Improved Smart Eye Tracker Communicator Using Low Resolution Webcam

Allana L. dos S. Rocha
University of Pernambuco
Pernambuco, Brazil
Email: alsr@ecomp.poli.br

Leandro H. de S. Silva
University of Pernambuco
Federal Institute of Paraíba
Email: lhss@ecomp.poli.br

Bruno J. T. Fernandes
University of Pernambuco
Pernambuco, Brazil
Email: bjtf@ecomp.poli.br

***Abstract*—Eye tracking is a tool presented in many applications ranging from scientific research to commercial applications. One of them is assistive technologies that aim to help people with some disabilities, including communication. However, the applications usually require specific hardware components or a high computational cost. This work proposes the Smart Eye Communicator II (SEC-II), an evolution of a previously presented algorithm to detect the pupil center and the user's gaze direction in real-time, using a low-resolution webcam and a conventional computer without a need for calibration. In SEC-II, a face aligner, which gets a canonical face alignment based on translation, scale, and rotation, has been added to the system. Likewise, strategies using eye coordinates were implemented to find the dominant eye. By implementing these new approaches, the algorithm achieved 86% accuracy, even under variable and non-uniform environmental conditions. Moreover, a graphical interface was implemented connecting the SEC-II to the internet and allowing users to express their desires and watch online videos chosen by themselves.**

I. Introduction

The concept of eye-tracking refers to a set of technologies that allow you to measure and record the eye movements of an individual, in relation to the position of the head, when showing a stimulus in a real or controlled environment, determining in which areas you fix your attention (volume of generated visual fixations) [1]. Ranging from understanding human activities and behaviors to improving human-machine interaction, eye-tracking strategies have several applications relevant to different areas [2]. Among them is the communication of people through Assistive Technology (AT) [3].

In order to improve and make people's lives easier, eye-tracking apps such as AT are created [4]. To maintain Activities of Daily Living (ADL) these technologies consist of assistive, adaptive and rehabilitation equipment [5]. As one of the solutions, alternative forms of communication emerged, helping these people to interact and participate in society.

As a possible solution to the problem of interaction and communication that some diseases generate in people, custom or invasive hardware methods, such as the Eye Tribe [1] and Tobii EyeX, which use cameras in the visible or infrared spectrum to determine the gaze direction through processing digital imaging and machine learning strategies, suffer from problems with camera positioning and lighting variation [6].

Even knowing that eye-tracking systems that do not need to be properly calibrated for use are more comfortable for users [7], many applications such as EYECAN [8] that, in addition to requiring specific hardware for their use, still need to be calibrated. Knowing that this type of hardware allows little or no head movement while using the system, which causes discomfort to users.

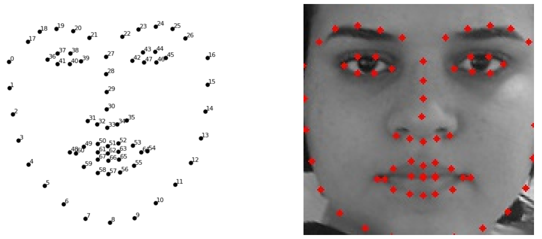
Today, Tobii® is the world's leading provider of AT for eye-tracking. Optical mice allow the use of a computer by people with, for example, Amyotrophic Lateral Sclerosis (ALS), enabling access to the internet in addition to communication with other people. However, these systems are expensive and beyond the financial reach of many people with ALS and other illnesses.

These types of factors prevent eye-tracking from becoming a more widespread technology, preventing it from being available to anyone with a reasonable camera, for example, a webcam. The objective of this work is to make eye-tracking technology more accessible and intuitive, so that people with ALS and other diseases can live more comfortably and independently.

This work proposes the improvement of a calibration-free digital image processing strategy to identify the gaze direction [9], based on five states: right, left, up, center, and eyes closed. The system has been improved to apply facial alignment strategies, before the steps of capturing the center of the pupil, as well as a new technique to search and use the coordinates of the "dominant eye", which influences the classification of the gaze direction. A database of 2,908 images was used, the same one created for the SEC [9], as it covers images with people in different positions, skin tones, and lighting. Accuracy of 86 % and 0.89 of the F1 score metric was achieved. This evolution maintains the requirements proposed in the previous work, that is, no calibration or specific hardware is required. This strategy is low-cost and can form a TA system to facilitate communication with people with disabilities, such as those with ALS.

