Continuous Non-Invasive Eye Tracking In Intensive Care

Ahmed Al-Hindawi[1,2], Marcela P Vizcaychipi[2,3] and Yiannis Demiris[1]

Abstract— Delirium, an acute confusional state, is a common occurrence in Intensive Care Units (ICUs). Patients who develop delirium have globally worse outcomes than those who do not and thus the diagnosis of delirium is of importance. Current diagnostic methods have several limitations leading to the suggestion of eye-tracking for its diagnosis through *in-attention*. To ascertain the requirements for an eye-tracking system in an adult ICU, measurements were carried out at Chelsea & Westminster Hospital NHS Foundation Trust. Clinical criteria guided empirical requirements of invasiveness and calibration methods while accuracy and precision were measured. A non-invasive system was then developed utilising a patient-facing RGB camera and a scene-facing RGBD camera. The system's performance was measured in a replicated laboratory environment with healthy volunteers revealing an accuracy and precision that outperforms what is required while simultaneously being non-invasive and calibration-free The system was then deployed as part of CONfuSED, a clinical feasibility study where we report aggregated data from 5 patients as well as the acceptability of the system to bedside nursing staff. To the best of our knowledge, the system is the first eye-tracking systems to be deployed in an ICU for delirium monitoring.

I. INTRODUCTION

Delirium is an acute confusion state that is a fluctuating, usually reversible, cause of cerebral dysfunction that manifests clinically with a wide range of neuropsychiatric abnormalities. This state can occur in any acutely unwell patient but occurs with high incidence on Intensive Care Unit (ICU) owing to the acuity of diseases [1]. An estimate of the incidence of delirium in acutely unwell patients is 20% and has been reported as high as 80% [2], [3].

The development of delirium in patients leads to increased risk of dying, increased hospital length of stay, increased cost of stay and worse cognitive scores compared to patients who do not [4], [5], [2]. Thus, the contemporaneous diagnosis of delirium is of paramount importance. Current methods of diagnosis are laborious and often miss episodes of delirium due to the disease being time-variant, or are only applicable to a small subset of patients.

To aid in the timely diagnosis of delirium, we propose a non-invasive system that relies on eye-tracking as a biological signal of *inattention* as a surrogate marker for delirium. The system is continuous, provides data in real-time, does not require patient involvement for calibration, is

[1] Personal Robotics Laboratory, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, United Kingdom. a.al-hindawi@imperial.ac.uk. YD is supported by a Royal Academy of Engineering Chair in Emerging Technologies.

[2] Magill Department of Anaesthesia, Chelsea & Westminster Hospital NHS Foundation Trust, London SW10 9NH, United Kingdom.

3 Anaesthetics, Pain Medicine and Intensive Care, Department of Cancer & Surgery, Imperial College London, London SW7 2AZ, United Kingdom.
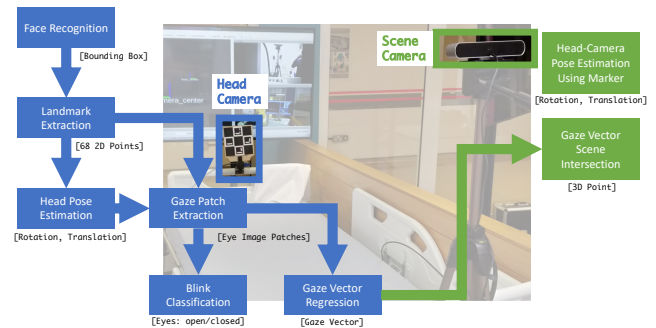
Fig. 1: The proposed system for eye tracking in ICU. Two cameras are used, an RGB patient facing camera (head-camera, blue) and a perspective facing RGBD camera (scene-camera, green) connected to a laptop (outside the viewpoint) running ROS. The flow of information through the system is demonstrated with the head-camera and scene-camera running concurrently.

accurate and does not limit or restrict movements. Empiric measurements in an adult ICU led to a set of requirements around invasiveness, calibration, accuracy, and precision. Under replicated laboratory conditions, the performance of the system was shown to be suitable for use in an adult ICU where it outperforms requirements. Following ethical and governance approvals for a clinical feasibility study (CONfuSED[1]), the system was deployed on patients in ICU to test its suitability. We report on descriptive statistical data on the initial 5 patients. We also report on the acceptability of the system to the bedside nursing staff.

