September 4th, 2020

Aiden Stevenson Bradwell
Unit 2, 122 Argyle Avenue
Ottawa, Ontario, K1B 4P4

Melanie Sarrazin
University of Ottawa COOP Office
75 Laurier Avenue East
Ottawa, Ontario, K1N 6N5

RE: Report of the Summer 2020 Coop 2901 Term

Dear University of Ottawa COOP Office,

I hope that this finds you well and healthy. I am a third-year student in a Bachelor of Science with Major in Computer Science and second Major in Psychology, and I have just completed my first four-month COOP term following the completion of semester 2B. Please find attached my Type I end-of-term report titled **Determining Gaze Estimation Using RGB Cameras**. Its contents are solely my own.

Being hired by the National Research Council, my placement this semester was in the Computer Vision & Graphical Interface Team, working under the supervision of Manuela Kunz (Manuela.Kunz@cnrc-nrc.gc.ca). In this position, I was given the task of converting HD webcam images of a user into accurate gaze estimations to determine the on-screen object being viewed. My days consisted of experimenting with potential solutions, and designing a final application to put the solutions to use for a future longitudinal medical study.

My experience this work term was incredibly positive and has provided me with a wealth of new methods and skills. Through daily research, I have refined my ability to both search for relevant studies, as well to find implementations of said studies for analysis. As well, I have gained insight into the fundamentals of image processing, with special regards to advanced image searching methods, and geometric analysis of images in python. Thank you for taking the time, and I look forward to any feedback you are willing to provide.

Aiden Stevenson Bradwell
300064655

Estimating and Analyzing Gaze Points Using RGB Laptop Webcams

2901 COOP Report

National Research Council, Ottawa, ON, CA

Aiden Stevenson Bradwell
University of Ottawa, Ottawa, Ontario, Canada
ABRAD060@uottawa.ca
300064655

Table of Contents

1.0 Academics

Entering my third year at the University of Ottawa, I am currently enrolled in an Honours BSc Major in Computer Science and Major in Psychology. My co-op terms are to take place in the computer science field, with a personal interest of mine being in entrepreneurship. The computer science field has an extreme range of professions, all with the common underlying task of programming however the intensity, language, and complexity of this code varies on the placement. During this summer 2020 work term I have had the great fortune to be placed at the National Research Council (NRC) on the Computer Vision & Graphics team, a job which consisted of being presented with a scientific theory and having full self-management to explore the legitimacy of said theory.

2.0 Expectations: Circumstances & Limitations

Having interviewed for and accepted this position pre-COVID, the realities of this co-op term varied greatly from the initial expectations. With the initial intentions being for me to work within a clinic using an industry standard eye-tracking device, the COVID-side-effect of lab access being minimized took its toll. As such, access to the device was removed and an alternative task for the summer was presented to me, with this task being to…

-

Aim to produce an interactive program for subjects to use, consisting of 5 major activities, each designed to invoke varying eye-movement responses, using a USB webcam to execute both real time gaze-estimation & analysis.

-

Alongside access to labs being removed, all National Research Council staff were to work from home, meeting in the morning and once a week for check-ins. While removing a large portion of the benefits from the term, such as networking, meeting new colleges, and exploring others' projects, COVID-19 provided me with an additional look into the potential of working remotely, or working from home in future careers. This came with perks such as being able to return to Windsor-Essex for the second half of the term.

3.0 Challenges Faced

Throughout the summer, a series of hurdles presented themselves, each with a unique solution requiring consideration. One such of these challenges was to produce an eye tracking device without UV lights.

-

As shown in Figure 1, industry standard eye-tracking devices take advantage of a standard characteristic of the human eye where the pupil absorbs UV light, while the rest of the eye reflects

it. As a result, cameras which operate by shining UV lights, and then measuring the light reflected can be highly effective at finding the pupil. These devices are often then either mounted on the user's head, or setup in very specific lab circumstances where the user's head is locked in place.

-

Figure 1: Representation of Industry Standard Quality as compared to RGB Webcam images [6]

Designing an eye-tracking device which does not use UV light, and instead relies on a USB RGB laptop webcam involves a different type of approach. Without a fixed head distance, fixed head position, fixed camera angle, or consistent lighting//background conditions, making assumptions on where or how the user is positioned is not possible. Compared to lab-based devices, this limits what can be done with regards to converting the pupil's position to a gaze-position on the screen.