II. Smart Eye Communicator II

An evolution of the Smart Eye Communicator (SEC) [9] software, called SEC-II, was proposed. SEC-II is an algorithm that proposes the detection of the gaze direction through a facial aligner, mathematical calculations, and facial landmarks predictors, which locate the contours of the eyes and eyelids, as shown in Figure 1.



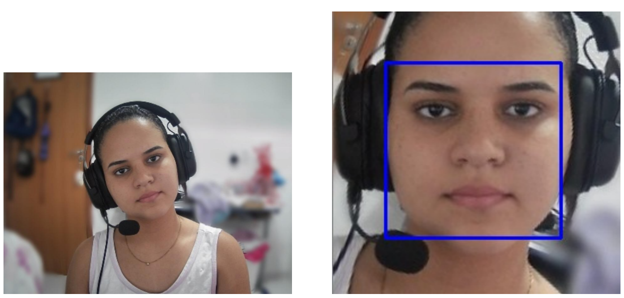
(a) All 68 face landmarks (b) Landmarks points in a real image

Fig. 1. Face landmarks and points detection in real image

A. Face alignment

Prior to facial recognition, we used a facial alignment method that uses the facial landmarks of the eye regions to obtain a normalized rotation, translation, and scaled representation of the face [10].

We use the normalization method so that the facial identifier can benefit from applying facial alignment. This procedure increases the accuracy of the facial recognition model and the robustness of the algorithm, giving the patient more flexibility, since even if his head is not centered, it will be positioned correctly. Figure 2 illustrates how effective alignment is.



(a) Original (b) Aligned

Fig. 2. Facial alignment

Considering that ALS presentations can include neck muscle weakness and head drop [11], the inclusion of this method in the system is of great value and importance. Since most people with this disease cannot position their heads centrally.

B. Pupil center coordinates

We can see in Figure 3, the flowchart demonstration of the SEC-II. Initially, when getting the image in RGB scale, it is converted to greyscale and passed to the face aligner, which obtains a canonical alignment of the face based on translation, scale and rotation. Then, steps such as clipping, application of Adaptive Histogram Equalization (CLAHE), histogram equalization, subtraction of pixel units, extraction based on color range, close and open morphological applications are performed on the images. All these processes are done in order to find the center of the pupil, because after all these applications the centroid is calculated using the image moment.

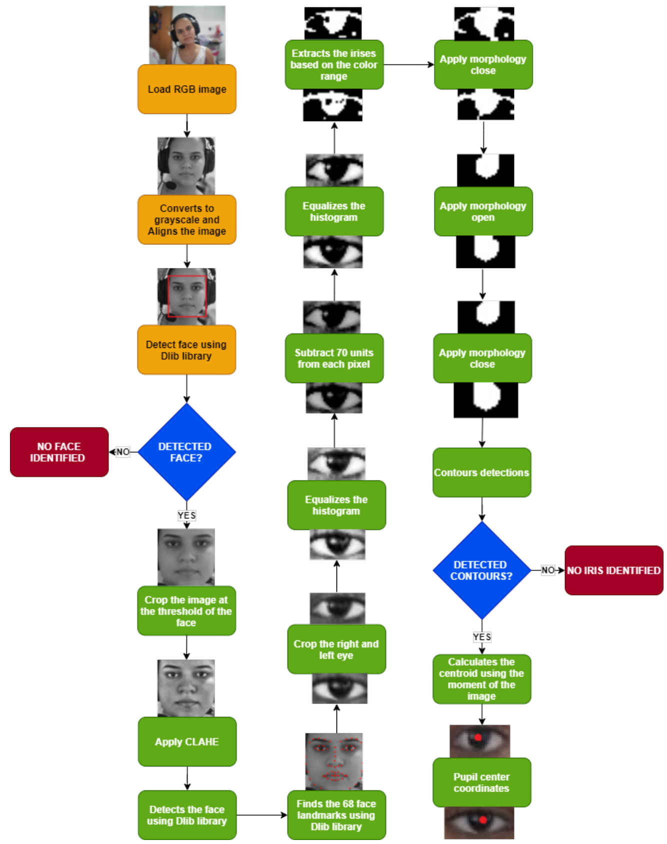


Fig. 3. Flowchart of the Smart Eye Communicator

C. Gaze direction classification and Dominant eye

As in the previous version of the [9] system, the classification of the gaze direction is based on the vertical and horizontal coordinates of the center of the pupil from the landmarks, after trying to identify the ”Right”, ”Left” and ” Up”, and none are identified, the system moves to the ”dominant eye” identification attempt step.