II. RELATED WORK & DIAGNOSTIC METHODS

Manual assessment methods, specifically the Confusion Assessment Method in ICU (CAM-ICU), are currently the most widely adopted systems in use for the diagnosis of delirium [6], [7]. The test relies on *inattention* as its primary method of diagnosis of delirium.

Electroencephalogram (EEG) is of limited use owing to technological limitations but more recent attempts use deep neural networks for the diagnosis of delirium in a clinically representative population [8], [9], [10]. EEG, however, requires the placement of electrodes on a patient's scalp for a prolonged period to ensure adequate signal acquisition hampering its clinical safety.

Eye-tracking technology has had success in inferring *covert* internal mental states [11], [12]. Clinically, eye-tracking is established in numerous fields where its success

[1]https://clinicaltrials.gov/ct2/show/NCT04589169

has led to the suggestion that it can be used for the diagnosis of delirium through *inattention* [13], [14]. The application of eye-tracking in ICU has been limited by the technological limitations of currently available systems. To the knowledge of the authors, no suitable system has been developed or implemented that facilitates the acquisition of eye movements from delirious patients in a clinical setting.

III. SYSTEM REQUIREMENTS

ICU presents a challenge for eye-tracking technology. The mixture of patient, disease, environment, staff and medical equipment regulations all coalesce together placing the following clinical empirical requirements:

a) Non-Invasive: Due to the nature of patients and their disease in ICU, a non-invasive system is required owing to the need for clinical hands-on care. Having a device that instruments the patient will interfere with clinical care and is thus potentially unsafe. This also infers that a device that facilitates free-viewing is required – a system that does not restrict the patient for signal acquisition.

b) Calibration-Free: In delirium, patients are unable to follow commands making calibration through command-following difficult. The system must be able to track the patient's gaze without an explicit calibration step that maps eye movements to scene gaze positions.

c) Accurate: The required accuracy of a system is dependent on its domain use; In ICU, accuracy facilitates the identification of *which* object the patient is looking at and correlates with the *spread* of objects in the patient's perspective.

d) Precise: Precision, defined as the reproducibility of measurement, is important as low precision leads to uncertainty on the gaze estimate. Low uncertainty facilitates subsequent analysis through the reliable classification of eye movements. In this context, the *size* of objects from the point of view of the patient correlates with precision. High precision leads to reproducible measurements which leads to low spread and the ability to identify small objects.

e) Performant: Eye movement research suggests that fixations typically last between 0.1 - 0.5 seconds giving a minimum required frequency of 20Hz by Nyquist criteria [15]. This system would ideally run on commercially available hardware.

IV. METHODOLOGY

To meet the challenges set out by the empirical requirements, we have designed a system that decouples the regression of the patient's eye and head pose from where they are looking in the scene through the use of a dual-camera system as per Figure 1. One camera located in front of the patient at the foot end of the bed facing the patient (termed the head-camera) and another behind the patient facing the same direction as the patient (termed the scene camera). The head-camera is a Red-Green-Blue (RGB) camera whereas the scene-camera is required to be a Red-Green-Blue-Depth (RGBD) camera. The decoupling of the eye gaze regression and gaze inference through the use of two cameras satisfies the first requirement of *Non-invasiveness* as per Section III.
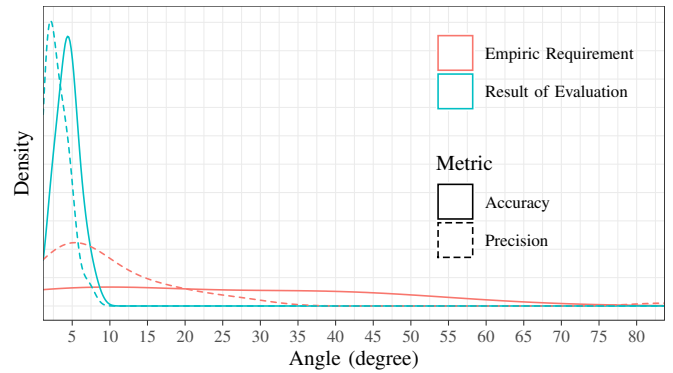