3.1 Accurate Location of the Pupil

A considerable portion of the term was dedicated to locating the user's pupil with the highest level of accuracy, with low resolution web-cams. This was approached as a segmentation problem, meaning to take the mathematical data of the image and search it for lines, and edges. We then sought to find the best way to improve the resolution without reducing the frames per second (fps) of the program. As shown in **figure 2,** as you begin increasing the fps with RGB cameras, you begin to lose quality. With low quality images, you are met with intense pixelization, which in turn removes the ability to locate the pupil; However, if you increase the quality too greatly you no longer has enough data points to analyses their eye movements. The cause of this is the computing power of the camera, where if one requires an HD image, the camera needs more time to capture the image, and thus produces less images over a set period than tis lower resolution counterpart.

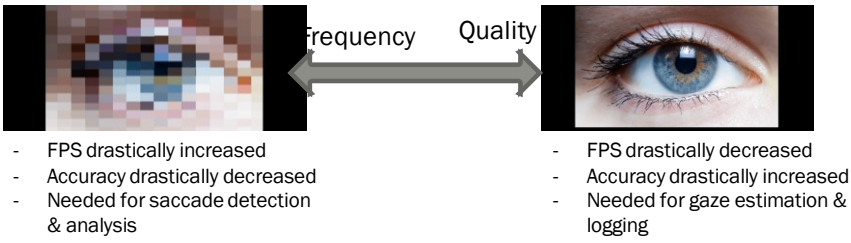

Figure 2: Tradeoff representation between quality and frequency [6]

Several Convolution Neural Networks were implemented to try bypassing the segmentation issues occurred. These had relative success, with many having flawed blinking and directional detection. In the end however, the advantages of using a neural network vastly overshadowed the flaws, and we abandoned the segmentation approach for a neural network approach.

3.2 Finding the Gaze Vector

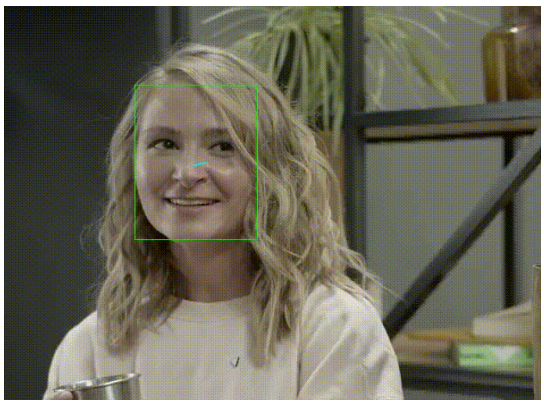

Figure 3: Gaze-Vector Visual Representation [5]

Upon locating the eye, it became important to track the eye movements from their initial position. Head drift, where the user moves their head slightly during the process of completing the activities, began to appear immediately. With most eye-movement detection systems relying on a fixed head position, where they have restrained both rotations and movements, simply tracking the current location of the pupil would not be enough to track the user's gaze.

-

As such, a neural network was applied to predict the Gaze-Vector (a numerical representation of which angle the user is looking as shown in **figure 3**) of the user. This vector has stored within it the x-slope, y-slope, and z-slope, where slope is defined in radians. This was trained using a dataset of participants in natural lighting settings, at various head positions. As such, it could accurately predict the angle of the user's gaze no matter the positioning of the user's head, or their position of the webcam. This angle vector will then be used to estimate where on the screen the user is looking, as explained in the following.

3.3 Converting Gaze into Screen Coordinates

At the start of the program, the user is asked to look into each corner of their laptop screen, and then at each edge, informing the system when they have done so. Storing the gaze data for later comparison, the program then calculates where the user is looking on the screen throughout the rest of the activities based off the ratio between the furthest left gaze vector and the furthest right **(FIGURE 1).** For example, if the user is staring at the center of the screen, the difference between their gaze vector and the vector stored for the left side of the screen would be half the difference between the right side and left side of the screen. **(FIGURE 2)**. The gaze positions are stored as points with x & y values.

