International Journal of Electrical and Electronics Engineers
Volume 16, Issue No. 01, Jan-June 2024
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

A Review and Analysis of Eye-Gaze Estimation Methods

Ravi kumar[1],Dr. Silki Baghla[2],Er. Veena Rani

[1]Research Scholar, [2]Associate Professor, [3]HOD ECE

JCDM College of Engineering Sirsa, Haryana (India)

ABSTRACT

Human gaze estimation plays a major role in many applications in human-computer interaction and computer vision by identifying the users' point-of-interest. The revolutionary developments of deep learning have captured significant attention in the gaze estimation literature. Gaze estimation techniques have progressed from single-user constrained environments to multi- user unconstrained environments with the applicability of deep learning techniques in complex unconstrained environments with extensive variations. This paper presents a comprehensive survey of the single-user and multi-user gaze estimation approaches with deep learning. The state-of-the-art approaches are analyzed based on deep learning model architectures, coordinate systems, environnntal constraints, datasets and performance evaluation metrics. A key outcome from this survey realizes the limitations, challenges, and future directions of multi-user gaze estimation techniques. Furthermore, this paper serves as a reference point and a guideline for future multi-user gaze estimation research.

Keywords- Gaze Estimation, Deep Learning ,Machine Learning, Eye Gaze

I. INTRODUCTION

The path taken and destination taken, attention paid, feelings experienced, and encounters had. Estimating the direction in which a person's gaze is directed is a method that is regularly employed to gain a deeper comprehension of human behavior and cognition. Numerous studies have been conducted to investigate the methods that can be used to track the position and direction of eye gazing, which is necessary for a variety of fields, including cognitive and social behaviour .In addition, the situations in which gaze estimation takes place can be categorized as either restricted or controlled, or unconstrained or wild. Constrained environments are those that have a set of parameters that cannot be changed, such as the amount of illumination, the number of subjects, or the range of head angles. On the other hand, environments that have a significant amount of flexibility in their parameters are referred to be unrestrained environments. It is abundantly obvious that as a result of the broad use of gaze estimating technology across a wide variety of application areas, gaze estimation has evolved farther into unconstrained environments, hence surpassing limited environment settings. Even though there are a few different solutions for eye gaze estimation, some of them suffer from drawbacks such as high costs, the requirement of manual interventions, unreliability, and inaccurate performance in actual deployments. In addition, the performance of some older methods is hindered by elements such as poor image quality and lighting circumstances. In situations like these, eye gaze estimation methods that are based on Deep Learning (DL) come into play as a result of the inherited benefits they provide, which include learning from previously collected data, automation, a flexible procedure, high accuracies, and improved decision making. These prevalence DL-based techniques have proven to be effective in improving performance in eye gazing applications. Techniques that are model-based and techniques that are appearance-based are the two primary categories that can be used to classify approaches to human gaze estimation. Model-based methods, by their very nature, call for specialized equipment, such as near-infrared (NIR) cameras, in order to manually regress the eye's features and construct a geometric model [2].This strategy is only applicable to the individual using it and cannot be used in open contexts [3]. In contrast, approaches that are based on appearance do not require dedicated equipment and are not restricted to situations that have certain constraints placed on them. Conventional appearance-based approaches and appearance-based methods combined with deep learning are the two categories into which these methods can be categorized after being subdivided. Because of their application and reliability in unconstrained situations, approaches for estimating gaze that are based on deep learning techniques have experienced a boom in attention in the eye-tracking literature over the past decade. This interest has been driven by the fact that these methods are based on deep learning. Deep learning-based methods have numerous advantages over traditional appearance-based methods. These advantages include the capacity to extract high-level gaze properties from images and the capability to build a non-linear mapping function directly from an image to eye gaze [4].