Fig. 2: Density plots demonstrating the required accuracy and precision from empirical measurements and the system's metrics as evaluated in the replicated environment in the laboratory.

a) Scene-Camera: The scene-camera performs head-camera pose estimation and gaze vector-scene intersection. The head camera's pose is measured using a ChAruCo board which is fixed superior to the head-camera – a once only calibration procedure yields a static pose between the head-camera and its board [16]. The resolution of the scene camera is dependent on its ability to resolve this marker reliably - we utilise a 1920x1080 pixel resolution at 30Hz. Once the gaze-vector has been estimated from the head camera, it is the scene camera that resolves that vector as a point in the scene. This is done using a ray-octree intersection that is created from the point cloud received from the scene camera. [17].

b) Head-Camera: The patient-facing *head-camera* carries out multiple steps that ultimately result in the head-pose and gaze-vector of the patient relative to the head camera. The head camera's pose as estimated from the scene camera creates a transformation tree that facilitates the location of the patient's gaze relative to the scene camera which acquires the point of view of the patient.

The first step is face detection. Following experimental evaluation between various face detectors, the Single Shot Scale-invariant Face Detector (S^3FD) face detector was used for its robust detection at extreme head poses [18]. The output is a bounding box around potential faces and the confidence of each detection. The camera specification is linked to its ability to reliably produce a high-resolution image of the face - we utilise a 1024x768 pixel resolution at 50Hz.

Following face-detection, the patient's head-pose is then estimated by solving a perspective-n-point problem that aims to align a generic 3D model of a human's head with the extracted landmarks from the patient [19]. Multi-Task Cascaded Convolutional Networks for Joint Face Detection and Alignment (MTCNN), Fine-Grained Head Pose Estimation Without Keypoints (HopeNet), and Face Alignment in Full Pose Range: A 3D Total Solution (3DDFA) were evaluated for head-pose estimation using a composite score comprised of accuracy, performance at inference time and utilisation of

Variable	Validation (n=6)	CONfuSED (n=5)
Age	25 - 32	18 - 73
Gender (male = 1)	5	3
Vision Corrected (Glasses = 1)	2	2

TABLE I: Demographics data of participants in validation experiment and patients in CONfuSED.

GPU Memory. 3DDFA had the lowest composite score and hence was chosen. Its output is 68 2D landmark points in image space that consistently outline the facial features of the patient's face. By knowing the camera calibration matrix and having a predefined model, a direct least squares method is utilised following a Random Sample Consensus (RANSAC) scheme to obtain the pose of the patient's head relative to the head-camera [20], [21]. The 3D model's size is parametrised by the patient's inter-pupillary distance with 0.06 metres used as the default. The landmarks extracted from the previous stage also result in the extraction of two image patches of the patient's eyes which are then used for gaze regression. Real-Time Eye Gaze Estimation in Natural Environments (RT-GENE), a neural network-based model that regresses gaze vector from eye patches, previously developed in our laboratory, was chosen for its published accuracy in natural viewing environments while Real-Time Blink Estimation in Natural Environments (RT-BENE) is used for blink detection following improvement through a data augmentation scheme and the use of a ResNet-18 as the backbone for increased speed and accuracy [?], [22]. RT-GENE was modified in this work to optimise inter-process communication and reduce latency between image capture and gaze estimation. The resolution of the camera (1024x768) results in eye-patches that are down-sampled for the consumption through RT-GENE/RT-BENE at a clinically safe distance of approximately 2 metres. The use of RT-GENE and RT-BENE in this pipeline satisfies the requirement for the system being *Calibration-Free* as per Section III.