4.0 Analysis of Gaze Points

The collected gaze points are stored in a text document, and upon the program being completed are analyzed for 3 gaze features. As well, where applicable the tasks have a data representation calculated from the gaze data of said task. These can then be combined with human analysis to analyze for changes in eye-function and accuracy over time between executions of the test. The data types analyzed are as follows.

4.1 Saccades

When the eye darts between two objects, the motion from one object to another is known as a saccade. The intensity, speed, and accuracy are biomarkers for cognitive decay. As such, tasks were implemented to both invoke and measure saccade movements between flashing points. Reason for this task is reliant on previous studies in the fields on soliciting such a response from the user's vision.

-

If a saccade travels past the intended target, and then corrects itself from the mistake through a much smaller saccade, it is referred to as a correction-saccade. Correction-saccades are a key marker of the accuracy one has for their eye-movements. They often go unnoticed, and a user who makes large mistakes in their saccades without immediately applying a correction saccade is demonstrating interesting behaviour.

-

4.2 Anti-Saccades

If one is asked to focus on an object, while motion occurs behind it, they may experience an anti-saccade, defined by an unintentional saccade away from the point the person is focusing on. A correction saccade is then implemented to return to the original focus target. This can be triggered by requesting the user focus upon a point, and then introducing distracting alternative points on the screen. This would then cause the user to accidentally look away from the target and rapidly correct themselves. A user who makes large anti-saccades and is slow to correct is displaying abnormal eye-movement behaviour.

4.3 Fixations

When the gaze of a person is focused upon a single point for an extended period, it is referred to as a fixation. Saccades occur between two fixations, and anti-saccades occur when a fixation is broken unintentionally. The occurrence of a fixation triggers locking neurons in the brain, reducing the likelihood of anti-saccades. Requesting a user to stare at a point or still object is enough to trigger a saccade.

5.0 Relevance to Program

5.1 Computer Science (Primary Module of Study)

The computational processing of imagery was a field I had not looked into prior to beginning this position. Fortunately, this position has given me a new insight into the fundamentals of the field, such as the over encompassing way which images are seen internally. Computer vision stores images a 2d array (list of lists) consisting of the color information for each pixel. As such, it does not see shapes, objects, or patterns without implementing external algorithms upon this array.

-

Image processing entails taking this array of pixels, and editing them through blurring, removing certain data, or filtering the data based on its' values. Libraries designed to read, process, and interoperate video data through processes such as Hough Circles were implemented in our system, as well as including advanced processing techniques explored throughout the term.

-

During the process of prototyping, I was also required to implement a Convolutional Neural Network (CNN), trained through the TensorFlow library, using a large data source collection from real-world visual data. This dataset consisted of tens of thousands of images where people's gaze angles were hand marked. The process of seeing how to train a CNN was very interesting and has opened a potential field to look into for future placements.

5.2 Psychology (Secondary Module of Study)

Throughout the term, there was an underlying study of psychology implemented into the design of the main program. As such, I have explored a section of psychology that I had not previously considered: Neural-Visual interaction. Examples include that function and meaning of saccades, anti-saccades, and fixations on both an analysis and psychological level were explored.

-

As well, a detailed research process into cognitive decay in patients across numerous studies was required, introducing me to the research aspect of psychology, as well as implementing these studies into a physical project.

-

This coop term has gone a distance to show to me that the two fields of study I have chosen, while seeming to have no connection, can be combined into interesting projects in both research and real-world implementations of programming. As such, I have begun looking into further explorations of psychology-computer science topics.

5.3 Method of Researching

A staple portion of this semester was designated to reading through research papers on psychological signs of cognitive decline, algorithms used in object detection, and convolutional networks.

-

One of the key things I've noted is that the type of research paper varies between the different topics. Psychology papers are extremely text heavy, with focus put into the background of the topic and the implications of the results. There are often few visuals, and the visuals included are data representations of test results.

-

Computer Science literature seems to be less text-heavy, with emphasis put of visual examples over textual. There is often very little background information from my experience, and there is far more emphasis put on the implications on the research. Computer science research papers seem to take a tone of self-importance as well.

6.0 Performance Objectives

Retrospectively reviewing my performance objectives, I find that I have achieved each in its own way. As such, this placement has gone a long way in achieving what I had intended for it.

