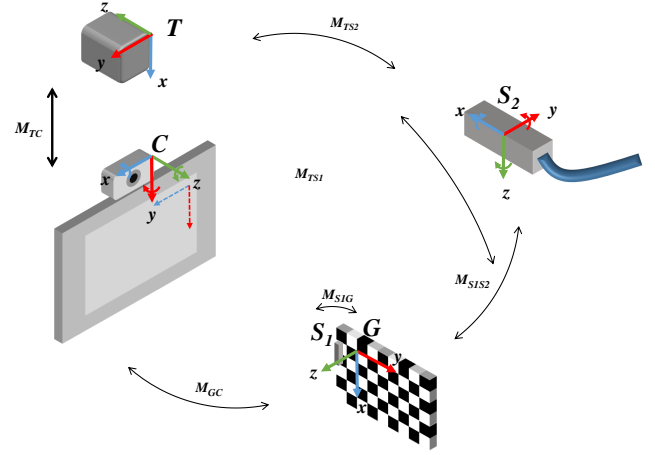


Figure 3: System elements involved in the setup calibration procedure. Sensor 1, S_1 is attached to the calibration grid while sensor 2, i.e. S_2 , will be introduced in the pointing device to estimate the transformation between the grid, G and S_1 . The final objective is to estimate M_{TC} as a function of intermediate system transformations.

specific position but the accurate transformation between the grid G and the the sensor S_1 is unknown.

To this end a second sensor is employed, i.e. S_2 . This sensor is introduced into the pointing device to be used as a marker. In a specific grid position the pointing device including S_2 is employed to mark forty points of the grid (see figure 4). Since S_1 is registered at the same time the position of the points with respect to S_1 can be calculated straightforwardly employing the transformation $M_{S_2 T}^j * M_{TS_1}$, where $M_{S_2 T}^j$ represents the position of S_2 with respect to the transmitter for the j^{th} point ($j = 1 \dots 40$) while M_{TS_1} represents the inverse transformation of S_1 with respect to the transmitter, which is now static as the grid does not move.

The grid is a regular pattern of black and white squares of size 30 x 30 mm. Thus, the coordinates of the forty points with respect to the grid origin G can be easily calculated. Procrustes analysis calculates a linear transformation (translation, reflection, orthogonal rotation, and scaling) between two set of points to minimize distances between the clouds of points. In this manner the transformation between the forty points according to S_1 and to G is calculated. Hence, as result of the Procrustes analysis the transformation between the two systems is calculated, i.e. M_{GS_1} . This transformation should be identical for any position of the grid.

Consequently, for each one of the grids employed during the camera calibration a transformation between the transmitter and the camera is calculated as:

$$M_{TC}^i = M_{TS_1}^i \cdot M_{S_1 G} \cdot M_{GC}^i \quad (1)$$

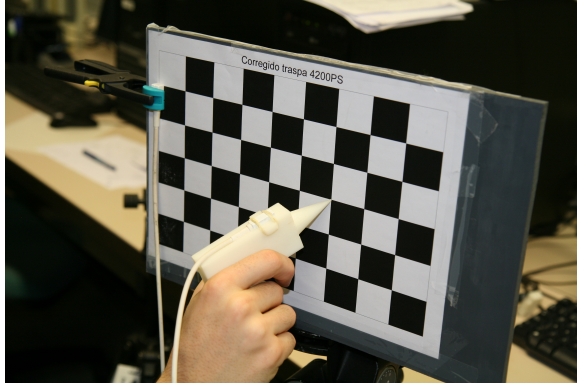


Figure 4: The pointing device carrying S_2 is used to mark forty points in the calibration grid while S_1 is registered. In this manner, the position of the forty points with respect to S_1 and to T can be easily calculated.

where $M_{TS_1}^i$ represents the position of S_1 with respect to the transmitter for the i^{th} position of the grid and M_{GC}^i is the position of the same grid with respect to the camera. Thus, we obtain a camera-transmitter transformation for each grid position which ideally should be equal to each other independently of the the grid position, i.e. the relative pose between the camera and transmitter does not vary. The distribution $\{M_{TC}^i\}$ with $i = 1...50$ is post-processed in order to detect outliers and to calculate a consensus value for M_{TC} . A summary of the elements involved in the setup calibration is sketched in figure 3.