The required accuracy and precision from an eye tracker in ICU were measured in situ. Measurements of the scene from the point of view of 5 patient from 5 different environments were undertaken across our ICU. Accuracy, as defined as the difference between a target location and the measured gaze relative to the user's head correlates with the identification of *which* object in the scene the patient is looking at and thus, given a collection of objects in the scene, the average *distance* between objects is of interest. The distances between objects in the viewpoint of the patient and their relative distance to the patient's head were measured. The average of these measurements then form the minimum required accuracy of the system. To ascertain a similar requirement for precision, the size of objects and their distance to the patient's head was measured. In this context, the size of the object in the scene correlates to repeatability – the smaller the size of the object, the higher the precision required for the identification of the object the patient is looking at. The average of these measurements then forms the required precision. Formally:

$$Accuracy = \frac{1}{N} \sum_{i=1}^{N} cos^{-1}(|gaze_i| \cdot |target|) \quad (1)$$

$$Precision = \sqrt{\frac{1}{N} \sum_{i=1}^{N-1} cos^{-1}(|gaze_{i+1}| \cdot |gaze_i|)^2} \quad (2)$$

where $\cdot$ is the vector dot product and $|x|$ is vector normalisation to unit length. The gaze vector is measured to start from the head pose's transform and extends to a unit length. The target vector is measured to also start from the head pose transform and extends to the target.

To measure the performance of the system, six healthy, vision-corrected, non-delirious volunteers were recruited to undergo testing of the accuracy and precision of the system (Table I). An ICU replicated environment was created under laboratory conditions and the bed was inclined at 30 degrees. The participant's head-pose and gaze vector were regressed from the head camera placed on the bottom left-hand side of the screen. Each participant was asked to look at the centre of an AruCo marker that would appear at a random location of the screen that was 0.1 metres in size; 40 samples were taken per marker. The experimental endpoint was defined as full coverage of the screen by target markers. For each marker location, outliers consisting of the first 10 and last 10 measurements were discarded concentrating on the middle 20 measurements to remove gaze points leading into and out of the target marker. For each marker location, the accuracy and precision were calculated as per Equations 1 & 2. For visualisation, median accuracy for each marker is then rasterised onto an image with the resolution of the screen which is then smoothed with a Gaussian kernel.

As part of a clinical feasibility pilot trial that aims to look at the use of this system on the detection of delirium in ICU, the system was deployed on 5 patients. We report aggregated descriptive statistical plots to demonstrate the performance of the system in patients suffering from episodes of delirium compared to episodes without delirium.

In assessing the acceptability of the system clinically, a questionnaire was deployed to the staff that were nursing the patients recruited into CONfuSED (Table II) [23].
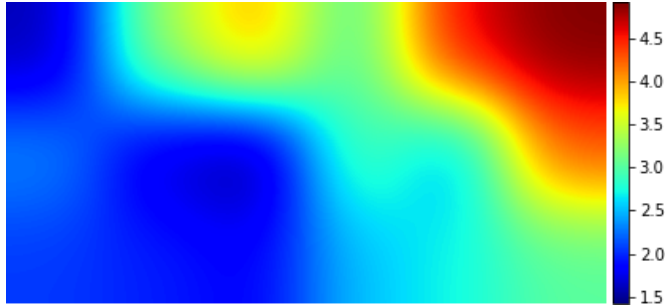
V. RESULTS

By utilising two cameras placed at a clinically safe distance, the system meets the clinical requirement of being non-invasive. RT-GENE and RT-BENE ensure that patient calibration is not required for the accurate and continuous measurement of gaze throughout the patient's ICU stay.

Figure 2 demonstrates the distribution of required accuracy and precision measurements. The required accuracy & precision density plot was of measured objects from the perspective of the patient in an ICU bed-space. Pairwise distances were measured as the shortest distance between objects. Size/Distance measurements were then converted to angles from the perspective of the patient as they are
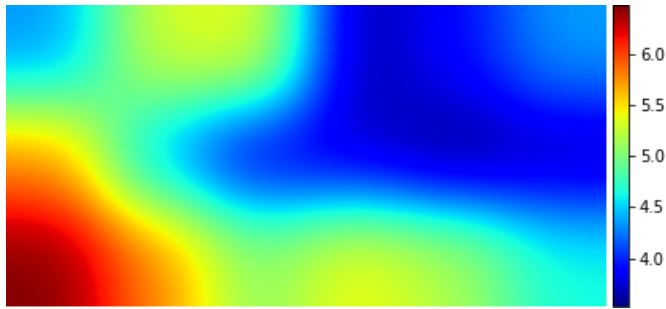
(a) Accuracy result per marker aggregated over all participants

