



Article

An Advanced Solution Based on Machine Learning for Remote EMDR Therapy

Francesca Fiani ¹, Samuele Russo ² and Christian Napoli ^{1,3,4,*}

¹ Department of Computer, Control and Management Engineering, Sapienza University of Rome, 00185 Roma, Italy; fiani@diag.uniroma1.it

² Department of Psychology, Sapienza University of Rome, 00185 Roma, Italy; samuele.russo@uniroma1.it

³ Institute for Systems Analysis and Computer Science, Italian National Research Council, 00185 Roma, Italy

⁴ Department of Computational Intelligence, Czestochowa University of Technology, 42-201 Czestochowa, Poland

* Correspondence: cnapoli@diag.uniroma1.it

Abstract: For this work, a preliminary study proposed virtual interfaces for remote psychotherapy and psychology practices. This study aimed to verify the efficacy of such approaches in obtaining results comparable to in-presence psychotherapy, when the therapist is physically present in the room. In particular, we implemented several joint machine-learning techniques for distance detection, camera calibration and eye tracking, assembled to create a full virtual environment for the execution of a psychological protocol for a self-induced mindfulness meditative state. Notably, such a protocol is also applicable for the desensitization phase of EMDR therapy. This preliminary study has proven that, compared to a simple control task, such as filling in a questionnaire, the application of the mindfulness protocol in a fully virtual setting greatly improves concentration and lowers stress for the subjects it has been tested on, therefore proving the efficacy of a remote approach when compared to an in-presence one. This opens up the possibility of deepening the study, to create a fully working interface which will be applicable in various on-field applications of psychotherapy where the presence of the therapist cannot be always guaranteed.



Citation: Fiani, F.; Russo, S.; Napoli, C. An Advanced Solution Based on Machine Learning for Remote EMDR Therapy. *Technologies* **2023**, *11*, 172. <https://doi.org/10.3390/technologies11060172>

Academic Editors: Mario Munoz-Organero, Gang Kou, Shuai Ding, Li Luo, Tian Lu and Yogesan Kanagasingam

Received: 24 October 2023
Revised: 23 November 2023
Accepted: 4 December 2023
Published: 6 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: machine learning; eye tracking; distance detection; camera calibration; trajectory estimation; EMDR; mindfulness

1. Introduction

In the last decade, requests for virtual psychotherapy have greatly increased, with one of the main new additions being the introduction of machine learning in trauma treatment. Machine learning could potentially assist in personalizing treatment approaches, analyzing patterns in patient responses and optimizing interventions based on individual needs. However, it also presents the benefit of increasing the efficiency of remote psychotherapy. The latest trend in the fruition of psychotherapy, especially during and after the COVID outbreak, has slowly evidenced the need for new approaches that can be operated even in situations where the physical presence of a therapist cannot be guaranteed. One such therapy is Eye-Movement Desensitization and Reprocessing (EMDR) therapy [1], a widely recognized treatment mainly for Post-Traumatic-Stress-Disorder-(PTSD)-suffering patients, which strictly requires the physical presence and evaluation of a therapist, due to some constraints in the formalized protocol of the therapy.

In EMDR, the process of “exposure” differs from simple exposure therapy. In traditional exposure therapy, the patient is gradually exposed to the feared or traumatic stimuli in a controlled manner, allowing them to confront and process the associated emotions. By contrast, EMDR incorporates bilateral stimulation, such as eye movements, during the exposure phase. The eye movement component in EMDR is believed to facilitate the processing of traumatic memories. Francine Shapiro, the developer of EMDR, proposed

that bilateral stimulation helps integrate and desensitize distressing memories. The exact mechanism is not fully understood, but it is theorized to be related to the natural information-processing capacities of the brain.

The EMDR protocol follows eight phases:

1. patient history;
2. preparation;
3. assessment;
4. desensitization;
5. installation;
6. body scan;
7. closure;
8. re-evaluation.

EMDR can also be applied to individuals experiencing mild distress. The protocol is adaptable, and the treatment steps can be modified to suit the needs of individuals with milder symptoms. The flexibility of EMDR allows therapists to tailor their approach, based on the specific requirements and comfort levels of each patient, ensuring effective and personalized treatment even in cases of mild distress. There are many reports in the literature of successfully applied EMDR therapies to mild-symptoms patients [2–4]. However, despite the mildness of the condition at hand, the EMDR therapeutic protocol is structured in the same manner, therefore requiring the already-introduced eight phases.

The desensitization phase of the protocol [5] is the core of EMDR, and it involves the patient focusing on a target memory while engaging in bilateral stimulation (such as eye movements) to reduce the emotional intensity associated with the memory. When performed through visual stimulation, phase 4 requires following closely the movement of the therapist's hand. Currently, the visual attention of the patients must be manually verified by the therapist, a task which can be imprecise due to the fatigue caused by the long duration of the stimulus administered by the therapist. Thanks to the most recent machine-learning techniques and, in particular, gaze recognition and tracking, we can automatically detect the engagement of the subjects in the desensitization task in the produced virtual environment. This allows us to analyze the subject's adherence to the prescribed movements, offering to the psychotherapist the possibility of evaluating the patient's reactions as assessed by the algorithm.

This study presents a preliminary work for the development of a computer-based system that will allow the performance of self-induced mindfulness meditation through the use of advanced and up-to-date machine-learning techniques. The task will be executed by administering a visual bilateral stimulation identical to the one used during the desensitization phase of the EMDR protocol, therefore proving the repeatability of our system for future EMDR use. Instead of focusing on trauma processing, we will focus on measuring and possibly lowering the stress levels of the involved subjects. The system will automatically gather data regarding the patient, measure the distance of the patient from the screen and ask the patient to reposition themselves to an ideal position, calibrate the camera to improve the quality of the eye tracking and, finally, execute the mindfulness protocol by performing visual bilateral stimulation through the use of a virtual light bar executing on the screen. The whole stimulation will be tracked by a computer-vision-based algorithm for eye tracking, which will generate a trajectory for the subject's gaze and compare it to the expected one, to determine if the patient is correctly executing the task.

Importantly, this study highlights that the role of the psychotherapist is not replaced or marginalized in any way. On the contrary, the therapist plays a pivotal role as the facilitator and guide throughout the virtual exercises. The virtual platform serves as a tool for constant guidance and supervision by the therapist, guaranteeing a personalized and supportive therapeutic experience. This collaborative approach leverages the synergy between technology and human expertise, enhancing the overall effectiveness of the virtual psychotherapy sessions. The virtual-therapy platform prioritizes user comfort, allowing patients to stop the exercises or the sessions at any moment if they experience discomfort

or distress. Moreover, patients have the option to switch to a mode where they can see and interact with the therapist directly, as in a normal video call, fostering a sense of familiarity and comfort. Obviously, in this latter case, it would be possible to take advantage of the technological support offered by our system. This flexibility allows us to tailor the therapeutic experience, based on the needs of the patients.

Several works are available in the literature, in regard to virtual interfaces for psychotherapy, especially in the context of the use of machine learning for their implementation. The applications and specific techniques are incredibly wide, with several applications being for diagnosis, screening and support of Alzheimer's disease and, in general, cognitive impairment. Such works tackle mainly visual, auditory and motor tasks, with a special emphasis on the mnemonic part, both with simpler methods, such as decision trees [6], and with more advanced ones, such as CNNs and GANs [7]. There are also some works tackling virtual interfaces specifically for eye tracking, with one of the most significant ones employing CNNs and transfer learning to identify Alzheimer's disease through the use of visual memory tasks that require analyzing a set of sequential images with slight differences between them [8]. Few-to-no works, however, tackle the problem of virtual interfaces specifically for visual BLS with applications to EMDR and the mindfulness protocol. Other available approaches for remote EMDR exist in the form of individual light bars or kits containing light bars, headphones and tactile stimulators; however, they present several drawbacks:

1. They often require specific software installations, and are not integrated in the hardware available to users;
2. They can be counterintuitive for some users, requiring long instruction manuals and a high number of cables to function;
3. They have high costs and are, therefore, not available to all users;
4. The light bars and EMDR kits are often shipped from a limited number of countries (e.g., the USA), therefore being hard to access worldwide.

In addition to physical hardware, several applications, both commercial and non-commercial, also exist online, to perform remote EMDR with the aid of a virtual light bar in the form of a moving dot. However, all the approaches available only provide a visual interface, with the requirement for the therapist to manually check the patient's execution of the task on a video call. Due the current infrastructures available, however, the data transmission is not completely real-time, and delays in the transmission of the patient's video can cause reduced efficacy of the treatments, due to the slowed-down response of the therapist to the subject's distractions. Our instrument, on the other hand, is not only real-time and integrated in the hardware already possessed by the subjects, with the possibility of executing the application on any device provided with a screen and a camera, but will also be actuated with the aid of machine learning. This will effectively contrast any possible delays in transmission of the video data to the therapist by performing a verification of the patient's adherence to the prescribed movements directly on the client side, with the therapist performing a validation on the automatically registered distractions. In our work, we made use of three different machine-learning techniques that were joined to create a smooth architecture: namely, distance detection, camera calibration and eye tracking. They provided a solid infrastructure compared to those already available, able to simulate a virtual light bar and jointly verify the subject's adherence to the stimulation.

In regard to distance detection, many approaches available in the literature simply rely on hardware measurements, but many works, especially in regard to face detection, prefer metrics based on facial features. The main features used concern eyes—for example, measurements of the distance between eyes' corners [9] or between the pupils [10]. Eyes are, however, not the only possibly usable feature, with another popular option being the measurement of the distance between face landmarks, such as the nose and the mouth [11]. Due to the absent repeatability of such a tailored approach, other metrics that do not rely on facial features exist, such as through the measurement of the distance of an object of known dimension on the scene. An excellent example of use which can also bypass the

standardized-dimension problem is Dist-YOLO, a modification of YOLOv3 [12] that adds to the network the possibility of measuring the distance of the bounded objects [13]. Finally, another recent option uses point-cloud data to estimate distance [14], creating a Bird's Eye View (BEV) representation that is used as the input of a standard YOLOv4 network [15].