Once the transformation between the transmitter and the camera has been calculated the positions registered by the transmitter can be easily transformed into camera coordinates. To follow, the positions of the points on the gazing surface with respect to the camera are calculated. To this end, the pointing device is again employed following the procedure shown in figure 5. The 3D points obtained after the marking procedure are further processed to be adjusted to the plane formed by the points on the gazing surface using also Procrustes analysis.

The quality of the calibration is measured by projecting the points of the calibration grid onto the captured image. The points of the calibration grid are projected based on the calibration information obtained according to:

$$G_C^i = M_{CT} \cdot M_{TS_1}^i \cdot G_{S_1} \quad (2)$$

where G_{S_1} (already known) and G_C^i represent the coordinates of the calibration grid points with respect to S_1 and the camera, i.e. C respectively. The grid of points G_C^i can be transformed into image (pixel) coordinates by using the intrinsic camera parameters and thus, be compared with the image captured by the camera. In an ideal calibration scenario the coincidence between the points calculated using the geometrical information and the ones in the image should be perfect. Figure 6 shows the result of a calibration. From the experiments carried out in order to measure the accuracy it can be deduced that our error remains below 1 mm for 3D data.

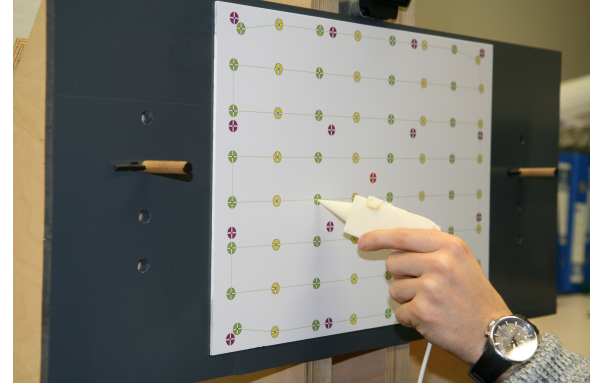


Figure 5: In order to calculate the position of the test points contained in the gazing surface the pointing device is used. After the marking procedure the position of the points is calculated with respect to the camera by using M_{TC} .

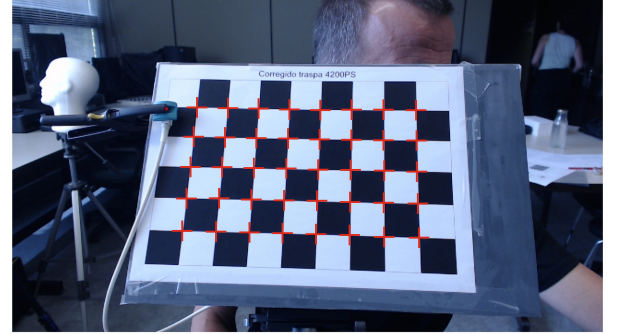


Figure 6: Red crosses represent the projection of the grid points onto the image by using calibration information. The distance in pixels between the red crosses and corner points can be used as measure of the goodness of the calibration procedure.

All the calibration information of the system is stored in the database.

It has to be clarified that in order to contribute to the precision of the data, specific calibrations have been performed for sensors S_1 and S_2 in order to locate accurately the origin of the sensors within their encapsulation which is not detailed here for simplicity.

4 USER RECORDING

Once the hardware of the system has been calibrated, user's sessions are to be recorded. Theoretically, a single hardware calibration should be enough as far as the relative position of system elements remains static. However, before any user is recorded, the calibration was tested and the system had to be recalibrated if necessary according to the procedure in the previous section.