Deep convolutional neural networks (DCNN) have been utilized in almost every deep learning-based gaze estimation approach due to their ability to map image features directly, handle large-scale datasets, and learn complex non-linear mappings when confronted with significant head-pose variations, eye occlusions, and illumination conditions. This is because of DCNN's ability to map image features directly, handle large-scale datasets, and learn complex non-linear mappings. The appearance-based approaches with deep learning that are the primary topic of this research can be further subdivided into two subcategories based on the number of subjects. These subcategories are referred to as single-user gaze estimation and multi-user gaze estimation, respectively. The demand for multi-user gaze estimating strategies is on the rise, notwithstanding the major shift in gaze estimation techniques towards applications in unconstrained environments. As of the year 2021, the time-shifting and space-shifting single-user gaze estimate study method has been used to the investigation of a select few of these methods.

This paper presents an overview of the most recent developments and advancements that have been made in eye gaze estimation research methods and procedures. We investigate the use of the most recent techniques for deep learning (DL), as well as helpful public datasets and other research methodologies that are connected. According to the findings of this survey and the lessons learnt from them, eye gaze applications are progressing with the use of DL methods due to the hereditary benefits. In addition, the findings of this study provide direction for following a DL-based procedure for eye gaze estimate, which can serve as a point of reference. In addition, we highlight the difficulties that have been encountered and the potential future paths of study for eye gaze estimation in a variety of applications. As a result, one of our goals is to provide researchers and developers with insights that can be helpful in the production of eye gaze estimation apps that are successful and efficient employing DL techniques [5].

- **Overview of Multi-User Gaze Estimation**

Eye contact between humans is a sort of active, natural connection that acquires information from a visual scene. Even while it is very unobtrusive and straightforward in compared to gesture and word, it is nonetheless able to convey a plethora of information regarding human actions. In eye gazing research, eye movements are analyzed in great detail in terms of the type of movement, how it functions, and the qualities it possesses. The analysis of a user's eye movements can be utilized to obtain information about the user's attention, purpose, and cognitive activities.Fixations, saccades, smooth pursuit, scanpath, gaze length, blinking, and changes in pupil size are the basic categories that are used to describe these kinds of eye movements [6].Fixations are the brief pauses that occur between movements in which the eyes remain still and study a situation. They have the slowest movement rate and are beneficial for reading, paying attention, and scanning large amounts of detailed information. Saccades, on the other hand, have the highest movement speeds and are useful for visual search. Saccades can be thought of as "jumping" from one location to another. These are motions that happen simultaneously with both eyes, and they take place between fixations. Eye-tracking movements known as smooth chases are employed in order to follow moving targets of interest. Before the eyes get at their destination, they travel along a scan path, which consists of a series of alternating eye fixations and saccades. 2D and 3D gaze are the two categories that can be used to describe the dimensionality of eye contact. When calculating eye stare in 2D, only the direction of gaze from one eye is required, but in order to calculate eye gaze in 3D, both the direction of gaze from both eyes and their depth of glance are required [7].

II. RELATED WORK

There have been many studies that have concentrated on eye gaze estimating research; however, there are only a few survey studies that examine expanding elements in the literature that is focused on DL approaches. These papers are available. Several of the research have examined several strategies for gaze estimation, such as methods based on models, methods based on appearance, methods based on deep learning, and methods based on convolutional neural network (CNN). For example, we have investigated these strategies with a particular emphasis on model-based approaches. They have presented their work within the framework of the following five categories: 1) 2D regression, 2) 3D model, 3) Appearance-based, 4) Cross Ratio-based, and 5) Shape based approaches. In a similar manner, researchers have analyzed recent developments in computer vision, such as deep learning, in order to assess gaze estimation techniques and divide them into two categories: 1) Geometric-based methods, and 2) Appearance-based methods. From a different vantage point, [24] has demonstrated a variety of various deep learning-based gaze estimate approaches that concentrate on CNNs. A significant number of these research have conducted additional analyses of the calibration methods, performance evaluation criteria, devices and platforms, and datasets found in the existing gaze estimate literature. However, the majority of the research do not explore these methods from the point of view of multi-user gaze estimation, taking into account issues such as unconstrained ambient conditions, gaze target fluctuations, and coordinate systems.