To perform camera calibration, the most common technique used in the literature is the use of dots positioned at strategic points of the screen called points of regard, with their configuration depending on the type of application [16]. The more common schemas are three-point calibration, generally used for single-line text reading, five-point calibration and nine-point calibration, mainly used for wide-vision applications with five-point calibration being a simplified nine-point calibration and, finally, thirteen-point calibration, mostly used for paragraph reading.

In regard to eye tracking, finally, we have a wide variety of approaches already available in the literature. Eye tracking can be tackled not only with machine learning, but also with the Internet of Things and cloud computing [17]. While the Internet of Things and cloud computing mainly focus on improving the reliability of the system and computation speed, machine learning instead offers several flexible approaches. Symbolic regression, in particular, shows great results in improving the calibration phase [18]. Better approaches are available in the form of neural networks, such as Convolutional Neural Networks (CNN) and Deep Neural Networks (DNN), but they require a large amount of data that are not easily available. Among them, many pre-existing deep CNNs work optimally for vision and perception tasks. Some possibilities are VGGNet [19], which has also been compared to AlexNet, to check their accuracy in fine-tuning [20]; ResNet [21]; U-Net [22], which is generally combined with DenseNet and obtains similar results to SegNet [23]; and, finally, Graph Convolutional Networks (GCN) [24], a relatively new approach that joins classic CNNs to Graph Neural Networks (GNN), with the advantage of being able to naturally fuse different groups of features, to have more stable training.

In addition to standard CNN approaches, a relatively new approach to eye tracking is the use of Recurrent Neural Networks (RNN) [25], which improve predictions by analyzing the previous frames of the captured video, to perform a contextual analysis, especially in a mixed approach (CNN + RNN). Finally, Generative Adversarial Networks (GAN) have also shown good results, with CycleGAN being particularly efficient in eye tracking, due to solving the problem of standard GANs in tracking accuracy by introducing a double processing of the image, therefore making accuracy trackable [26].

The already-mentioned approaches can also extend to gaze tracking, a subclass of the eye-tracking problem, which accurately estimates pupil position and the direction of the gaze, allowing the identification of the point on the screen where they are focusing their attention [27]. There are several gaze-tracking and estimation methods [28], which can be categorized as 2D-mapping-based methods, 3D-model-based methods and appearance-based methods. Standard vision approaches are also functional as a solution to the gaze-tracking problem, with much emphasis being put on the pre-processing and analysis of the frames, and various papers have tackled the problem with vision approaches.

This work had two main aims. Firstly, we wanted to verify the efficacy of our eye-tracking algorithm in reconstructing a trajectory generated by fast eye movements in following a moving target and catching the distractions of the subjects. Secondly, after verifying the quality of the approach, we wanted to validate our thesis that a virtual approach can be comparable to an in-presence one and performs better than other relaxation methods, such as diaphragmatic breathing or a distraction task, such as filling in a questionnaire. Our work confirmed that the virtual approach performs excellently in decreasing the stress levels of the subjects selected for this study compared to our control group and diaphragmatic breathing ones. Moreover, our algorithm, when operated under the correct environmental conditions, is able to correctly discern the number of distractions of the patients. This effectively makes our approach suitable to be tested on EMDR therapy, given the similarity in tasks between the desensitization phase of the protocol and the mindfulness meditation,

with the possibility of furthering the work, to generate a fully virtual interface for the whole EMDR protocol.

This study was also attentive to potential constraints, obstacles and ethical concerns related to the utilization of virtual interfaces in psychotherapy. In the results section, we provide a detailed overview of the measures taken to ensure ethical standards and address any potential issues. Additionally, the discussion section includes a thorough exploration of the limitations and challenges encountered during the study. We acknowledge the importance of ethical considerations in virtual therapy and propose remedies and strategies to mitigate these concerns. Our aim is to contribute to the ongoing dialogue on ethical practices in virtual psychotherapy and to offer insights for future research in this domain.

The article will be structured as follows: Section 2 will analyze the development of the application, with a particular focus on the data extraction and the choice of algorithms for distance detection, calibration and eye tracking. Section 3 will briefly introduce our testing environment and subjects, presenting the results obtained. Section 4 will focus on the analysis and comparison of the obtained data, with a mention of the future directions of this work. Finally, Section 5 highlights the conclusions of this study.

2. Materials and Methods

Our interface was developed to account for three different study groups. Group 1 performed a desensitization task, comprised of a dot-following exercise followed by controlled breathing (see in Figure 1). The dot-following was thoroughly analyzed by our eye-tracking algorithm. Group 2 performed diaphragmatic breathing, as instructed by the interface. Finally, Group 3 filled in a questionnaire.

The algorithm described in this section was implemented in the context of a web application developed with HTML, CSS, PHP and JavaScript. Communication between the web application and the Python scripts responsible for the machine-learning part was implemented via the Flask framework. Our system operates in real-time, ensuring that tasks are executed without noticeable delays. It is important to note that the current state of data transmission and processing within our system is either below or equivalent to network latency. In other words, the time taken for data to be transmitted and processed is comparable to or slightly faster than the delays experienced in the network. The data regarding the subjects was stored in an SQL database interfaced with the web application, and all communication with the database was automatically performed by the algorithm through PHP. The database was structured as follows:

1. db_group1 contained two tables: info, which contained the id number of the subjects (primary key, therefore unique), their ages, their genders and their initial stress levels; text, which contained the id number of the subjects (not unique, we had multiple lines per subject), the repetition number, their impression for the repetition of the task and, finally, the number of distractions;
2. db_group2 was identical to db_group1, but the text table was instead called task and did not contain the number of distractions;
3. db_group3 only contained the info table, as we were not interested in the results of the questionnaires for the purpose of this work.

The aim of the application was to generate a graphic interface, to perform the testing of our algorithm on psychotherapy and, in particular, on a self-induced mindfulness meditative state. We wanted to compare the effects on lowering stress of a visual-bilateral-stimulation protocol to the effects induced by a distraction task, such as filling in a questionnaire, and to another type of self-induced meditation, such as diaphragmatic breathing. Given the simplicity of our two control tasks, which required no complex coding, due to the absence of machine-learning intervention, we focused only on the implementation of the dot-following section of the interface. Our algorithm could be divided into three consecutive steps, with the block diagram for the application shown in Figure 2:

1. distance estimation of the subject from the screen;
2. calibration of the camera;

3. generation of an eye trajectory and extraction of the number of distractions.

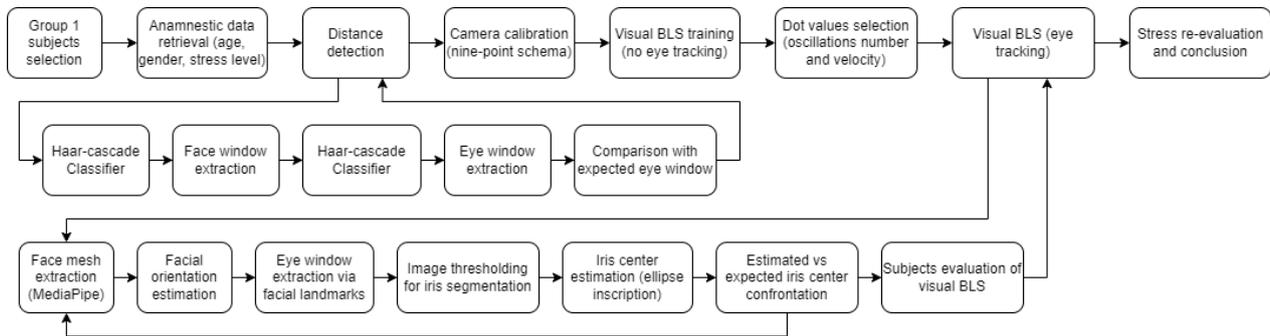


Figure 1. Block scheme for the workflow of the experiment for Group 1. Group 1 performs a visual bilateral stimulation actuated by means of the oscillation of a dot on the screen.

The first step was the distance estimation. This step fundamentally had to be performed before any operation with eye tracking, in order to ensure that the subject was at an optimal distance from the screen, so as to make the eye tracking more precise. Being too far would make pupil movements too small and unnoticeable for the algorithm, while being too close would force the subject to move their head to follow the movements of the dot on the screen. Our application, especially from the perspective of EMDR future use, is strictly required to have no head movements and for the visual stimulation to be followed only with the eyes. We ascertained that, with a window of dimension 1707×803 (measured on a Google Chrome window of a 17" PC), a distance of around 50 cm was sufficient to solve both requirements; therefore, we considered these values as our baseline.

We chose to use as our metric the dimension of the eye of the subject, which was extrapolated through eye detection. The estimated dimension was compared to the expected eye dimension, in correlation to the distance from the screen. To extract the eye dimension we chose to employ cascade classifiers, a class of machine-learning algorithms used in ensemble learning, mainly for face recognition. In particular, we chose to use Haar-cascade classifiers, first introduced in 2001 as a fast and efficient way to perform object detection [29]. This approach uses Haar-like features extracted from the training images by employing three different rectangle features, as shown in Figure 2. The regions make use of the natural lighting properties of a face, with the white regions indicating high luminosity and the black regions indicating low luminosity, to identify features in an image. The features are then confronted with the expected one (e.g., a face or eyes), to check if the feature corresponds to the one we want to extract.

We used two different pre-trained Haar-cascade classifiers from the OpenCV open library available on GitHub, one for the face detection and one for the eye detection. We first segmented the face window with the first classifier, and then used it as a baseline to extract the eye window with the second classifier. Usually, with this algorithm, we either detect both eyes but with slightly different dimensions (e.g., ± 5 pixels) or only one, in cases of peculiar lighting conditions. It is sufficient for our application that only one eye is visible and measurable, however, so we simply considered the first recognizable eye as the one we would use for our distance check. The Haar classifier generates eye windows with a width:height ratio of 1:1 (i.e., squares); therefore, we could use either dimension of the eye to verify the distance from the screen. Moreover, the bigger the screen was, the smaller the eye needed to be; therefore, their values were inversely proportional. We could easily determine the required height of the eye through a simple proportion:

$$window_width : 1707 = \frac{1}{eye_height} : \frac{1}{7} \quad (1)$$

with $1/7$ being the expected height of the eye's bounding box. This value was then compared to the height estimated by the Haar classifier, in order to determine the correctness

of the subject's distance from the screen, and only when the two values coincided did we confirm that the subject was at the optimal distance from the screen. Compared to the State-of-the-Art approaches, the proposed solution is much easier to implement, requires no training, due to using pre-trained models, and offers high precision with a low computation time. Moreover, it presents an innovative use of Haar-cascade classifiers not only applied to object detection, but also in the context of distance detection by employing the dimension of the generated bounding boxes.

