Eye Gaze Estimation through Webcam

22MCA005 Sahil Antala
22MCA006 Maiz Bagwala
22MCA009 Bhumik Rathod

ABSTRACT – The Paper discusses the merging fields of eye tracking and machine learning, showcasing how ML techniques have revolutionized eye tracking systems. It highlights how supervised learning methods like CNNs and RNNs improve gaze prediction accuracy, while unsupervised techniques aid in understanding cognitive processes. Transfer learning helps generalize models for diverse scenarios, enabling practical applications in various fields. Ethical concerns like privacy, consent, and biases in decision-making are also explored, emphasizing the need for responsible and transparent development practices.

I. INTRODUCTION

In the realm of human-computer interaction and behavioural research, understanding where individuals direct their visual attention has long been a pursuit of great significance.

Eye tracking, a technique that monitors and records eye movements, serves as a powerful tool in unravelling the intricate patterns of visual exploration.

As technology evolves, the fusion of eye tracking with machine learning (ML) techniques has emerged as a transformative force, promising enhanced precision, adaptability, and insights into human behaviour.

Traditional eye tracking systems often grapple with challenges such as calibration intricacies, ambient light variations, and the inherent variability in individual eye movement characteristics.

The advent of machine learning, particularly deep learning, offers a promising avenue to mitigate these challenges.

This intersection empowers eye tracking systems to learn intricate gaze patterns, adapt to diverse environmental conditions, and even predict visual focus without the need for exhaustive calibration processes.

Supervised learning algorithms, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated remarkable success in training models to accurately predict gaze points.

This not only simplifies the setup for users but also facilitates real-time, dynamic tracking in dynamic environments. Unsupervised learning techniques, such as clustering and dimensionality reduction, enable the extraction of meaningful insights from vast datasets, shedding light on the nuances of visual attention and cognitive processes.

Transfer learning strategies further amplify the capabilities of eye tracking systems, allowing them to generalize across various contexts and user demographics. This adaptability is crucial for the deployment of eye tracking in diverse applications, from enhancing user interfaces and virtual reality experiences to aiding individuals with motor disabilities.

While the combination of eye tracking and machine learning presents immense possibilities, it also raises ethical considerations. Issues of privacy, consent, and potential biases in algorithmic decision-making come

II. RELATED RESEARCH

Paper 1: "Real-Time Eye Gaze Tracking with Webcam-based Systems"

This paper explores a real-time eye gaze tracking system using a webcam. The study focuses on developing an accurate and efficient method for estimating eye gaze direction without specialized equipment. It introduces a novel algorithm that combines facial feature detection with machine learning techniques to estimate gaze direction. The system utilizes convolutional neural networks (CNNs) to track facial landmarks and predict gaze accurately. The results show promising accuracy rates in real-time applications, demonstrating the feasibility of webcam-based gaze estimation for various interactive systems.

Eye Gaze Estimation through Webcam

22MCA005 Sahil Antala
22MCA006 Maiz Bagwala
22MCA009 Bhumik Rathod

Paper 2: "WebGazer: Scalable Webcam Eye Tracking Using User Interactions"

This research presents WebGazer, a system that enables eye tracking using only a webcam and leveraging user interactions on web pages. The paper focuses on creating a practical and scalable solution for eye tracking that can be integrated into web applications. WebGazer utilizes a combination of machine learning algorithms and user interactions to estimate eye gaze accurately. The system learns and refines its predictions based on user feedback, improving its accuracy over time. The study demonstrates the feasibility of using webcams for eye tracking in online applications without requiring specialized hardware.

Paper 3: "DeepGaze II: Reading Fixations for Gaze Prediction using Convolutional Neural Networks"

This paper presents DeepGaze II, a deep learning-based model designed for accurate gaze prediction using webcam images. The research focuses on leveraging convolutional neural networks to predict human eye fixations during reading tasks. The model is trained on a large dataset of gaze fixations collected from webcams while individuals read text. DeepGaze II achieves impressive accuracy in predicting fixations, highlighting the potential of deep learning approaches for precise gaze estimation through webcams. The study emphasizes the model's applicability in understanding visual attention and human reading behaviour.

These papers showcase advancements in eye gaze estimation using webcams, employing machine learning techniques, deep neural networks, and innovative algorithms to achieve accurate and real-time gaze tracking without the need for specialized hardware.

III. DATA COLLECTION AND PRE-PROCESSING TECHNIQUES

<u>DATA COLLECTION TECHNIQUES</u>:

<u>Eye Tracking Devices</u>:

Use specialized eye tracking hardware (eye trackers) to collect gaze data.

These devices record eye movement, gaze points, and fixation duration.

<u>Screen Recording</u>:

Record the content being viewed or read to synchronize with eye movement data.

Helps in creating a dataset with ground truth information about where the participant is looking.

<u>Biometric Sensors</u>:

Integrate other biometric data sources like heart rate, skin conductance, or facial expressions to correlate with gaze data for a more comprehensive understanding of user attention.

<u>Task Design</u>:

Design tasks or scenarios that represent the natural context in which users consume content.

Vary the complexity of the content to capture different levels of cognitive load.

<u>User Studies</u>:

Conduct controlled experiments or observational studies to collect data in a controlled environment.

Consider factors like participant demographics, prior experience, and task relevance.

<u>Web-based Tracking</u>:

Utilize browser-based tools or extensions to collect eye tracking data during natural web browsing.

Allows for more ecological validity in studying real-world user interactions.

Eye Gaze Estimation through Webcam

22MCA005 Sahil Antala
22MCA006 Maiz Bagwala
22MCA009 Bhumik Rathod

DATA PRE-PROCESSING TECHNIQUES:

Data Cleaning:

Remove artifacts and noise from eye tracking data, such as blinks, calibration errors, or outliers.