(b) Precision result per marker aggregated over all participants

Fig. 3: Aggregated results for all participants demonstrating the accuracy and precision per marker where X & Y axes represent the ArUco marker location on screen in Pixels. The sensitivity of the camera's position relative to the viewer is evident where the highest accuracy and precision is when the user's head pose is directed at the camera.

sitting in an ICU bed. This reveals that an accuracy of 15.4° and a precision of 6.4° is thus required for a clinically useful system. Figure 2 also demonstrates the accuracy and precision results from the replicated laboratory ICU across all participants pooled across all marker locations. Median accuracy is 4.4° with a median precision of 2.7° as defined by Equations 1 & 2 respectively. These results exceed empiric requirements making the system's performance suitable for use in ICU.

In the laboratory experiment, grouping by markers, Figure 3 displays aggregated results across all participants where each marker's median accuracy and precision are rasterised as an image the size of the marker and then smoothed with a Gaussian kernel. As the utility of this eye-tracking system is in ICU, a median error of 4.4° is lower than the 15.4° . The figure demonstrates the sensitivity of the camera's placement relative to the user's head as the error of the system is not uniform; median accuracy is lowest at the bottom left which, from the perspective of the head-camera, represents the most challenging eye image to regress as the eye is mostly closed. Median precision is worse in the top-right where the eye is the most open but the head is rotated giving only 1 eye to regress a gaze vector thus giving the least precision.

Optimisation of the pipeline through internal validation experiments led to acceptable performance on commercially available hardware. The total rate of 28Hz is of sufficient performance to yield to eye movement analysis.

Figure 4 demonstrates the measurements that the system generates on patients in ICU. The system was successfully deployed on 5 patients where some of whom developed episodes of delirium. Blinks, eye and head angles are displayed as aggregated density plots. Extreme head angles are as a result of occlusions of the patient's face - examples of such occlusions are the patient's hands occluding their face, nurses interrupting the line of sight between the camera and the patient and nasogastric tubes. These periods of occlusions could be mediated by increasing the threshold for face detection above the current value of 0.6 thus increasing the specificity of detections ignoring periods of occlusion.

The figure also demonstrates the aggregate statistics from the datasets used to create RT-GENE and RT-BENE alongside. The distributions of blinks and gaze data are similar across the patients and the dataset indicating likely accurate acquisition.

The questionnaire outlined in Table II was conducted with 11 nursing staff, all of whom reported "No" to every question, indicating the acceptability of the system in clinical practice.

VI. CONCLUSION

Delirium is a common occurrence in ICU with tests that are laborious or too restricted. Eye-tracking has been suggested to diagnose *inattention* as a surrogate marker of delirium but to date (and best of our knowledge), no system has met the requirements to be suitable for use in ICU. We thus developed a system that meets empiric and measured criteria as being non-invasive, calibration-free and exceeds the required accuracy and precision all while running on commercially available hardware.

Current and future work focuses on the use of the signals acquired by this system to automate the diagnosis of delirium.

REFERENCES

[1] K. D. Krewulak, H. T. Stelfox, E. W. Ely, and K. M. Fiest, "Risk factors and outcomes among delirium subtypes in adult ICUs: A systematic review," *Journal of Critical Care*, vol. 56, pp. 257–264, Apr. 2020.

[2] S.-M. Lin, C.-y. Liu, C.-h. Wang, H.-c. Lin, C.-D. Huang, P.-y. Huang, Y.-f. Fang, M.-h. Shieh, and H.-p. Kuo, "The impact of delirium on the survival of mechanically ventilated patients." *Critical care medicine*, vol. 32, no. 11, pp. 2254–9, 2004.

[3] M. A. Pisani, K. L. B. Araujo, P. H. Van Ness, Y. Zhang, E. W. Ely, and S. K. Inouye, "A research algorithm to improve detection of delirium in the intensive care unit." *Critical care (London, England)*, vol. 10, no. 4, p. R121, Aug. 2006.

[4] O. Clancy, T. Edginton, A. Casarin, and M. P. Vizcaychipi, "The psychological and neurocognitive consequences of critical illness. A pragmatic review of current evidence," *Journal of the Intensive Care Society*, vol. 16, no. 3, pp. 226–233, Aug. 2015.