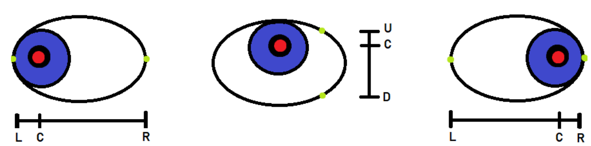
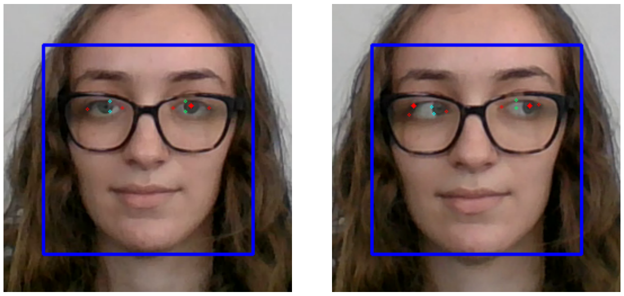


Fig. 4. The position of the reference point in the eye (the red spot in the center), in relation to the pupil (the black circle).

A method has been developed to solve the ”divergent pupil centers” problem, that is, when the pupil center is not captured correctly, in one eye, and the two centers are misaligned, as shown in Figure 5.

The logic used to identify which ”dominant eye” in the image is to identify, from the eye's coordinates, figuratively shown in the image 4, the eye in which the center is likely to be correct. That is, if after going through all the conditions mentioned above, no direction is classified, then the distance between the centers of the pupils is measured, so that if the distance is less than 65, then it is checked if the distance



(a) Center direction (b) Left direction

Fig. 5. Demonstration of error when capturing one of the pupil centers

between the coordinates verticals of C and U of the eye is at most 30%, and if the distance of the horizontal coordinates of C is between those of L and R, then the distance between L and R needs to be greater than 65%. If the above conditions are true for the right eye, the direction will be ”Right”. If not, the same process is done for the left eye, if it meets all the requirements the direction will be ”Left”. Finally, if no alternative is satisfied, the direction will be classified as ”Center”.

III. Experiments and Results

A. Dataset

To evaluate the performance of the SEC-II, the same dataset of the initial version [9] was used. Bearing in mind that the base is composed of 2,908 frames, encompassing females and males, with different lighting conditions in different locations, with or without glasses, and in different positions in relation to a laptop webcam. The division of the dataset frames is shown in Table I.

TABLE I
Number of frames for each gaze direction.

Gaze direction	Frames
Center	899
Left	749
Right	469
Up	383
Close	408
Total frames	2908

To obtain a quantitative assessment of the algorithm's robustness, we tested the algorithm's performance under different imaging conditions (Figure 6). Thus, some changes were applied to the videos for this evaluation, described in Table II.

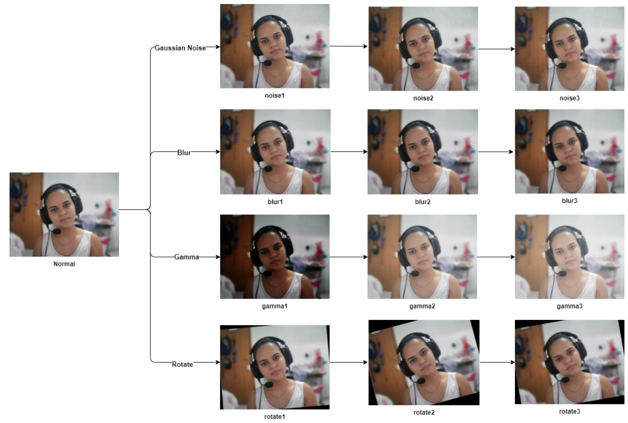


Fig. 6. Image with transformations applied.

TABLE II
Transformations applied to the dataset.

Transformation	Variations
Gausian noise	Mean = 0.01 (noise1) Mean = 0.03 (noise2) Mean = 0.05 (noise3)
Blur	Kernel = 3x3 (blur1) Kernel = 5x5 (blur2) Kernel = 9x9 (blur3)
Gamma	Gamma = 0.5 (gamma1) Gamma = 1.5 (gamma2) Gamma = 2.0 (gamma3)
Rotation	Degrees = 5 (rotate1) Degrees = 10 (rotate2) Degrees = 15 (rotate3)

B. Results

Seeking to evaluate the original dataset, that is, without any transformation, Figure 7 shows the F1 Score metric, which combines Precision and Recall to bring a number that indicates the overall quality of the model even with datasets that have disproportionate classes for every direction of look. Achieving an average of 0.83 F1 scores for the 5 directions. Given the often challenging lighting and positioning features of the dataset, it is reflected in the graph through outliers. As in the previous version of the system, the highest average F1 score remained with the left-hand drive, rising to a score of 0.98.