- **Gaze Estimation**

The measurement of human eye gazing is what's known as "gaze estimation," and it's used to determine things like a person's intent and interest [8].Researchers in the 18th century employed invasive eye-tracking techniques to examine people's eye movements [9]. This is where the history of human gaze estimation and eye-tracking began. However, thanks to advancements in digital signal processing and computer vision, non-invasive gaze estimate methods that make use of the specific, physical properties of the eye are becoming increasingly popular

International Journal of Electrical and Electronics Engineers
Volume 16, Issue No. 01, Jan-June 2024
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

[10].The key elements required for this endeavor have been provided by the photometric and kinematic characteristics of the human eye [11].Two metrics that are utilized in gaze estimation are the direction of the glance and the point of look. The gaze direction is determined by the visual axis, which is different from the optical axis, as shown in Figure 3. Properties of the eye, such as the pupil and corneal reflection, are extracted from eye regions and are utilized in the determination process at the application level [12].Following this, the definition of the gazing point is the intersection of the glance direction and the surface of the item being viewed.[13]

Before the development of methods based on computer vision, gaze estimating approaches relied on identifying patterns of movement such as fixations, saccades, and smooth pursuits (Young & Sheena,

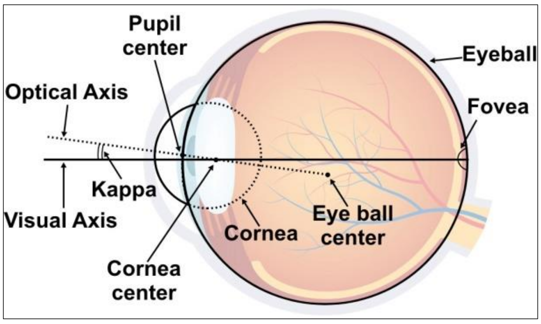

Figure1: Model of a human eye ball

1975). Today, these techniques rely on analyzing data generated by computers. There are three categories that can be used to organize computer vision-based approaches: (1) 2D eye feature regression methods, (2) 3D eye model recovery method, and (3) Appearance-based methods [14].These techniques estimate the gaze by making use of eye images and video data in conjunction with the geometric model properties of the eye. To be more specific, the first two methods detect geometric aspects of the eye like corneal reflection and the center of the pupil, and then they develop an eye model in order to estimate gaze [15].The literature refers to both of these strategies as model-based strategies, despite the fact that they are fundamentally different. In the third method, an attempt is made to estimate gaze by considering the photometric appearance of the eye [16].Methods that are based on appearance do not require the assistance of such specific instruments for gaze measurement, in contrast to model-based methods, which do require the support of dedicated devices such as infrared cameras. The remote eye tracker is normally kept at a distance of 60 centimetersfrom the user, whereas the cameras on the head-mounted eye tracker are typically put on a frame of glass. In general, there are two sorts of devices that are utilized in these methods: (1) the remote eye tracker and (2) the head-mounted eye tracker.There are four different types of user interfaces that are used for gaze estimation. These interfaces include active, passive, single, or multi-modal. Active interfaces make use of the user's gaze in order to activate a function, whereas passive interfaces make use of obtained gaze data in order to determine the amount of interest or attention exhibited by a user. Techniques for estimating gaze can be broken down into two categories, 2D gaze estimation and 3D gaze estimation, according to the type of coordinate system that is employed. The vast majority of work has been presented for estimating in two dimensions, whereas just a few research have concentrated on estimation in three dimensions for the purpose of producing reliable results in real-world contexts.

International Journal of Electrical and Electronics Engineers
Volume 16, Issue No. 01, Jan-June 2024
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

- **Multi-User Gaze Estimation**

A growing interest in gaze estimation in unconstrained contexts has been observed in the past decade as a result of the rapid use of deep learning-based methods in gaze estimation techniques. Because of this adaption, the idea of multi-user gaze estimation has been researched and implemented in a variety of different application domains [18]. Multi-user gaze estimation, as opposed to the more common practice of single-user gaze estimation, is typically necessary in open environmental contexts such as shopping, public meetings, and public events. As a result, it requires gaze estimate methods that are accurate, don't require a lot of overhead, and are fast. Studies that have already been conducted on multiple users' gaze patterns can be divided into two distinct categories: those that share time and those that share space.The number of users is spread out evenly across a given time period using the time-sharing method. The space sharing method, on the other hand, can process several users at the same time. Due to their lack of scalability and overall stability, time-shifting methodologies have not received a lot of attention in the relevant academic research.[19]