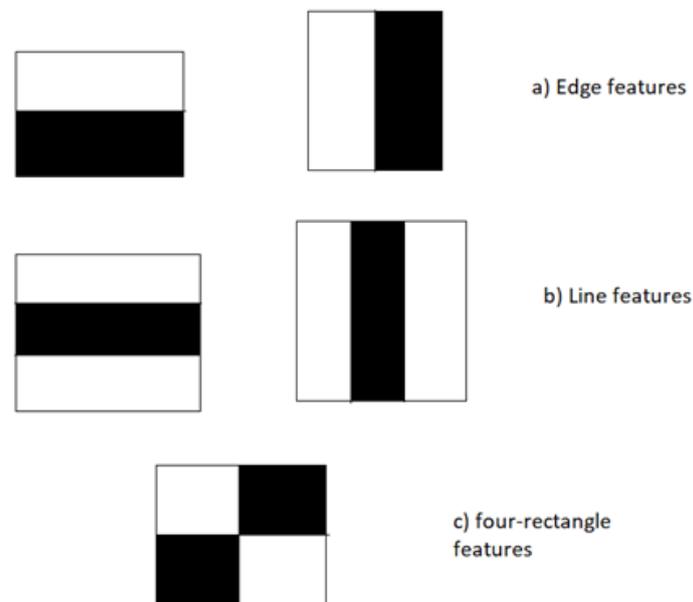


Figure 2. (a) Edge features are used mainly for eyebrows and nose segmentation. (b) Line features are used mainly for lips and pupil segmentation. (c) Four-rectangle features are used for diagonal segmentation, e.g., if the face is tilted.

Once the distance calibration had been completed, we needed to perform camera calibration, in order to improve the performance of the eye tracking. Camera calibration is a fundamental step in computer vision, a subclass of machine learning. While for this specific application we only tested horizontal movements, we also needed to check for distractions of the subjects along the vertical axis; therefore, a simple three-point calibration schema was not sufficient for our task. We instead used a nine-point calibration schema, which was more suited to our needs. This schema, in fact, calibrated the algorithm to recognize screen-wide eye movements, therefore allowing us to catch distractions even on the vertical and diagonal plane with higher accuracy. Moreover, as we were working on a big screen, a screen-wide calibration was necessary, to be able to catch the wide pupil movements that the subjects performed. The calibration worked by showing, for 5 s each, a still dot on the screen in the positions marked in the schema in Figure 3, preceded by a 2 s fixation cross, to identify the spawn point of the dot. The chosen dot color was a brownish yellow, while the background was black, to make the dot pop up more and to reduce the possibility of reflections in the subjects' irises, which could hinder the process. The dot was also much smaller compared to the one used during the dot-following exercise to reduce pupil wandering. The idea was to simulate the conditions that would be replicated during the experiment and also for the calibration phase, in order to calibrate the eye tracker under similar conditions. The eye tracker captured the pupil position of the subject, generated an estimate of the average fixation point for each calibration point, and then compared it to the overall average fixation point. The procedure for the pupil-position extraction was identical to the one used for the eye tracking, and will therefore be explained later on. The

result was that the camera was re-calibrated to make the predictions of the fixation point more precise by estimating the center of the eye-bounding box through the nine calibration points. While this implicated an increase in the accuracy of the eye tracker, it must be noted that the precision was not influenced, as it was only dependent on the hardware quality. With our algorithm, calibration and eye tracking were performed together, so at this stage we simply automatically recorded singular videos for the calibration.

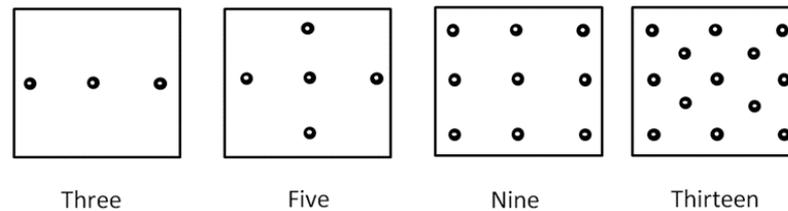


Figure 3. The main four schemas used for camera calibration. In our algorithm, we used the nine-point one.

The final step was the eye tracking and the consequent analysis of the extracted data. We worked on recordings that were captured in real-time during the execution of the visual BLS. The application fully used OpenCV to perform computer-vision operations on the extracted frames. In particular, the algorithm extracted the face mesh, which identified the location of the person in the frame with the main landmarks (eyes and lips), identified the irises of the subject for both eyes by generating both a mask and a bounding box and, finally, performed the estimation of the gaze trajectory for both eyes, generating a prediction for both eyes singularly and a joint trajectory.

We used MediaPipe, a machine-learning library with many pre-trained solutions for computer vision, to first generate a face mesh on the given frame. Two deep neural networks, a detector and a regressor, generated a mesh of 468 3D points identifying each face in the scene. The mesh was then used to determine the orientation of the face and if the user was looking at the screen. We simply checked that the amplitude of each angle $\{XZ, YZ, XY\}$ was included in the window $(-29, +29)$ degrees. We then extracted the eye-bounding boxes by using the eye landmarks. Finally, we performed image thresholding, generally used for image segmentation and separation of subject and background [30], and the connected component analysis, to identify the biggest connected component and to segment the iris. In particular, we used a low-pass filter to filter out the lighter areas of the eye windows. Depending on the threshold, we had different behaviors, as shown in Figure 4. If the threshold was too low, many of the lighter areas were not considered as part of a connected component and, consequently, the computation of the biggest connected component returned a region smaller than the real iris. On the contrary, if the threshold was too high, the computed biggest connected component included areas of the eye outwith the iris, especially if the eye was not properly illuminated. With a median value and proper illumination, however, we had near-perfect segmentation of the irises for both eyes, which, in turn, lead to smoother movement detection and gaze-trajectory reconstruction. The segmentation of the iris through inscription in an ellipse also allowed us to identify the center of the iris, which was used to reconstruct the eye trajectory. The resulting iris center for each frame was compared to the expected one, and if the estimated iris center did not fall within the vicinity of the expected position, we identified a distraction. We calculated the vicinity by considering the dimension of the dot on the screen and an allowed margin of error due to the speed of the moving dot. Note that we only counted a mismatched position as a distraction if the position of the iris center was incorrect for more than 15 consecutive frames, i.e., for longer than 0.5 s. This corresponded, on average, to a sequence of three fixations plus one cascade. The reason was that, as we were working with a moving target, we needed to consider the subjects' reaction times, in registering that the position of the dot had changed between frames.

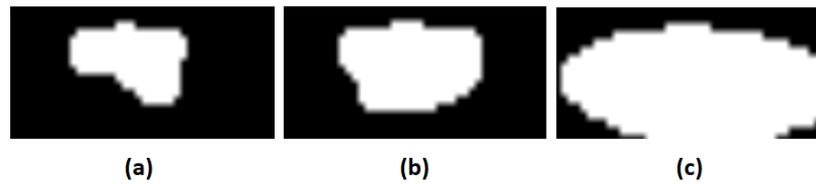


Figure 4. (a) Iris segmentation with low threshold. (b) Iris segmentation with medium threshold. (c) Iris segmentation with high threshold.

While numerous solutions for gaze tracking exist in the literature, including specific applications for eye-controlled PCs, the presented approach introduces two key innovations in the context of attention analysis. The first novelty lies in the method employed to compare the expected gaze direction to the estimated one. Unlike the majority of approaches in the literature, which directly compare the point of regard on the screen, our method involves a direct comparison between gaze trajectories. This introduces two crucial advantages. The primary one is the elimination of the need to perform a change of reference system from eye coordinates to screen coordinates, thereby reducing the likelihood of estimation errors. Consequently, the second advantage is a reduction in computation time, enabling a faster assessment of distractions. Given its potential future application in Eye-Movement Desensitization and Reprocessing (EMDR), it is essential that our application can identify distractions as quickly as possible.

On the other hand, the second novelty pertains to the application of eye tracking in measuring attention on a moving target. Few works in the literature address this aspect, with one of the most recent employing Hidden Markov Models (HMMs) as a solution to the problem [31]. In comparison to this solution, which—although well-established in the literature—is computationally intensive, our infrastructure is much lighter, easily reproducible, and swift. A single-frame analysis requires only a few milliseconds, enabling near-instantaneous tracking of the iris center position and, consequently, its correspondence to the position of the moving dot on the screen.

3. Results

3.1. Preparation

The subjects for this study were divided into three groups. The data collected with our infrastructure were treated in accordance with ethical concerns, with the video recordings for the patients being visioned only by the examiners. Moreover, the anagraphic data of our subjects (age and gender) stored in our SQL database were associated with a randomly generated code, rendering it impossible to reconstruct the identity and data of specific test subjects, in accordance with privacy. Due to the execution of the application completely on the user side, we also eliminated any possibility of data leakage in transferring data. In any case, all the people involved were of age and of varied demographics, as shown in Table 1. Concerning the exclusive use of a virtual interface for psychotherapy, which could raise doubts, such as autonomy issues or a lessened competence of the algorithm compared to human expertise, all the experiments were performed in the presence of a therapist and an operator. These professionals guided the subjects throughout the experiment and intervened as needed, addressing any concerns or issues that may have arisen. This approach ensured a supportive and supervised environment, mitigating concerns related to the exclusive reliance on the virtual interface. They also provided validation for the efficacy of the task by manually checking the eye movements of the involved subjects in performing the visual BLS. Finally, the subjects were free to interrupt the experiment at any time, in case of discomfort.

Table 1. Demographics and gender of the subjects involved in the study.

Group	Average Age	Males	Females	Others
Group 1	42	1	3	0
Group 2	53.75	2	2	0
Group 3	31.5	1	2	1 ¹

¹ Genderfluid.