The first sensor, i.e. S_1 , is attached to the top of user's head and remains fixed during the whole recording. The second sensor within the pointing device is used to mark 3D points of user's head.



Figure 7: The pointing device with a sensor is used to mark the four eye corners and the nose tip of the user. A mannequin is used in the figure to represent the individual.

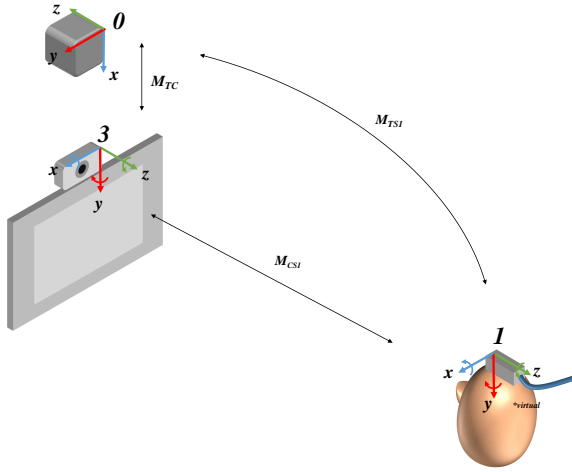


Figure 8: The user wearing S_1 sensor on the top of the head is recorded while gazing alternative points.

Four eye corners and nose tip are marked and saved in the database with the aim of having a simplified head model for each user in the dataset. This procedure is sketched in figure 7. The procedure allows to obtain a plain model for the head with respect to S_1 considered to be the origin of the head coordinate system. A drawing of a recording session is shown in figure 8.

Alternative sessions are recorded for each user. The requirements employed in order to design the sessions have been the following:

- Being tolerance to head position one of the main problems to overcome when working with eye tracking technology, selected user's displacements are chosen.
- When working with high resolution systems quasi-static head position is assumed due to the fact that the accuracy decreases if the head is moved on the one hand and because of the narrow Field of View (FoV) of this type of systems on



Figure 9: Sample images extracted from I2Head database.

the other hand. Sessions requiring static head are included in order to measure the effect of head movement with respect to the calibration position but, additionally, sessions in which the user is allowed to move the head freely are also introduced. This scenario is more feasible when working in low resolution scenarios.

- One of the main lacks regarding the evaluation of many gaze estimation algorithms is the fact that they employ the same grid for calibration and testing purposes. A gaze estimation method should demonstrate its generalization ability to be useful, i.e. should demonstrate that it can estimate gaze correctly in points not included in the calibration and to prevent overfitting situations. To this end different sessions including alternative grids of points, i.e. 17 or 65 points, are selected.

No chin rest is employed in any of the sessions. The sensor registers user position during all the sessions together with the time stamp. In the same manner, for any gazed point images are recorded for which the registration time is saved. Hence, employing a careful synchronization procedure user images and the HPE information from the sensor can be paired. Light conditions were not controlled, however, no complex variations of light or wild images have been considered in the database.

Several users participated in the construction of the database, the recording procedure was tiring and tedious in order to obtain reliable and valuable data for the eye tracking community. Finally, twelve valid users were recorded in alternative sessions for each of them. A sample of the images contained in the database can be observed in figure 9.

5 I2HEAD DATASET STRUCTURE

I2Head dataset is a public database containing data from twelve users in a desktop machine setting.

For each user, 8 videos are recorded under controlled movements. In the central position four sessions are recorded being this position located at 60 cm from the camera. The user is asked to keep the head static in the first two sessions during which the 17-points

Table 1: The following table summarizes the eight sessions recorded for each user. The characteristics for each session are provided in the columns. The first column shows the name of the session, the second one indicates the grid, the third one describes the free or static head condition while the last one shows the position of the user.

name	points	static/free	position
17pointsfree	17	free	centered
17pointsstatic	17	static	centered
65pointsfree	65	free	centered
65pointsstatic	65	static	centered
17pointsbwd	17	static	5 cm backwards
17pointsfwd	17	static	5 cm forwards
17pointsleft	17	static	5 cm to the left
17pointsright	17	static	5 cm to the right