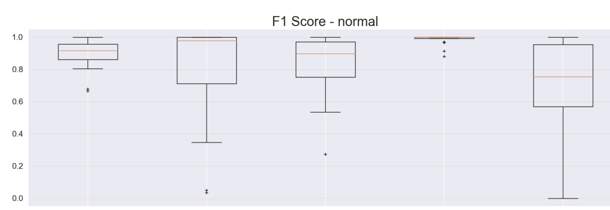


Fig. 7. F1 Score for the 5 directions obtained with original dataset.

Figure 8 reports the confusion matrix for the five directions. With the evolution of the SEC, the system obtained an average of 86% of correct answers, obtaining an improvement compared to its previous version, which presented an accuracy of 81.9%. In the new version, the direction with the highest hit rate was the center, with 813 frames correct out of 1058 frames. The system also got an improvement in the Up directions rating going from 294 corrected frames to 346, and in the right direction going from 551 to 566.

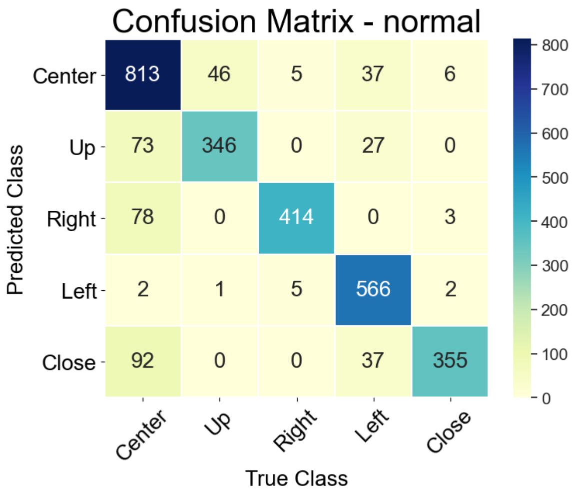


Fig. 8. Confusion matrix for the five directions.

Seeking to assess the robustness of the system, Figure 9 shows the boxplot plots of the weighted average of the F1 Score for each applied transformation and the original dataset.

It can be seen, after the analysis carried out, that the biggest impact factor, as in the previous version, were the noise applications. What was expected, given the low definition of the images transmitted by the webcam used. The results of the videos applied to gamma3 also differ considerably from the average of the other cases. However, the other transformations maintained a rate of change close to the results without changes in the images.

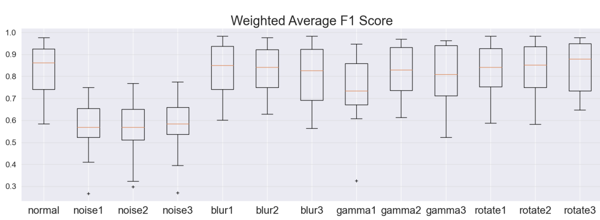


Fig. 9. Weighted average F1 Score of the transformations.

IV. Application

Using the Smart Eye Communicator II software, as shown in the Figure 10, the user can operate the graphical interface by moving the checkbox in the direction of the eye. The framework provides options to express the following needs: thirst, hunger, clearing saliva, neck pain, youtube, itching, pain, position changes, shortness of breath, and turning BiPAP (turning on Bi-level Positive Airway Pressure). These options can be easily customized according to the needs of the patient.

Confirmation is given by looking up and, and once the needs are met, the selected phrase will be played in the audio. After

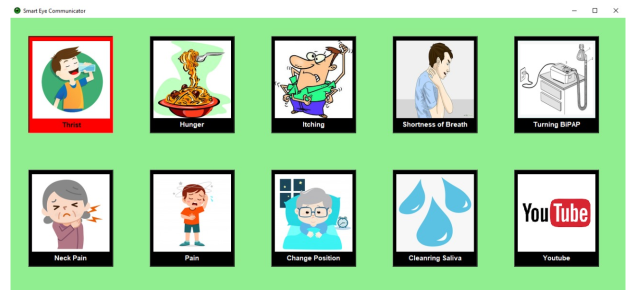


Fig. 10. Graphic interface.

the user expresses the expected demand, he can blink 3 times to end the program.

Also, if the computer running the software is connected to the internet, the user can browse Youtube using eye movement. The movement to the right and left worked as the ”Tab”/”Shift + Tab” key on the keyboard, while the upward movement works as ”Enter”. This step follows the same logic for ending the program (the project will be available online).