- **Gaze estimation approaches**

The various methods of gaze estimate that are now in use can be grouped into one of two primary groups: appearance-based techniques and model-based techniques. The cornea, optical axes, and visual axes are only few of the ocular components that are accounted for in model-based gaze estimation techniques. These techniques make use of a geometric model of the eye to determine the subject's direction of look. Although model-based approaches of gaze estimate are more accurate, using them often requires personal calibration from each participant, which can be a time-consuming process. Appearance-based approaches typically call for user eye images to be used in order to immediately develop a mapping function that goes from eye image to gaze estimation [21].Since the mapping is done directly on the image of the user's eye, appearance-based approaches often do not require camera calibration and geometry data. Conventional appearance-based methods and appearance-based methods combined with deep learning are the two categories that may be used to classify appearance-based methods. The abstract notions of these two types of approaches are illustrated in Figure 2 and Figure 3, respectively.

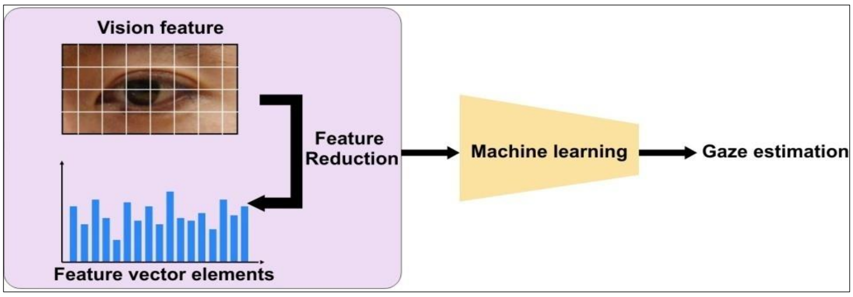

Figure 2: Conventional appearance based methods.

International Journal of Electrical and Electronics Engineers
Volume 16, Issue No. 01, Jan-June 2024
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

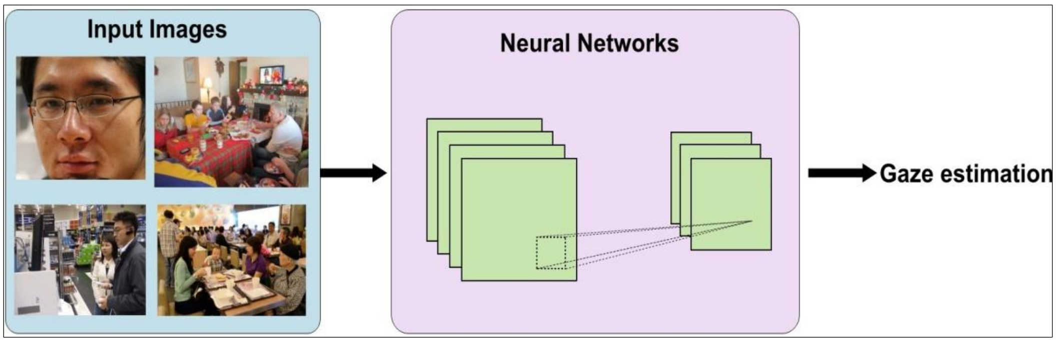

Figure 3: Appearance based methods with deep learning.

- **Conventional Appearance Based Methods**

In conventional appearance-based techniques, full images are treated as features, and eye gaze is inferred directly from the images themselves. Traditional approaches that are based on appearance have traditionally made use of mapping functions such adaptive linear regression, k-nearest neighbor, random forest regression, artificial neural networks, gaussian processors, and support vector machines.a technique known as adaptive linear