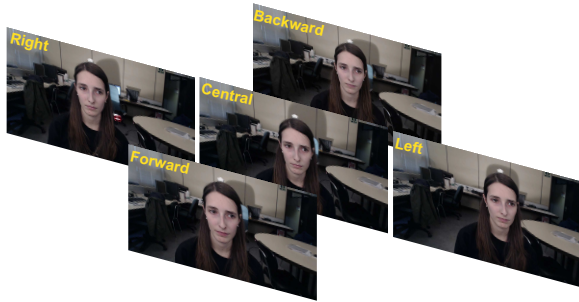


Figure 10: The image in the middle represents the centered position while left and right images represent lateral movements. The lower and upper images represent forward and backward movements respectively.

grid (static) and the 65-points grid (static) are recorded. To follow in the next two sessions the user is allowed to move the head in a free fashion while the 17 and 65 points grids are recorded. In the remaining four sessions the 17 points grid is exclusively employed changing the position of the user. The user is moved 5 cm approximately in forward, backward, leftward and rightward directions. During these sessions the user is asked to remain static. The range of gaze angles is approximately $\pm 20^\circ$. In table 1 the recorded sessions are summarized. Figure 10 shows the translation performed by the same user.

For each user setup calibration data are provided such as, camera calibration parameters, transmitter-camera information (M_{TC}), and the position of the points in the gazing surface with respect to the camera. This information can be different for each user if the setup calibration was to be repeated. Additionally, the simplified head model for the user is provided together with the data. Knowing the head data, i.e., head model and pose, permits to project 3D facial points of the head model in the image providing the 2D face landmark positions for each session if needed. Moreover, apart from the 2D gaze information included in the database, the user of the dataset can easily calculate the 3D gaze direction for each image according to head information. The eight sessions are given for each user classified according to the fixation points, i.e. 17 or 65. Each

Table 2: Images in I2Head database.

Users	Sessions	Images per point
12	6 (17 pts), 2 (65 pts)	10
Total	$12 * (6 * 17 + 2 * 65) * 10 = 27840$	

fixation point is divided into the number of images acquired and the HPE data recorded for the corresponding fixation while gazing the point during 1 sec. A filtering process has been carried out using the raw images in order to remove blinks and keep the best images for each fixation. Intraface [De la Torre et al. 2015] software is used to calculate the iris center corresponding to the alternative images captured. The iris center position distribution has been analyzed for each fixation point and using the mean value and standard deviation an outlier removal procedure has been carried out. An image is considered to be an outlier when the distance between iris center and the mean iris center positions is large in comparison with the standard deviation, e.g. an image containing a blink. For each fixation point the ten images in which the iris center was closer to the mean are automatically selected. In this manner, ten images per fixation are provided in the database free of blinks. In table 2 the total of images in the database is shown.

6 CONCLUSIONS

In this paper the I2Head database has been presented to the scientific community. The database contributes clearly to the state-of-the-art datasets devoted to low resolution gaze estimation together with other outstanding examples such as EYEDIAP and MPII Gaze. This database aims to be a solid framework in order to evaluate not only gaze but also head pose estimation algorithms. From our point of view, head data is key in low cost gaze estimation thus, a special emphasis has been paid in order to improve the head information included. The design of the database has been made, under the premise that head pose information is relevant, in order to estimate gaze and knowing the sensitivity of the accuracy with respect to user's movement. Hence, obtaining reliable and highly accurate head pose and calibration data has been one of the main objectives of our work. Moreover, the careful calibration and user recording procedure permits to assure a high accuracy of the data. The publicly available I2Head dataset contains images, gaze and HPE data of 12 users gazing different grids from alternative positions.