The first group performed the complete mindfulness protocol, consisting of a dot-following-plus-respiration task. During the whole BLS, the previously calibrated eye tracker monitored the movements of each subject's eyes, to make sure the subject's gaze did not stray from the movement of the dot. After each visual BLS, they were asked to perform diaphragmatic breathing, to return to a state of calm, and they had to state their opinions on the experience before repeating the set. In total, each subject performed 10 stimulations of this type, for a total time of around 5 min of exercise.

The second group performed a simplified task that removed the visual BLS and was, therefore, comprised only of the diaphragmatic breathing part, an exercise which was, in itself, a meditation and relaxation technique. The subjects performed five repetitions of diaphragmatic breathing of around 6 s each, followed by an evaluation of the sensations felt. In this case too, the subjects performed 10 sets of stimulations of this type for a total of around 5 min of exercise. This group worked as a middle ground between a control group and a visual-task group.

Finally, the third group was our control group, and they did not perform any kind of visual or respiration task in order to provide an estimation of the lowering of stress, in case no visual bilateral stimulation was administered to the subjects. Instead, they were subjected to questionnaires with no relation whatsoever to the experiment, with a duration similar to that of a standard session for another group (around 5 min). They served as a baseline for our study, and they provided the data for patients who do not receive any kind of treatment. For our study, we chose to make them fill in a questionnaire of 25 items on general knowledge of English.

All the subjects were asked to give an estimate of their stress level at the beginning of the exercise, on a scale from 1 to 5, and to re-evaluate the stress level at the end of the exercise, also giving an overall evaluation of their experience with the experiment. The tests were all realized in-presence and with the same apparatus, to have consistency in the results, with the database for the subjects' data stored locally on the PC. The textual responses given by the subjects during the exercise were analyzed, to verify the quality of the relaxation at the increase of the number of repetitions of the exercise. Finally, the obtained graphs were manually compared to the recording of the eye movements, to validate them by checking the match between the estimated iris oscillation and the physical ones recorded in the videos.

In order to not have any interference in the eye and iris detection, all the glasses-wearing subjects for Group 1 were asked to remove them. They executed the exercise in a room with natural lights of medium intensity, to simplify eye recognition and iris segmentation for the algorithm. All the subjects performed the exercise while sitting on a chair and with the PC in front of them at eye level. The subjects rested their backs against the chairs, in order to reduce the oscillations of their bodies and to increase the quality of the eye tracking. Regardless of the group they belonged to, they were asked to sit as centered as possible in front of the screen and in a relaxed and open position, to favor relaxation during the test. In Figure 5, we show the experimental setup for our study. Despite the more complex experimental setup used for this study, it must be noted that our framework can be run on any PC and does not require any specific apparatus nor a specific screen dimension to function correctly.

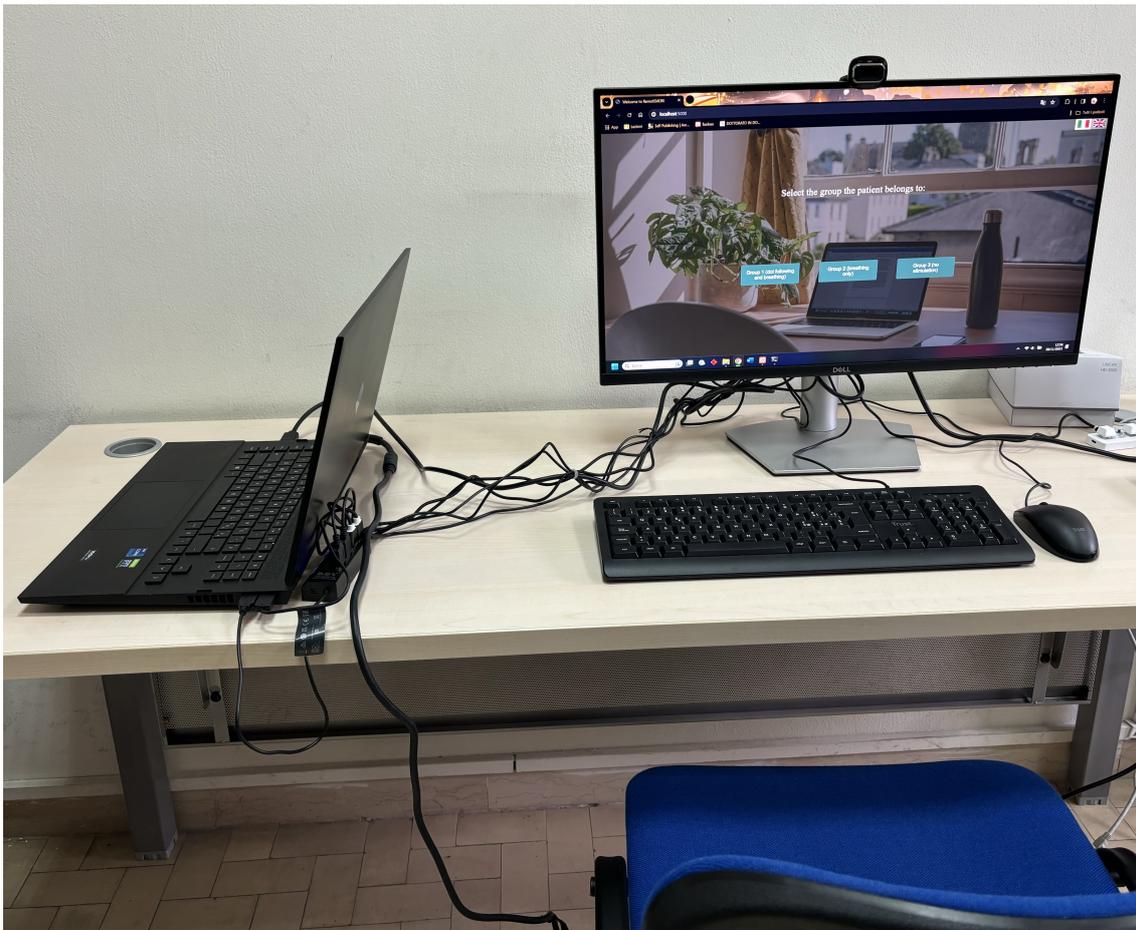


Figure 5. Experimental setup for our study. The subjects had at their disposal an adjustable (both in height and in rotation) 27" monitor with an external webcam with 30 fps, a mouse and a keyboard to interact with the web application. The chair they performed the experiment on allowed the subject's back to fully rest against it, in order to minimize accidental oscillations of the subject during the execution of the dot-following. As shown in the picture, the room had a uniform natural-colored lighting that did not negatively impact on the darkness of the image. The light was generated by artificial lighting.

Depending on the group, the subjects were then debriefed about the task they were to perform, before the start of the experiment. To not interfere with their perception of the experienced relaxation, however, they were not informed about the group they belonged to, the existence of more than one test group, nor the hypothesized difference in efficacy between the exercises of the groups. The authors of this work guided the subjects during their interactions with the interface, manually intervening to operate the algorithm when needed. In any case, the screen was never moved between runs of the same experiment, in order to not alter the distance of the subject from the screen. Moreover, the subjects were asked to move as little as possible during the whole experiment, to avoid changing their position compared to the calibrated data, with particular attention to not moving during the recordings of their ocular movements.

No training was required for the machine-learning algorithms employed in our experiment, as the chosen neural networks (Haar-cascade classifier and MediaPipe's mesh extractor) were all pre-trained. The only validation present was that of the therapist, who—thanks to his expertise—confirmed the correct execution of the dot-following of the subjects, and helped us to design suitable control tasks to which to compare the results of the algorithm. Therefore, we directly skipped to the testing phase. However, before applying the

interface on the test subjects, we performed some trials, which are explained in more detail in the following subsection.

3.2. Data Collection

The first trial was a simple test, where we performed two willing movements, one in which we closely followed the dot movement and one in which we purposely injected distraction. Both movements were performed with velocity 5 and 10 oscillations of the dot. The expected iris-center trajectory for both eyes is shown in Figure 6. For all the graphs shown in this manuscript, on the horizontal axis we have the expired time and on the vertical axis the x coordinate of the iris center, where 0 corresponds to the center of the eye, +1 to the left side (the right side in the registered mirrored frame) and -1 to the right side (the left side in the registered mirrored frame).

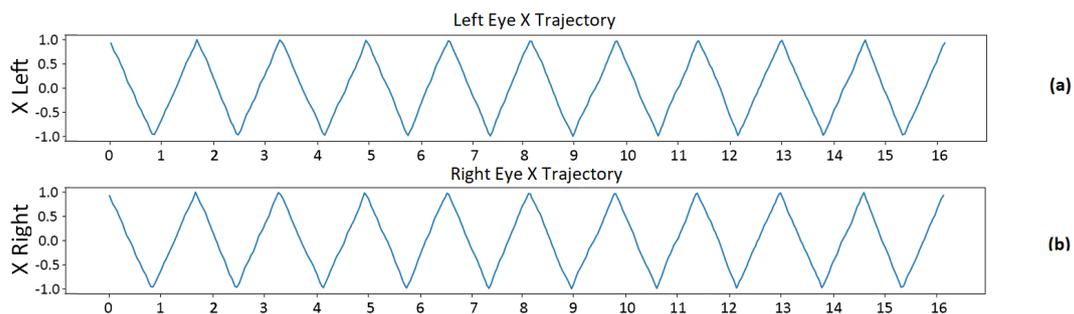


Figure 6. (a) Expected iris-center trajectory for left eye. (b) Expected iris-center trajectory for right eye. In both cases, the expected trajectory was generated by calculating, for each frame, the expected displacement of the iris center's x coordinate according to the current position of the dot.

As shown in Figure 7, we verified the trajectory for the first test. As we can see, the trajectory closely followed the expected one, with no distractions signaled by our algorithm. This was possible, despite the trajectory slightly diverging from the expected one, as we permitted some position errors due to the dimension of the dot and the necessity to follow a moving object. In Figure 8, on the other hand, the trajectory diverged at three points from the expected one. We first see a fixation on the left side, followed by a fixation towards the center-right and a final fixation towards the left. The algorithm correctly identified three distractions, despite the very noisy measurement for the right eye caused by shadows over the eye. We therefore concluded that the algorithm was correctly able to identify distractions during the dot-following task and that we could safely proceed to our experimental phase with the selected subjects.