Question	Answer
Did this study affect your ability to care for your patients?	Yes/No
Did this study affect your interactions with patients/families?	Yes/No
Did this study affect your interactions with other nurses?	Yes/No
Did this study affect your interactions with doctors?	Yes/No

TABLE II: This questionnaire used assess the acceptability of the use of cameras within this study on day to day activities of the nursing staff on ICU.
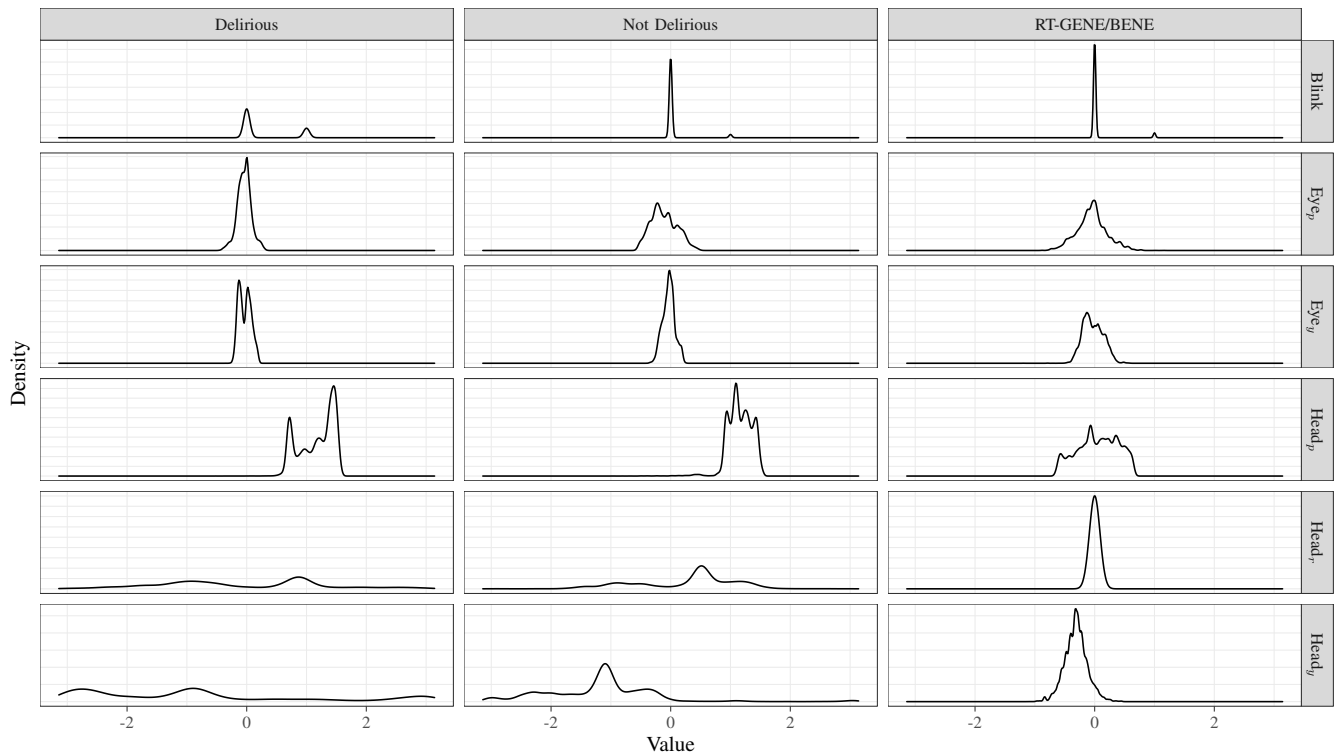
Fig. 4: Density plots of aggregated data from 5 patients. Some experienced episodes of delirium. Eye and Head refer to Eye and Head pose angles from the system while subscripts y, p and r refer to yaw, pitch and roll respectively. Blink is thresholded as per RT-BENE. Blink and Gaze statistics are of similar distributions between RT-GENE/BENE and that of non-delirious patients. Of note, no $Head_r$ data is present in RT-GENE-RT-BENE datasets as faces are processed to remove roll prior to inference.

[5] J. McCusker, M. Cole, M. Abrahamowicz, F. Primeau, and E. Belzile, "Delirium Predicts 12-Month Mortality," *Archives of Internal Medicine*, vol. 162, no. 4, p. 457, Feb. 2002.