It has to be assumed that the solution for low resolution gaze estimation is still a challenge. Further work and research are required. The objective pursued with this database is to be a perfect framework to obtain solid conclusions based on real data about gaze estimation methods for low cost systems using controlled head movements. In other words, the database permits to evaluate HPE and gaze estimation methods in a reliable manner. Ground truth values for the head position and the gazed points are available together with the corresponding images, making it possible to evaluate the contribution of each source of error to the final gaze estimation.

ACKNOWLEDGMENTS

The authors would like to thank Iosu Inda, Rebeca Echeverria, Saray Agorreta, Jose Javier Bengoechea and Mikel Ariz for their collaboration in the construction of the I2Head database.

We would like to acknowledge the Spanish Ministry of Economy, Industry and Competitiveness for their support under Contracts TIN2014-52897-R and TIN2017-84388-R in the framework of the National Plan of I+D+i.

REFERENCES

- Mikel Ariz, José J. Bengoechea, Arantxa Villanueva, and Rafael Cabeza. 2016. A Novel 2D/3D Database with Automatic Face Annotation for Head Tracking and Pose Estimation. *Comput. Vis. Image Underst.* 148, C (July 2016), 201–210. <https://doi.org/10.1016/j.cviu.2015.04.009>
- Fernando De la Torre, Wen-Sheng Chu, Xuehan Xiong, Francisco Vicente, Xiaoyu Ding, and Jeffrey Cohn. 2015. Intraface. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, Vol. 1. IEEE, 1–8.
- H. Deng and W. Zhu. 2018. Monocular Free-Head 3D Gaze Tracking with Deep Learning and Geometry Constraints. In *2017 IEEE International Conference on Computer Vision (ICCV)*, Vol. 00. 3162–3171. <https://doi.org/10.1109/ICCV.2017.341>
- Kenneth Alberto Funes Mora, Florent Monay, and Jean Marc Odobez. 2014. EYEDIAP: A Database for the Development and Evaluation of Gaze Estimation Algorithms from RGB and RGB-D Cameras. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*. ACM, New York, NY, USA, 255–258. <https://doi.org/10.1145/2578153.2578190>
- Qiong Huang, Ashok Veeraraghavan, and Ashutosh Sabharwal. 2017. TabletGaze: Dataset and Analysis for Unconstrained Appearance-based Gaze Estimation in Mobile Tablets. *Mach. Vision Appl.* 28, 5-6 (Aug. 2017), 445–461. <https://doi.org/10.1007/s00138-017-0852-4>
- Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. 2016. Eye Tracking for Everyone. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Paivi Majaranta, Hirotaka Aoki, Mick Donegan, Dan Witzner Hansen, and John Paulin Hansen. 2011. *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies* (1st ed.). Information Science Reference - Imprint of: IGI Publishing, Hershey, PA.
- Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. 2014. Learning-by-Synthesis for Appearance-Based 3D Gaze Estimation. In *CVPR. IEEE Computer Society*, 1821–1828. <http://dblp.uni-trier.de/db/conf/cvpr/cvpr2014.html#SuganoMS14>
- Lech Świrski and Neil A. Dodgson. 2014. Rendering synthetic ground truth images for eye tracker evaluation. In *Proceedings of ETRA 2014*. 219–222. <http://www.cl.cam.ac.uk/research/rainbow/projects/eyerender/>
- R. Valenti, N. Sebe, and T. Gevers. 2012. Combining Head Pose and Eye Location Information for Gaze Estimation. *IEEE Transactions on Image Processing* 21, 2 (2012), 802–815. <https://ivi.fnwi.uva.nl/isis/publications/2012/ValentiTIP2012>
- Erroll Wood, Tadas Baltrušaitis, Louis-Philippe Morency, Peter Robinson, and Andreas Bulling. 2016. Learning an appearance-based gaze estimator from one million synthesised images.. In *ETRA*, Pernilla Qvarfordt and Dan Witzner Hansen (Eds.). ACM, 131–138. <http://dblp.uni-trier.de/db/conf/etra/etra2016.html#WoodBM0B16>
- Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2015. Appearance-Based Gaze Estimation in the Wild. *CoRR* abs/1504.02863 (2015).