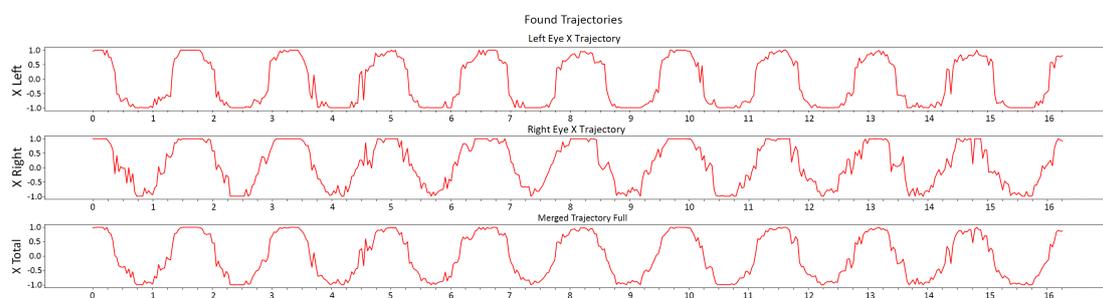


Figure 7. Evaluated trajectories for the left-eye, right-eye and joint trajectories in a dot-following test. The presence of some small peaks (e.g., at 4.50 s) in all the graphs was generally due to imprecise iris segmentation, often caused by environmental lighting, especially in the right eye, which was slightly more shadowed during the execution. The joint trajectory showed a correspondence of the peaks to those of the expected trajectory, with the divergences falling in the allowed error window.

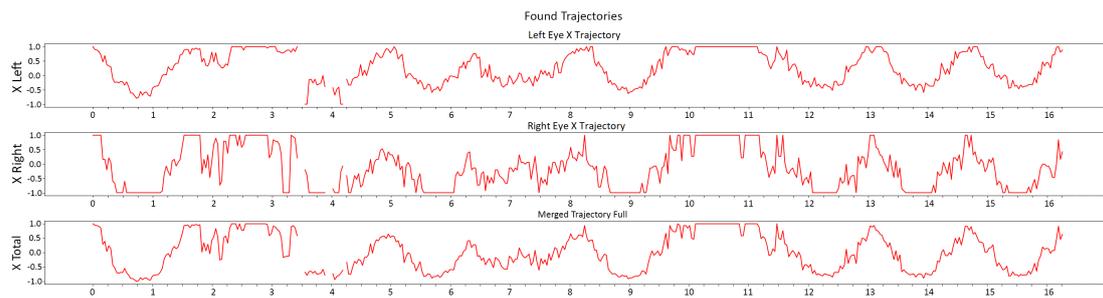


Figure 8. Evaluated trajectory for the left-eye, right-eye and joint trajectories on a distraction test. The discontinuities in the trajectories were due to skipped frames, attributable to blinking, to shifting the gaze from one direction to another and/or to inability to recognize the eyes or irises, due to shadows. Peaks of the x in 1, e.g., from around 1.50 s to 3.50 s, correspond to fixations to the left side of the screen (mostly with values saturating in 1), while peaks of the x around -0.3 , from around 6.25 s to 8.25 s, indicate fixations to center-right of the screen. The downward peaks in the right-eye trajectory were due to heavy shadows over the right eye, causing the perceived iris center to shift. The initial and final sections of the trajectory follow quite closely the expected one, with a margin of error, indicating correct dot-following (correlated to no measurement of a distraction).

In Tables 2–4, we present the results for our study group. It is important to note that this study was not only focused on verifying the efficacy of a virtual mindfulness protocol compared to other approaches, but also on testing the efficacy of our eye-tracking algorithm for tracking fast eye movements. In particular, we also wanted to verify if asking glasses-wearing subjects to remove their glasses reduced or interfered with their involvement in the exercise. For this reason, the participants selected for Group 1 were mainly glasses-wearing subjects. Subject 702011934 suffered from presbyopia, subject 7436290 from astigmatism and subject 33124157 from heavy myopia. While subject 702011934 showed no difficulty in following fast movements, subject 7436290 had difficulties with fast movements and asked to slow down the dot. With this small adjustment, however, she showed no difficulties in performing the task. Subject 33124157, on the other hand, showed an incredibly high number of distractions due to being unable to properly see the dot movement even at low speed.

In regard to Group 2 and 3, the selected participants had no particular requirements, other than being as balanced as possible in their gender and age distribution, to reflect as much as possible a wider sample of subjects. It is of note that subject 721212524, however, was overweight and had some breathing difficulties, which required taking small pauses between trials, to let the subject readjust their breathing. Moreover, the subjects for Group 3 were all proficient in English (C1+ level), despite English not being their mother tongue and, therefore, had no particular difficulties with the questionnaire.

Table 2. Group 1 Results.

Subject ID (Initial Stress Level -> Final Stress Level), Age, Gender	Trial	Evaluation	Distractions
773886486 (3 -> 1), 25, F	1	I emptied my mind.	11
	2	I felt calmer.	8
	3	I'm feeling more relaxed.	0
	4	I feel good.	4
	5	I feel calm.	0
	6	I'm okay.	0
	7	I'm feeling relaxed.	1
	8	I notice no change.	0
	9	I'm good.	0
	10	I'm feeling much better.	0

Table 2. Cont.

Subject ID (Initial Stress Level -> Final Stress Level), Age, Gender	Trial	Evaluation	Distractions
702011934 (2 -> 1), 61, M	1	I feel calm.	0
	2	I feel calm.	0
	3	Always calm.	1
	4	Always calm.	0
	5	Very relaxed.	2
	6	No change.	0
	7	Always the same.	2
	8	Very relaxed.	9
	9	Everything okay.	1
	10	Very calm.	2
7436290 (3 -> 1), 57, F	1	I'm not thinking about anything.	0
	2	I gritted my teeth.	0
	3	Very relaxed.	0
	4	Thought about two positive things, relaxed.	0
	5	Relaxed, harder to follow the dot.	0
	6	More relaxing than before.	0
	7	I thought about nothing.	0
	8	I'm dissociating a bit.	0
	9	I feel relaxed.	0
	10	Calm and relaxed.	0
33124157 (2 -> 1), 23, F	1	Relaxed, but a bit hard to follow.	0
	2	Better to follow.	7
	3	Relaxed, easy to follow now.	13
	4	More relaxed than before.	12
	5	Much better.	9
	6	As before.	18
	7	Very calm.	18
	8	I feel no tension at all.	14
	9	Relaxed.	11
	10	I feel okay.	17

Table 3. Group 2 Results.

Subject ID (Initial Stress Level -> Final Stress Level), Age, Gender	Trial	Evaluation
709896582 (4 -> 3), 25, M	1	I still feel nervous.
	2	I feel slightly more relaxed.
	3	I notice no change.
	4	I feel a slight chest oppression.
	5	I'm slightly calmer.
	6	I'm okay.
	7	I'm slightly unnerved.
	8	I'm calm.
	9	I'm thinking about the things I have to do.
	10	Same as before.

Table 3. Cont.

Subject ID (Initial Stress Level -> Final Stress Level), Age, Gender	Trial	Evaluation
721212524 (2 -> 2), 89, F	1	Everything good.
	2	I feel calm.
	3	It's a bit hard to breathe at the end of the exercise.
	4	Better than before.
	5	All good.
	6	Nothing in particular.
	7	A bit distracted.
	8	Slightly harder to breathe.
	9	Better after pausing, relaxed.
	10	Relaxed but slightly bored.
134487392 (3 -> 1), 45, F	1	Nothing in particular.
	2	Nothing to notice.
	3	I feel okay.
	4	I feel slightly calmer.
	5	Like before.
	6	I feel lighter in my chest.
	7	I'm calm.
	8	I'm feeling a bit lightheaded.
	9	Better than before, I'm less tense.
	10	I'm calm.
501047471 (3 -> 2), 56, M	1	Everything ok.
	2	Everything ok.
	3	More relaxed.
	4	All good.
	5	Relaxed.
	6	Same as before.
	7	Very relaxed.
	8	As before.
	9	Everything good.
	10	Everything is fine.

Table 4. Group 3 Results.

Subject ID	Initial Stress Level	Final Stress Level	Gender	Age
3642332	3	3	O ¹	25
34567683	4	4	F	53
248738499	5	5	F	24
723457394	3	3	M	24

¹ Genderfluid.

4. Discussion

The results obtained for the first experimental group show that there was a significant improvement in relaxation levels after performing the dot-following task, with the subjects being able to mostly focus on the task, even in cases where they were glasses-wearing people being asked to remove the glasses for the exercise. When we noticed outliers in the distractions, they could mostly be explained:

- subject 773886486's first high distraction numbers were due to a too-low threshold value for the analysis of the videos and iris segmentation, generally set to 70 in our lighting conditions, while in this case they were tested with lower thresholds and, therefore, constituted an outlier;

- subject 702011934's only high distraction number at trial 8 did not correspond to actual distraction in dot-following when compared to the video, which showed the subject to be properly following the dot, and the extracted trajectory shown in Figure 9;
- subject 33124157's generalized high distraction count was due to difficulties in seeing the movement and responding to it smoothly, due to glasses removal, with the trajectories showing clear signs of slowed-down eye movement, as exemplified in Figure 10, but a decent adherence to the task.

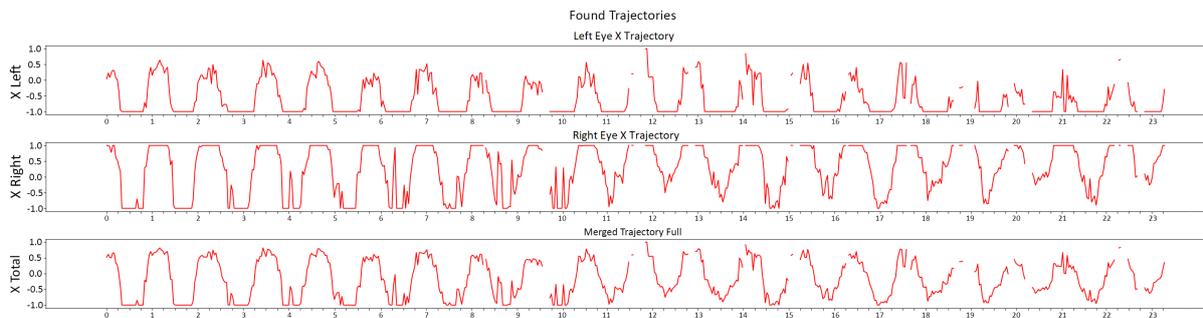


Figure 9. Evaluated trajectories for the left-eye, right-eye and joint trajectories for subject 702011934's trial 8. We can see that the trajectories mostly followed the profile of the expected one, coinciding with an extended version of the one shown in Figure 4. The measured distractions were probably a result of incorrect lighting or abnormal shadows, as the subject had slightly sunken eyes. Several saturations were present (both positive, e.g., from 0.75 s to 1.50 s in the right eye, and negative, e.g., 0.25 s to 0.75 s in the left eye), which may have constituted a measured distraction. We also had some skipped frames from half of the trajectory up to the end, which may also have contributed to the outlier. Neither problem was present in the other measurements of the same subject.