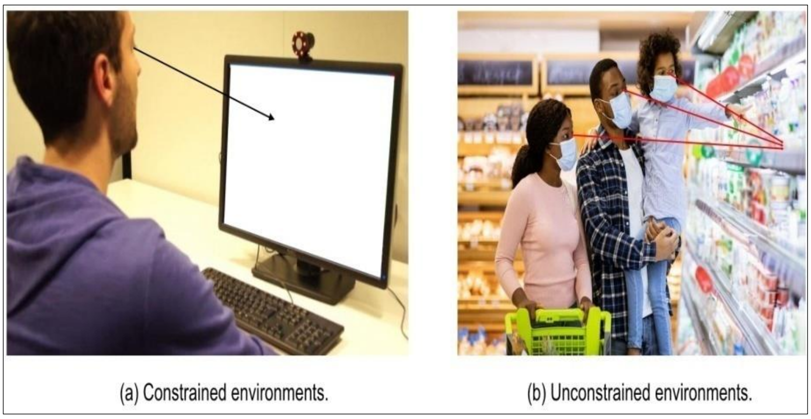

Figure 4: Constrained environment and Unconstrained environment

regression (ALR) has been described for mapping high-dimensional aspects of the ocular picture to low-dimensional gaze positions. This technique considerably decreases the number of training samples required for high accuracy estimate. k-Nearest Neighbors has become a typical approach in the conventional appearance-based method for forecasting gaze by using the mean of gaze angles from neighbor samples. This method is based on the concept that neighbors have similar gaze patterns. [22] have proposed a paradigm for gaze estimation that incorporates neighbor selection as well as neighbor regression. It makes considerable use of information regarding the posture of the head, the center of the pupil, and the overall look of the eyes [23].have suggested a way to learning the highly non-linear mapping function between the gaze information and the RGB eye picture appearances, including depth cues. This approach is built on an ensemble of trees that are gathered together in a single forest. a strategy that is based on particle swarm optimization with a BP neural network was proposed. These procedures are fraught with a great deal of difficulty. Most Conventional approaches that are based on appearance typically need either a fixed head stance or a restricted range of head movements, as shown in Figure 4(a). In addition, this technique struggles to account for subject differences, particularly when applied in an unrestricted setting.

III. CONCLUSION

The commercial, social, and medical health sectors are only few of the application domains that can benefit from eye gaze estimation technologies. This survey focused primarily on investigating the most cutting-edge methods currently employed in eye gazing research, with a particular emphasis on deep learning strategies. In this study, utilizing deep learning techniques, a critical analysis of the associated models in appearance-based gaze estimation approaches was carried out. In unconstrained environment situations, such as extreme head-pose variations, illumination conditions, eye and facial occlusions, appearance-based methods with deep learning perform more robustly than model-based methods and conventional appearance-based methods. This is due to the fact that appearance-based methods with deep learning combine machine learning techniques with traditional appearance-based approaches. In addition to this, they are able to learn a complicated non-linear mapping function directly from picture data to gaze without the need for a device that is specifically designed for this purpose. It was found that single-user gaze estimating methods have been extensively researched in both confined and unconstrained settings, obtaining near-human levels of performance in both types of situations. However, multi-user gaze estimation studies have only been investigated in a limited number of application domains, such as retail and crowd-behavior analysis. In addition, we have discussed the benefits and drawbacks of associated methodologies, as well as the characteristics of datasets that are readily available to the public. Finally, we have explored prospective future research topics and provided ideas for selecting eye gaze estimation methodologies. These can be helpful for researchers and developers working in the field.

REFERENCES

[1.] Akinyelu, A. A., & Blignaut, P. (2020). Convolutional Neural Network-Based Methods for Eye Gaze Estimation: A Survey. *IEEE Access*, *8*, 142581–142605. doi:10.1109/ACCESS.2020.3013540.

[2.] Bermejo, C., Chatzopoulos, D., & Hui, P. (2020). EyeShopper: Estimating Shoppers' Gaze using CCTV Cameras. *MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia*, (pp. 2765–2774). doi:10. 1145/3394171.3413683.

[3.] Cazzato, D., Leo, M., Distante, C., & Voos, H. (2020). When i look into your eyes: A survey on computer vision contributions for human gaze estimation and tracking. *Sensors (Switzerland)*, *20*, 1–42. doi:10.3390/ s20133739.

[4.] Chao, P., Kao, C.-Y., Ruan, Y.-S., Huang, C.-H., & Lin, Y.-L. (2019). Hardnet: A low memory traffic network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3552–3561).