V. Conclusion

Eye trackers are beneficial for Assistive Technology (AT). This type of eye tracking application suffers from variations in positioning, lighting, and poor webcam quality from conventional computers. In view of this problem, an evolution of the Smart Eye Communicator (SEC) software was proposed, called SEC-II. First, a facial aligner was implemented in SEC-II, intending to overcome the head tilt problem before the detection of the pupil center and increase the accuracy of the detection. Moreover, a new step to identify the ”dominant eye” was also implemented to improve the user's gaze direction classification. To overcome the lighting and head positioning issues, the SEC continues to divide the direction of gauze into five categories: right, left, up, center, and eyes closed. These gaze directions can be used as user input to control the software for communication purposes.

Algorithm evaluation was performed using the same database as the previous version. The results showed an evolution of precision, which is 86% for the five directions, reaching a precision of 98.32% in the left direction. Furthermore, a quantitative assessment of the algorithm's robustness was made, applying various transformations to the image, including noise applications, brightness and contrast increase and decrease, as well as axis rotation and blur applications, as was done in the previous version. The algorithm results in these transformations showed that the most significant impact factor in the results was the application of noise, which is plausible, considering the increased difficulty in extracting characteristics from the images.

Acknowledgment

The authors would like to thank the PDTE (Programa de Desenvolvimento Tecnológico e Extensão – POLI/UPE) for their support through the scholarship granted. We also thank CNPq (Conselho Nacional de Desenvolvimento Científico e

Tecnológico) and FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco) for their support.

References

[1] D. Venugopal, J. Amudha, and C. Jyotsna, "Developing an application using eye tracker," in *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. IEEE, may 2016, pp. 1518–1522. [Online]. Available: http://ieeexplore.ieee.org/document/7808086/

[2] A. T. Duchowski and A. T. Duchowski, *Eye tracking methodology: Theory and practice*. Springer, 2017.

[3] W. Chareonsuk, S. Kanhaun, K. Khawkam, and D. Wongsawang, "Face and Eyes mouse for ALS Patients," in *2016 Fifth ICT International Student Project Conference (ICT-ISPC)*. IEEE, may 2016, pp. 77–80. [Online]. Available: http://ieeexplore.ieee.org/document/7519240/

[4] P. S. Holzman, L. R. Proctor, D. L. Levy, N. J. Yasillo, H. Y. Meltzer, and S. W. Hurt, "Eye-tracking dysfunctions in schizophrenic patients and their relatives," *Archives of general psychiatry*, vol. 31, no. 2, pp. 143–151, 1974.

[5] M. Arbesman and K. Sheard, "Systematic Review of the Effectiveness of Occupational Therapy-Related Interventions for People With Amyotrophic Lateral Sclerosis," *American Journal of Occupational Therapy*, vol. 68, no. 1, pp. 20–26, jan 2014. [Online]. Available: http://ajot.aota.org/Article.aspx?doi=10.5014/ajot.2014.008649

[6] H. M. Elahi, D. Islam, I. Ahmed, S. Kobashi, and M. A. R. Ahad, "Webcam-based accurate eye-central localization," *Proceedings - 2013 2nd International Conference on Robot, Vision and Signal Processing, RVSP 2013*, pp. 47–50, 2013.

[7] J. R. Khonglah and A. Khosla, "A low cost webcam based eye tracker for communicating through the eyes of young children with ASD," *Proceedings on 2015 1st International Conference on Next Generation Computing Technologies, NGCT 2015*, no. September, pp. 925–928, 2016.

[8] R. Kaushik, T. Arora, Sukanya, and R. Tripathi, "Design of Eyewriter for ALS Patients through Eyecan," in *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*. IEEE, oct 2018, pp. 991–995. [Online]. Available: https://ieeexplore.ieee.org/document/8748520/

[9] A. L. d. S. Rocha, L. H. d. S. Silva, and B. J. Fernandes, "Eye-tracking algorithm for low webcam image resolution without calibration," in *Anais Estendidos do XXXIII Conference on Graphics, Patterns and Images*. SBC, 2020, pp. 187–190.

[10] D. L. Baggio, *Mastering OpenCV with practical computer vision projects*. Packt Publishing Ltd, 2012.

[11] A. R. Martin, R. Reddy, and M. G. Fehlings, "Dropped head syndrome: diagnosis and management," *Evidence-based spine-care journal*, vol. 2, no. 02, pp. 41–47, 2011.