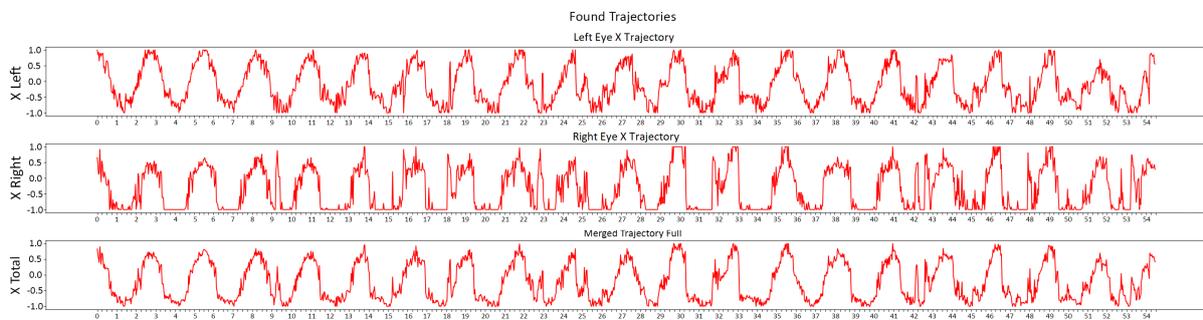


Figure 10. Evaluated trajectories for the left-eye, right-eye and joint trajectories for subject 33124157's trial 3. The trajectories were often discontinuous, due to jumps in the eye gaze (e.g., at around 33 s), as confirmed by the recorded video, and presented a high number of peaks, due to improper lighting (e.g., from 9.00 s to 9.50 s). However, when analyzing the joint trajectory, we can see that the profile was mostly faithful to the expected one, with the main source of errors derived from the middle-end part of the exercise, when the subject's eyes were more strained.

In most cases, however, the subjects were able to stay focused on the dot movement, despite it being performed on a screen instead of manually. The analyzed videos confirm the results obtained with the algorithm, and are also useful for validating the efficacy of the exercise. In fact, even in cases in which some small distractions were noticed, we can verify that the subjects were focusing on the movement, with subject 7436290 even entering a slightly dissociative state which amplified her focus on the dot. In her case, the visual BLS was nearly perfect, as exemplified in Figure 11. This confirms that their positive responses were attributable to the visual BLS performed on screen and not to conditioning due to knowing the aim of the experiment, a fact also enforced by the diminished efficacy of the other tested methods. We can, therefore, confirm our hypothesis that a virtual interface

shows the same results as a physical stimulation performed by the therapist by hand. With regard to the algorithm, faster movements were harder to track than slower ones, and were often those where we noticed the most errors in the estimation of position. This was partly due to the patients having slightly delayed responses to the movement, due to its speed, and also partly due to being asked to remove their glasses and, thus, having slightly more difficulty in following the dot.

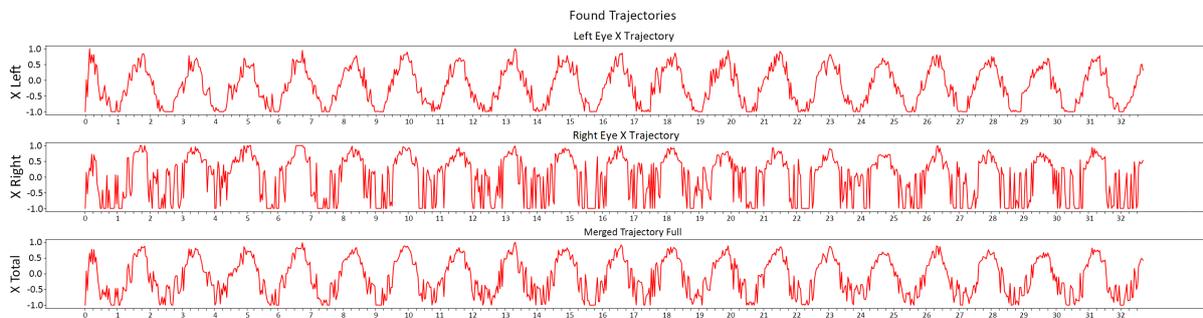


Figure 11. Evaluated trajectories for the left-eye, right-eye and joint trajectories for subject 7436290's trial 4. The trajectory for the left eye in particular and for the joint trajectory were nearly coincident with the expected one, considering an allowed window of error. The trajectory for the right eye was much noisier and presented several isolated peaks (which were filtered out by the algorithm), due to shadows in the image caused by the subject's hair darkening her right eye.

By comparison, diaphragmatic breathing did show some, albeit smaller and mostly physical, improvement in the relaxation of the subjects. However, sequences that were too long caused, instead, boredom, distraction from the task and, often, slight breathing fatigue. It is, therefore, advised to not have sessions which require a long sequence of diaphragmatic breathing for future studies, as they frequently induce hyperventilation. As confirmed by subject 721212524, sequences that are too long could result in breathing problems for people with breathing difficulties, asthma or suffering from panic attacks. Moreover, a virtual BLS performs better and is, therefore, preferable. Finally, for the control group all the subjects showed no improvement in stress levels and sometimes even noticed slight temporary worsening of stress, due to the questions being engaging and requiring thinking. This shows that simple distraction tasks, at least in the form of questionnaires, have no positive effect on the stress levels of the subjects.

A statistical analysis was conducted on the obtained data, due to the low number of subjects, in order to verify the statistical significance of the results. First of all, we evaluated the average and standard deviation of the final stress level and the difference between the final and initial stress levels for each group, as shown in Table 5. The table confirms the trend of Group 1 presenting on average the lower final stress level, with no measured deviation, while its average variation was the highest, with the lowest part of the deviation being, however, on a par with and even slightly lower than that of Group 2. In Table 6, on the other hand, we studied the Pearson correlation between the significant values of the study (method used, gender, age, initial stress level, final stress level and variation of the stress level). The correlation values analyzed confirmed that the strongest correlation was between the used method and the effect on the stress levels of the subjects, particularly concerning the final stress level (with strong positive correlation) and the variation of the stress level (with strong negative correlation). On the other hand, we did not verify any correlation between the method used, the results and the anagraphic data of the subjects; therefore, their results were not dependent on one another.

Table 5. Average and standard deviation (STD) for each group on the final stress level and variation of the stress level.

Group	Average Final Stress Level	STD Final Stress Level	Average Variation Stress Level	STD Variation Stress Level
Group 1	1	0	1.5	0.57735
Group 2	2	0.8165	1	0.8165
Group 3	3.75	0.95743	0	0

Table 6. Pearson correlation matrix for the values considered in the study. As the matrix was symmetrical, only the lower part of the matrix was reported. A high negative correlation implies that the values were strongly negative correlated—for example, in the correlation between the method and the variation of the stress level. This implies an increase in the method number (from 1 to 3, corresponding to the group number), corresponding to a lower variation of the stress level. By contrast, a high positive correlation implies the values were strongly positively correlated, as seen in the correlation between the method and the final stress level. A lower method number corresponded to a lower final stress level, confirming our results, in that the used method not only impacted on the effect on stress, but also that dot-following was the best approach.

	Method	Gender	Age	Variation Stress Level	Initial Stress Level	Final Stress Level
Method	1	-	-	-	-	-
Gender	0.31623	1	-	-	-	-
Age	-0.20412	-0.2582	1	-	-	-
Variation Stress Level	-0.76613	-0.32303	0.20851	1	-	-
Initial Stress Level	0.592	-0.074883	-0.29002	-0.22174	1	-
Final Stress Level	0.86424	0.14907	-0.32075	-0.76244	0.80001	1

We also performed a statistical significance test (chi-square test), to verify the obtained results between our algorithm and the control groups. We performed the test for two values: the final stress level (which had different values, depending on the analyzed case) and the variation of the stress level (which varied between 0 and 2; we did not consider negative variations and variations higher than 3, as they had never been measured on the current population sample). The chi-square test was repeated for three different pairs of groups, with the resulting chi-square being compared to one of the many tables for the correspondence between the chi-square distribution and the percentage points. In all the mentioned cases, we had 2 degrees of freedom for the variation of the stress level (we worked with 3×2 tables), while for the final stress level the number of degrees of freedom was variable. We considered a significance level of 0.05. The obtained results were as follows:

1. Group 1 (dot-following) and Group 2 (diaphragmatic breathing): For the final stress level, we obtained a chi-square of 4.8, with $p = 0.090718$ under 2 degrees of freedom (the final stress level only varied between 1 and 3 in the analyzed population sample). In regard to the variation of the stress level, we obtained a chi-square of 1.333, corresponding to $p = 0.513417$. We can conclude that we have no evidence against the null hypothesis for the variation of the stress level, which can be justified by the heavy dependency on the initial stress level, and weak evidence for the final stress level, which is in line with our hypothesis that diaphragmatic breathing did show advantages, albeit in a less effective way than dot-following.
2. Group 1 (dot-following) and Group 3 (questionnaire-filling): For the final stress level, we obtained a chi-square value of 8, with a percentage point of 0.046012 under 3 degrees of freedom (we see variations in 1 and between 3 and 5), while for the

variation of the stress level we obtained a chi-square value of 8 with $p = 0.018316$. In both cases, we have strong evidence against the null hypothesis, especially in regard to the lowering of the stress level, which confirms both our most important hypotheses (questionnaire-filling had no effect on lowering stress and was much less efficient than dot-following).