[6] E. W. Ely, S. K. Inouye, G. R. Bernard, S. Gordon, J. Francis, L. May, B. Truman, T. Speroff, S. Gautam, R. Margolin, R. P. Hart, and R. Dittus, "Delirium in Mechanically Ventilated Patients: Validity and Reliability of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU)," *JAMA*, vol. 286, no. 21, p. 2703, Dec. 2001.

[7] E. W. Ely, R. Margolin, J. Francis, L. May, B. Truman, R. Dittus, T. Speroff, S. Gautam, G. R. Bernard, and S. K. Inouye, "Evaluation of delirium in critically ill patients: Validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU)," *Critical Care Medicine*, vol. 29, no. 7, p. 1370, July 2001.

[8] A. W. van der Kooi, I. J. Zaal, F. A. Klijn, H. L. Koek, R. C. Meijer, F. S. Leijten, and A. J. Slooter, "Delirium Detection Using EEG," *Chest*, vol. 147, no. 1, pp. 94–101, Jan. 2015.

[9] H. Koponen, J. Partanen, A. Pääkkönen, E. Mattila, and P. J. Riekkinen, "EEG spectral analysis in delirium." *Journal of neurology, neurosurgery, and psychiatry*, vol. 52, no. 8, pp. 980–5, Aug. 1989.

[10] H. Sun, E. Kimchi, O. Akeju, S. B. Nagaraj, L. M. McClain, D. W. Zhou, E. Boyle, W.-L. Zheng, W. Ge, and M. B. Westover, "Automated tracking of level of consciousness and delirium in critical illness using deep learning," *npj Digital Medicine*, vol. 2, no. 1, pp. 1–8, Sept. 2019.

[11] P. V. Amadori, T. Fischer, R. Wang, and Y. Demiris, "Decision Anticipation for Driving Assistance Systems," in *IEEE International Conference on Intelligent Transportation Systems*, June 2020, p. 8.

[12] R. Wang, P. V. Amadori, and Y. Demiris, "Real-Time Workload Classification during Driving using HyperNetworks," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2018, pp. 3060–3065.

[13] P. Trillenberg, R. Lencer, and W. Heide, "Eye movements and psychiatric disease." *Current opinion in neurology*, vol. 17, no. 1, pp. 43–7, Feb. 2004.

[14] C. Exton and M. Leonard, "Eye tracking technology: A fresh approach in delirium assessment?" *International Review of Psychiatry*, vol. 21, no. 1, pp. 8–14, 2009.

[15] G. Feng, "Eye movements as time-series random variables: A stochastic model of eye movement control in reading," *Cognitive Systems Research*, vol. 7, no. 1, pp. 70–95, Mar. 2006.

[16] G. H. An, S. Lee, M.-W. Seo, K. Yun, W.-S. Cheong, and S.-J. Kang, "Charuco Board-Based Omnidirectional Camera Calibration Method," *Electronics*, vol. 7, no. 12, p. 421, Dec. 2018.

[17] A. S. Glassner, "Space subdivision for fast ray tracing," *IEEE Computer Graphics and Applications*, vol. 4, no. 10, pp. 15–24, Oct. 1984.

[18] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, "S3FD: Single Shot Scale-Invariant Face Detector," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 192–201.

[19] S. Ohayon and E. Rivlin, "Robust 3D Head Tracking Using Camera Pose Estimation," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 1, Aug. 2006, pp. 1063–1066.

[20] J. A. Hesch and S. I. Roumeliotis, "A Direct Least-Squares (DLS) method for PnP," in *2011 International Conference on Computer Vision*, Nov. 2011, pp. 383–390.

[21] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[22] K. Cortacero, T. Fischer, and Y. Demiris, "RT-BENE: A Dataset and Baselines for Real-Time Blink Estimation in Natural Environments," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019.

[23] M. MacMurchy, S. Stemler, M. Zander, and C. P. Bonafide, "Acceptability, Feasibility, and Cost of Using Video to Evaluate Alarm Fatigue," *Biomedical Instrumentation & Technology*, vol. 51, no. 1, pp. 25–33, Jan. 2017.