[5.] Cheng, Y., Wang, H., Bao, Y., & Lu, F. (2021). Appearance-based Gaze Estimation With Deep Learning: A Review and Benchmark, . (pp. 1–21). URL: http://arxiv.org/abs/2104.12668. arXiv:2104.12668.

[6.] Chennamma, H. R., & Yuan, X. (2013). A Survey on Eye-Gaze Tracking Techniques, *.4*, 388–393. URL:

[7.] Chong, E., Ruiz, N., Wang, Y., Zhang, Y., Rozga, A., & Rehg, J. M. (2018). Connecting Gaze, Scene, and Attention: Generalized Attention Estimation via Joint Modeling of Gaze and Scene Saliency. *Lecture*

Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), *11209 LNCS*, 397–412. doi:10.1007/978-3-030-01228-1_24. arXiv:1807.10437.

[8.] Chong, E., Wang, Y., Ruiz, N., & Rehg, J. M. (2020). Detecting attended visual targets in video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5396–5406).

[9.] De Silva, S., Dayarathna, S., Ariyarathne, G., Meedeniya, D., Jayarathna, S., & Michalek, A. (2021). Computational Decision Support System for ADHD Identification . *International Journal of Automation and Computing (IJAC)*, *18*, 233–255. doi:10.1007/s11633-020-1252-1.

[10.] De Silva, S., Dayarathna, S., Ariyarathne, G., Meedeniya, D., Jayarathna, S., Michalek, A., & Jayawardena, G. (2019). A rule-based system for ADHD identification using eye movement data. In *Proceedings of the Moratuwa Engineering Research Conference (MERCon)* (pp. 538–543). doi:10.1109/MERCon.2019.8818865.

[11.] Everingham, M., Gool, L. V., Williams, C. K. I., Winn, J., & Zisserman, A. (2009). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, *88*, 303–338.

[12.] Fang, Y., Tang, J., Shen, W., Shen, W., Gu, X., Song, L., & Zhai, G. (2021). Dual attention guided gaze target detection in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11390–11399).

[13.] Fischer, T., Chang, H. J., & Demiris, Y. (2018). Rt-gene: Real-time eye gaze estimation in natural environments. In

[14.] *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 334–352).

[15.] Gamage, G., Sudasingha, I., Perera, I., & Meedeniya, D. (2018). Reinstating dlib correlation human trackers under occlusions in human detection based tracking. In *Proceedings of the 18th International Conference on Advances in ICT for Emerging Regions (ICTer)* (pp. 92–98). doi:10.1109/ICTER.2018.8615551.

[16.] Ghani, M. U., Chaudhry, S., Sohail, M., & Geelani, M. N. (2013). Gazepointer: A real time mouse pointer control implementation based on eye gaze tracking. In *Proceedings of the 16th International Multi-Topic Conference* (pp. 154–159). IEEE.

[17.] Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs.

[18.] *International journal of industrial ergonomics*, *24*, 631–645.

[19.] Graves, A., Fernández, S., & Schmidhuber, J. (2005). Bidirectional lstm networks for improved phoneme classification and recognition. In *International conference on artificial neural networks* (pp. 799–804). Springer.

[20.] Guojun, Y., & Saniie, J. (2016). Eye tracking using monocular camera for gaze estimation applications. *IEEE International Conference on Electro Information Technology*, *2016-August*, 292–296. doi:10.1109/EIT.2016. 7535254.

[21.] Gwon, S. Y., Cho, C. W., Lee, H. C., Lee, W. O., & Park, K. R. (2013). Robust eye and pupil detection method for gaze tracking. *International Journal of Advanced Robotic Systems*, *10*, 98.

[22.] Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Huang, S., Brooks, M., Lee, M. J., & Asadi, H. (2019). Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *American Journal of Roentgenology*, *212*, 38–43.

[23.] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

[24.] Huang, Q., Veeraraghavan, A., & Sabharwal, A. (2017). Tabletgaze: dataset and analysis for unconstrained appearance- based gaze estimation in mobile tablets. *Machine Vision and Applications*, *28*, 445–461.

[25.] Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE transactions on vehicular technology*, *53*, 1052–1068.