3. Group 1 (dot-following) and Group 2 (diaphragmatic breathing) + Group 3 (questionnaire-filling): The results of the chi-square test for the final stress level showed chi-square = 8.4 and the corresponding percentage point being 0.077977. For the variation of the stress level, the result was 4.5 for the chi-square test, with a corresponding percentage point of 0.105399. In both cases, we have weak evidence against the null hypothesis, justified by the lower distinction in the effects between dot-following and diaphragmatic breathing. This shows that merging the two control groups had no significant effect on proving evidence, an obvious conclusion, as the two approaches were vastly different.

Finally, we also performed a verification on the full 5×3 table, for a more precise evaluation of the correlation between all the data. For the final value, with 8 degrees of freedom (the results varied between 1 and 5) we obtained a chi-square of 12.1333, with a corresponding p -value of 0.145368. The results show extremely weak evidence against the null hypothesis, in line with the results obtained for the last case mentioned above. In regard to the variation of the stress level, we obtained a chi-square value of 8.1333, which corresponded to a p -value of 0.086826. This shows weak evidence against the null hypothesis, which once again is in line with the previously obtained results. It must be noted that by considering a lower significance level (e.g., 0.1), however, many of the weak significances measured would instead be verified as strong significances, therefore proving more strongly the statistical significance of this study. Considering the very low number of subjects included in the study, we can safely assume that a bigger population would indeed be linked to a significant increase in the statistical significance for all the cases mentioned in this significance study.

The overall result of the performed tests, even on a small sample of subjects, proves our hypothesis that, even in a fully virtual setting, visual bilateral stimulation is effectively able to greatly reduce the patient's stress level and induce relaxation. Moreover, it is also more effective than techniques such as distraction, which was the worst performing one—sometimes even injecting more stress into the already-stressed subjects—or diaphragmatic breathing. Finally, the efficacy of the eye-tracking protocol in the correct environmental condition clearly showed that implementing virtual applications for BLS is not only feasible but also an excellent solution to aid the therapist when administering such stimulations in settings such as the mindfulness protocol or EMDR therapy.

Our current results do show some high requirements, especially in regard to environmental conditions and, in particular, lighting. In case of execution of the protocol in a user-operated setting, i.e., inside their home, a further increase of the capabilities of the system, to reduce environmental noise, might be necessary, so as to ensure that the algorithm can properly perform even when the user cannot guarantee optimal conditions. Our tests have, however, shown promising results in this direction, with good results even when lighting conditions are sub-optimal, as exemplified by the noisier graphs shown in this manuscript. In regard to other requirements, the presented algorithm already tackles several of them. The distance requirement is already automatically dealt with through distance estimation, which waits until the subject is at the correct distance from the screen before proceeding with the exercise. The machine-learning infrastructure is also light enough to allow for execution on most hardware, even on obsolete hardware, and works perfectly with basic cameras, such as PC-integrated ones, which nowadays all have 30 fps. Finally, in regard to head movements, we have no current options to ensure that the head movements of the subjects are nullified, other than relying on their ability to stay still. However, considering our infrastructure is only assistive for the therapist, we can rely on their expertise to catch head movements and signal them to the subjects. Alternatively, we

could in the future implement an analysis of head pitch, roll and yaw rotations, to ensure the subject is not moving during the exercise. Given our results, however, we can safely confirm that most fluctuations are not significant in the efficacy of the patient's state of mind in the virtual protocol we have implemented.

The future direction of the work will be focused on improving the quality of the eye-tracking algorithm by testing an estimation of the gaze trajectory and the distraction of the subject fully via neural-networks analysis of the frames. If possible, we also want to allow patients to not remove their glasses, because, as proved, some patients have noticeable difficulties in seeing without them. This could easily interfere with the efficacy of the treatment, in cases of EMDR application, where we require an extremely precise execution of the dot-following. We also want to verify if a natural-language-processing analysis of the textual responses can offer us some insight into the efficacy of the treatment, possibly allowing auto-tuning of the parameters for the visual BLS and predicting pre-emptively the expected re-evaluation of the subjects. Finally, we plan to apply this method to a tailored protocol, to verify the efficacy of a virtual approach not only for self-meditation but also for reducing trauma in preparation for an application to the EMDR protocol. In this new experiment, we will manually induce traumatic memories, to treat with desensitization in the subjects, with the possibility of reproducing the experiment in a full EMDR therapy session, in case of success.

5. Conclusions

This work is a first step in a new direction for the application of machine learning to psychology and, in particular, to psychotherapy, a field in which machine learning still has a very minimal impact, aside from common pathologies such as ASD, depression or Alzheimer's disease, and which has shown interesting results even in a small sample of 12 subjects. The research unfolds several advantages in the integration of virtual interfaces within psychotherapy. Firstly, it alleviates the potential discomfort associated with the therapist having to track eye movements using physical gestures, removing any awkwardness from the therapeutic process. Secondly, it introduces a measurable form of stimulation, enhancing the precision of therapeutic interventions. Additionally, the virtual interface is versatile, having applications within and outwith the EMDR protocol. In cases where traditional forms of stimulation may be contra-indicated, such as patients with trauma linked to physical violence, the virtual platform offers a viable alternative. This is crucial, as it minimizes the risk of reactivating traumatic experiences and addresses potential barriers to therapy, ultimately contributing to a more inclusive and effective therapeutic approach. Some of the barriers addressed by a virtual psychotherapeutic approach are:

1. Therapy can become accessible to a broader population, with the possibility of reaching patients with traveling difficulties, such as the elderly, individuals with mobility challenges, agoraphobics, hikikomori or those living in remote or isolated villages;
2. The use of eye-tracking technology provides additional control support beyond eye movement. The therapist could take advantage of our instrumental support to accurately check, within a hyper-human rate, whether the patient is correctly performing the prescribed movements;
3. The patient has the opportunity to attend the session while remaining in a familiar and more comfortable environment, compared to the therapist's study;
4. Virtual therapy eliminates the need for physical infrastructure, reducing the costs associated with maintaining a therapy space, and may result in more affordable treatment options for both patients and therapists;
5. Virtual therapy contributes to environmental sustainability by reducing the need for travel, thereby lowering the carbon emissions associated with commuting to in-person sessions. This eco-friendly approach aligns with the growing importance of promoting practices that minimize our carbon footprint and support a greener, healthier planet;
6. Virtual therapy proves advantageous in times of pandemics, ensuring the continuity of psychotherapeutic interventions when in-person sessions may be challenging or

restricted. This adaptability becomes crucial for maintaining mental-health support during global health crises;

7. The virtual platform enables the provision of psychotherapeutic interventions from a distance, allowing therapists to offer support remotely. This is particularly valuable in situations where face-to-face interactions are not feasible, providing a flexible and accessible mode of intervention. Some examples of a remote intervention protocol are natural disasters, earthquakes, terrorist attacks and war.

Regarding this last point, several studies on the efficacy of EMDR for early intervention have been presented by Shapiro, Lehnung, Solomon and Yurtsever [32–36]. These works, under the influence of Roger Solomon's works on critical incidents, formalized a specific Recent Traumatic Event (RTE) protocol by adapting the standard EMDR protocol for recent traumatic events. The RTE protocol was first introduced during the Lebanon war, to assist practitioners unfamiliar with it, in effectively addressing recent traumatic events. Finally, the need to develop a broader model and protocol for the treatment of recent traumatic events led to the development of the Recent Traumatic Episode protocol (R-TEP). This approach stems from the observation of limitations in existing protocols and aims to address the gaps in dealing with complex and multiple events, while proposing an expansion of interventions from a narrow focus on intrusive sensory images to a broader perspective encompassing a series of episodes, underscoring the need for an approach that embraces both aspects.

Our support technology is fully compatible and ready to be used for such applications, by providing an extraordinary asset to the psychotherapist, in order to offer support therapies and treatments in the immediacy of traumatic events, despite the geographical reachability of the subject and the logistic limitations that can result from disrupting or catastrophic scenarios. Our technology can therefore support the psychotherapist, in order to also deliver treatments in cases of disasters, pandemics and quarantine measures (e.g., COVID-19), natural calamities or other disruptive events that could inhibit the ability of the therapist to physically meet the patients, assuring the maximum possible level of assistance and health-care service despite the circumstances.

As previously mentioned, there is still much work to do, to consolidate this approach and to verify its applicability to EMDR, but the results we have obtained with this preliminary study confirm the validity of the experiment and that it should greatly improve the quality and scope of applicability of current approaches. In case of an application for the EMDR protocol, the interface will be enhanced, to follow precisely the eight steps. In particular: phase 1 to phase 3 will be carried out by means of a video call with the psychotherapist, with no intervention of any artificial intelligence algorithm, using the system as a simple telecommunication interface; phase 4, 5 and 6 will be carried out by the psychotherapist by taking full advantage of the support system in order to track the eye movements of the subject and assess their correctness; finally, phase 7 and 8, will only use the system as a telecommunication interface.

This project also saw many improvements in the pre-processing of images, in regard to the eye-tracking task: for example, taking into account people who wear glasses, and sub-optimal lighting conditions, which now are instead required to be controlled by experimenters and would introduce limitations for real-life uses. Moreover, accurate pre-processing through operations such as rendition to normalized gray-scale image (we pass from three channels to one channel) and downsizing to a standardized dimension [37], with also the possibility of applying some filters to simplify the identification of the pupil (e.g., contour extraction or gradient-magnitude map), could easily also improve the velocity of the analysis of the neural network, rendering an already considerably simplified process, due to the reduction of the frame rate, even faster and, therefore, closer to being real-time. We could also make the image even smaller and easier to process by extracting a window of the face or the eyes [38].

Moreover, the current literature offers few applications of the described methods directly to the EMDR task, with minimal exposition of the techniques used from an engi-

neering perspective. Therefore, it is expected that this research will provide a foundation for future studies and a reference, whose methods could prove to be an ideal direction to follow, to develop stable-EMDR-remote-therapy applications. Moreover, through the study of eye tracking and object-detection techniques, it is expected that new results will also be produced in that field, possibly setting a new baseline in the State-of-the-Art approaches currently available.