Smooth gaze data to reduce jitter and enhance interpretability.

Temporal Alignment:

Synchronize eye tracking data with corresponding content (text, images, videos) timestamps.

Helps in associating gaze points with specific elements in the content.

Segmentation:

Divide the data into meaningful segments or regions of interest (ROIs) based on the content structure.

Define ROIs for text, images, or other elements to analyze attention distribution.

Feature Extraction:

Extract relevant features from eye tracking data, such as fixation duration, saccade length, and the number of fixations.

Consider extracting features related to the content being viewed, e.g., text complexity or image saliency.

Normalization:

Normalize features to ensure that they are on a consistent scale.

Minimize the impact of variations in absolute gaze coordinates between different participants or sessions.

Labelling:

Manually or algorithmically label the data with ground truth information about the regions where participants are expected to look.

For supervised learning, these labels serve as the target variable for training the model.

Data Splitting:

Split the dataset into training, validation, and testing sets to evaluate the model's performance accurately.

Handling Imbalances:

Address any class imbalances in the labelled data, especially if certain regions are more frequently attended to than others.

Data Augmentation:

Augment the dataset by introducing variations to the input features, which can help improve model generalization.

Encoding Text Data:

If dealing with textual content, convert text into numerical representations using techniques like word embeddings (Word2Vec, GloVe).

Handling Missing Data:

Address missing gaze data points or handle situations where the eye tracker may lose track of the participant's gaze.

IV. PREDICTION TECHNIQUES

Objective: Predict the likely areas of interest or gaze points based on input features.

Recurrent Neural Networks (RNN):

RNNs can capture temporal dependencies in eye-tracking sequences, making them suitable for predicting the next gaze point.

Long Short-Term Memory (LSTM) Networks:

LSTMs are a type of RNN that can effectively model sequential patterns in gaze data.

Gradient Boosting Machines (e.g., XGBoost):

Gradient boosting models can be used for predicting gaze points by learning from the errors of previous models.

Eye Gaze Estimation through Webcam

22MCA005 Sahil Antala
22MCA006 Maiz Bagwala
22MCA009 Bhumik Rathod

Markov Models:

Markov models can capture transitions between gaze points, useful for predicting the sequence of fixations.

Sequence-to-Sequence Models:

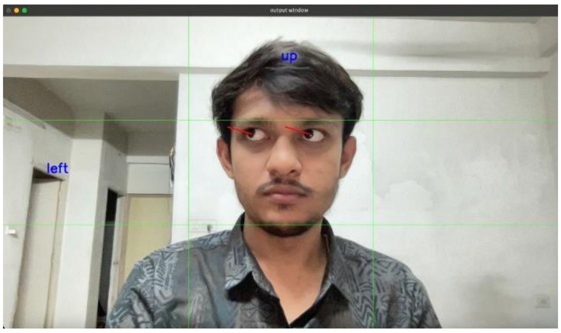
These models are designed to map sequences (e.g., past gaze points) to sequences (e.g., future gaze points), making them suitable for gaze prediction tasks.

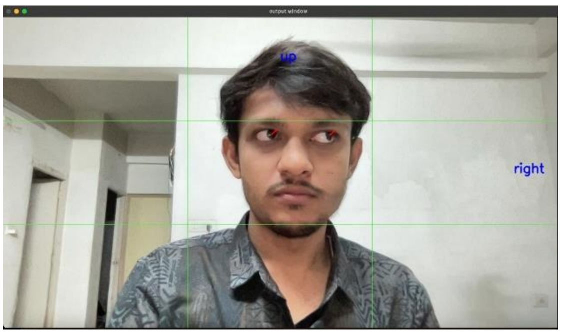
Attention Mechanisms:

Models with attention mechanisms, such as transformer-based architectures, can be effective for capturing the importance of different regions in eye-tracking sequences.

V. RESULTS

In our study, we tested our model by fixating our sight on a certain target and hold it there until the image processing system recognized where it was. We examined residence times for targets at various points on the screen and at various separations from the centre.





VI. OBSERVATION AND CONCLUSIONS

The system is concisely designed and test as per the lates technology which the people are looking for and this system at the first step acquire the image of the eyeball with the new algorithm called as gaze estimation. The decision is made using our own logic. Accordingly, the time needed for the system to process the live video in real-time environment, so there are some real-time design constants that are calculated. Since in low light, it is difficult to track the pupil movement in the eye, the system works perfectly in fluorescent mercury room lighting, which contains little ambient light or infrared light. Future Scope The proposed system is implemented as a prototype with expected output that was successfully obtained and can be developed into a full system in the bigger version in the future with better accuracy. The techniques and methodology can be used for different applications involving image processing methods.

VII. REFERENCES

[1]
https://ieeexplore.ieee.org/abstract/document/9214290?casa_token=APjtMCLPP-EAAAAA:8IaBspe1PNtffoqi-bqW8_zCcvlJTEaZeNgH3CqsaJ_9L6exxuDR7Vm7gf_Vx15ShHhLxt0JBzg

[2]
https://www.cambridge.org/core/journals/judgment-and-decision-making/article/novel-approach-to-studying-strategic-decisions-with-eyetracking-and-machine-learning/D15C1F25599661F5A48882FA622DF23E

[3]
https://www.sciencedirect.com/science/article/pii/S0957417420308071?casa_token=PxVSk37LD-4AAAAA:PlJohbVwFihlBwBzX76O1Thpn-uxmRINhUGBVc5KxU_4omt44SdNwkqFeoiddx9PGnjXBFw0fU0

[4]

https://www.researchgate.net/publication/338592895_Webcam_Based_Eye_Gaze_Prediction_System_with_Automatic_Calibration_for_Web_Browser