Author Contributions: Conceptualization, S.R. and C.N.; methodology, F.F. and S.R.; software, F.F.; validation, S.R.; formal analysis, C.N.; investigation, F.F. and S.R.; resources, S.R.; data curation, S.R.; writing—original draft preparation, F.F. and S.R.; writing—review and editing, S.R. and C.N.; supervision, C.N.; project administration, S.R.; funding acquisition, C.N. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was partially supported by the Age-It: Ageing Well in an Ageing Society project, task 9.4.1 work package 4 spoke 9, within topic 8 extended partnership 8, under the National Recovery and Resilience Plan (PNRR), Mission 4 Component 2 Investment 1.3- Call for tender No. 1557 of 11/10/2022 of the Italian Ministry of University and Research funded by the European Union—NextGenerationEU, CUP B53C22004090006.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the internal ethics committee of Sapienza University of Rome, as well as by the International Trans-disciplinary Ethics Committee for human-related research (protocol n. 2023030302, approved on 8 May 2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available in order to comply with the data privacy and protection regulation (mandatory for this experiment).

Acknowledgments: This work was developed at the *is.Lab()* Intelligent Systems Laboratory of the Department of Computer, Control and Management Engineering, Sapienza University of Rome, and carried out while Francesca Fiani was enrolled in the Italian National Doctorate in Artificial Intelligence run by Sapienza University of Rome in collaboration with the Department of Computer, Control, and Management Engineering.

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

References

- Shapiro, F. Eye Movement Desensitization and Reprocessing (EMDR) and the Anxiety Disorders: Clinical and Research Implications of an Integrated Psychotherapy Treatment. *J. Anxiety Disord* **1999**, *13*, 35–67; Erratum in *J. Anxiety Disord* **1999**, *13*, 621. [[CrossRef](#)] [[PubMed](#)]
- Mevissen, L.; Lievegoed, R.; De Jongh, A. EMDR treatment in people with mild ID and PTSD: 4 cases. *Psychiatr. Q.* **2011**, *82*, 43–57. [[CrossRef](#)] [[PubMed](#)]
- Valiente-Gómez, A.; Moreno-Alcázar, A.; Treen, D.; Cedrón, C.; Colom, F.; Perez, V.; Amann, B.L. EMDR beyond PTSD: A systematic literature review. *Front. Psychol.* **2017**, *8*, 1668. [[CrossRef](#)] [[PubMed](#)]
- Maxfield, L. Low-intensity interventions and EMDR therapy. *J. Emdr Pract. Res.* **2021**. [[CrossRef](#)]
- Hase, M. The Structure of EMDR Therapy: A Guide for the Therapist. *Front. Psychol.* **2021**, *12*, 660753. [[CrossRef](#)]
- Diaz-Asper, C.; Chandler, C.; Turner, R.S.; Reynolds, B.; Elvevåg, B. Increasing access to cognitive screening in the elderly: Applying natural language processing methods to speech collected over the telephone. *Cortex* **2022**, *156*, 26–38. [[CrossRef](#)] [[PubMed](#)]
- Dao, Q.; El-Yacoubi, M.A.; Rigaud, A.S. Detection of Alzheimer Disease on Online Handwriting Using 1D Convolutional Neural Network. *IEEE Access* **2023**, *11*, 2148–2155. [[CrossRef](#)]
- Haque, R.U.; Pongos, A.L.; Manzanares, C.M.; Lah, J.J.; Levey, A.I.; Clifford, G.D. Deep Convolutional Neural Networks and Transfer Learning for Measuring Cognitive Impairment Using Eye-Tracking in a Distributed Tablet-Based Environment. *IEEE Trans. Biomed. Eng.* **2021**, *68*, 11–18. [[CrossRef](#)]
- König, I.; Beau, P.; David, K. A New Context: Screen to Face Distance. In Proceedings of the 8th International Symposium on Medical Information and Communication Technology (ISMICT), Firenze, Italy, 2–4 April 2014; pp. 1–5.
- Li, Z.; Chen, W.; Li, Z.; Bian, K. Look into My Eyes: Fine-Grained Detection of Face-Screen Distance on Smartphones. In Proceedings of the 12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), Hefei, China, 16–18 December 2016; pp. 258–265.

11. Jain, N.; Gupta, P.; Goel, R. System to Detect the Relative Distance between User and Screen. *Int. Res. J. Eng. Technol. (IRJET)* **2021**, *8*, 687–693.
12. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
13. Vajgl, M.; Hurtik, P.; Nejezchleba, T. Dist-YOLO: Fast Object Detection with Distance Estimation. *Appl. Sci.* **2022**, *12*, 1354. [[CrossRef](#)]
14. Usmanhujaev, S.; Baydadaev, S.; Kwon, J.W. Accurate 3D to 2D Object Distance Estimation from the Mapped Point Cloud Data. *Sensors* **2023**, *23*, 2103. [[CrossRef](#)]
15. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
16. Kasprowski, P.; Harezlak, K.; Stasch, M. Guidelines for Eye Tracker Calibration Using Points of Regard. In *Advances in Intelligent Systems and Computing*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 284, pp. 225–236.
17. Klaib, A.F.; Alsrehin, N.O.; Melhem, W.Y.; Bashtawi, H.O.; Magableh, A.A. Eye Tracking Algorithms, Techniques, Tools, and Applications with an Emphasis on Machine Learning and Internet of Things Technologies. *Expert Syst. Appl.* **2021**, *166*, 114037. [[CrossRef](#)]
18. Hassoumi, A.; Peysakhovich, V.; Hurter, C. Improving Eye-Tracking Calibration Accuracy Using Symbolic Regression. *PLoS ONE* **2019**, *14*, e0213675. [[CrossRef](#)] [[PubMed](#)]
19. Akinyelu, A.A.; Blignaut, P. Convolutional Neural Network-Based Technique for Gaze Estimation on Mobile Devices. *Front. Artif. Intell.* **2021**, *4*, 796825. [[CrossRef](#)]
20. Vora, S.; Rangesh, A.; Trivedi, M.M. On Generalizing Driver Gaze Zone Estimation Using Convolutional Neural Networks. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 849–854.
21. Wong, E.T.; Yean, S.; Hu, Q.; Lee, B.S.; Liu, J.; Deepu, R. Gaze Estimation Using Residual Neural Network. In Proceedings of the 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Kyoto, Japan, 11–15 March 2019; pp. 411–414.
22. Chaudhary, A.K.; Kothari, R.; Acharya, M.; Dangi, S.; Nair, N.; Bailey, R.; Kanan, C.; Diaz, G.; Pelz, J.B. RITnet: Real-time Semantic Segmentation of the Eye for Gaze Tracking. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019; pp. 3698–3702.
23. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, S.; Chen, D.; Tang, Y.; Zhang, L. Children ASD Evaluation Through Joint Analysis of EEG and Eye-Tracking Recordings with Graph Convolution Network. *Front. Hum. Neurosci.* **2021**, *15*, 651349. [[CrossRef](#)]
25. Hwang, B.J.; Chen, H.H.; Hsieh, C.H.; Huang, D.Y. Gaze Tracking Based on Concatenating Spatial-Temporal Features. *Sensors* **2022**, *22*, 545. [[CrossRef](#)]
26. Rakhmatulin, I. Cycle-GAN for Eye-Tracking. *arXiv* **2022**, arXiv:2205.10556.
27. Majaranta, P.; Bulling, A. Eye Tracking and Eye-Based Human–Computer Interaction. In *Advances in Physiological Computing*; Springer: Berlin/Heidelberg, Germany, 2014.
28. Liu, J.; Chi, J.; Yang, H.; Yin, X. In the Eye of the Beholder: A Survey of Gaze Tracking Techniques. *Pattern Recognit.* **2022**, *132*, 108944. [[CrossRef](#)]
29. Viola, P.; Jones, M. Rapid Object Detection Using a Boosted Cascade of Simple Features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), Kauai, HI, USA, 8–14 December 2001.
30. Sivakumar, V.; Muruges, V. A brief study of image segmentation using Thresholding Technique on a Noisy Image. In Proceedings of the International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai, India, 27–28 February 2014; pp. 1–6. [[CrossRef](#)]
31. Kim, J.; Singh, S.; Thiessen, E.D.; Fisher, A.V. A hidden Markov model for analyzing eye-tracking of moving objects. *Behav. Res.* **2020**, *52*, 1225–1243. [[CrossRef](#)] [[PubMed](#)]
32. Shapiro, E.; Laub, B. Early EMDR intervention (EEI): A summary, a theoretical model, and the recent traumatic episode protocol (R-TEP). *J. EMDR Pract. Res.* **2008**, *2*, 79. [[CrossRef](#)]
33. Shapiro, E.; Laub, B. Early EMDR intervention following a community critical incident: A randomized clinical trial. *J. EMDR Pract. Res.* **2015**, *9*, 17–27. [[CrossRef](#)]
34. Lehnung, M.; Shapiro, E.; Schreiber, M.; Hofmann, A. Evaluating the EMDR Group traumatic episode protocol with refugees: A field study. *J. EMDR Pract. Res.* **2017**, *11*, 129–138. [[CrossRef](#)]
35. Yurtsever, A.; Bakalim, O.; Karaman, Ş.; Kaya, S.; Konuk, E. The effect of the online eye movement desensitization and reprocessing early intervention protocol (EMDR R-TEP) for the risk groups with post-traumatic stress symptoms during the COVID-19 pandemic. *Front. Psychol.* **2022**, *13*, 935782. [[CrossRef](#)]
36. Shapiro, E. EMDR Recent Traumatic Event (RTE) protocol guidelines (adapted from Roger Solomon’s workshops). *Electron. J.-EMDR-Isr.* **2003**. Available online: <https://www.emdr.org.il/wp-content/uploads/2018/05/marchproto.pdf> (accessed on 23 November 2023).

37. Dharavath, K.; Talukdar, F.A.; Laskar, R.H. Improving Face Recognition Rate with Image Preprocessing. *Indian J. Sci. Technol.* **2014**, *7*, 1170–1175. [[CrossRef](#)]
38. Piratla, N.M.; Jayasumana, A.P. A Neural Network-Based Real-Time Gaze Tracker. *J. Netw. Comput. Appl.* **2002**, *25*, 179–196. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.