6.1 Objective One

Produce a real-time computer-vision application usable for infield patient testing, which combines gaze tracking solutions with data analysis.

This goal worked as the overall objective for the summer. Encompassing the main task I had received, producing this software encompassed what I had worked on, and is the final product which I produced. Consisting of 5 tasks, each of which designed to invoke some visual response such as saccades, or fixations, and taking roughly 15 minutes to complete, this program then produces visual analysis of the collected data.

-

The first task consists of an initially still ball, which the user is asked to track as it moves across the screen. The gaze position of the eyes is then recorded as it does so, with the post analysis consisting of a retrospective saccade & correction saccade detection. The gaze points throughout the task are stored within a text document, while the detection of saccades & correction-saccades are stored within a separate document. Human review of the overall saccade accuracy is required as automated analysis is not currently implemented.

-

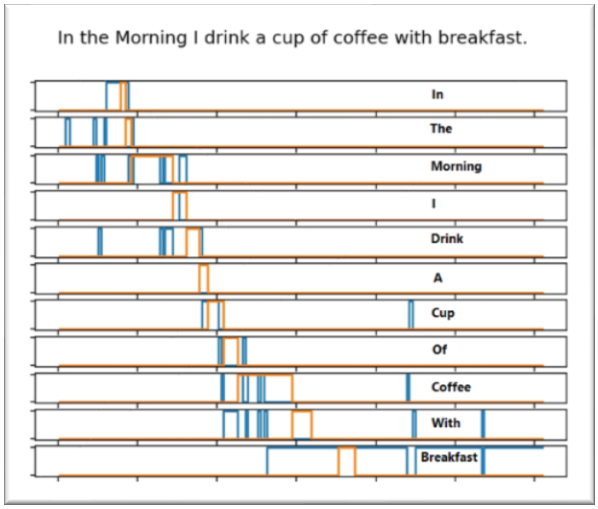

Figure 4: Visual representation comparison of reading time VS speaking time

Following this, the user is presented with a series of sentences. They are asked to read these sentences aloud, while the system analyzes which words they are looking at while doing so. A graph is produced automatically of which words they were gazing at while reading the sentence, and an overlaid human-determined analysis of when they read each word is applied. This figure dictates the amount of time and focus the user had placed on each word in the sentence, as well as which words were skipped over during the reading. The intention of said task is out of interest of which words were fixated on, and which words were overlooked due to their predictability, as demonstrated in **Figure 4**.

-

Continuing through the program, the user is presented with instructions to scan through an image in search of an object within it. The gaze information is stored, and then automatically analyzed to visualize where the user was looking as they scanned the image. An important piece of information is to check if they had looked at the object prior to acknowledging it, and how cognizant they were of doing so as shown in **Figure 5.**

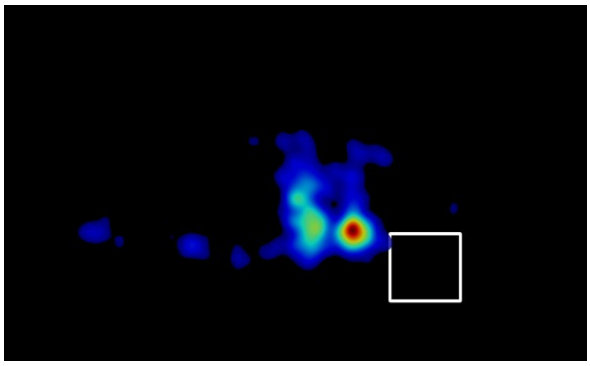

Figure 5: Heatmap representation of a subject's gaze path

Lastly, the user enters into a two-part task, each designed to invoke either saccades or anti-saccade. During the first phase, the user is requested to gaze between points which appear in one spot, and then begin disappearing and reappearing in alternative locations. The movement between these changing spots triggers rapid saccades of the user, and can be analyzed for intensity, accuracy, and reaction speed.

-

Pursuing this, a spot is projected onto the screen, with the user asked to stare into it. A series of red distraction circles are then displayed surrounding the initial dot, intending to pull the users gaze away from the dot and hence triggering an anti-saccade in the progress.

The intensity of this anti-saccade can then be analyzed for discrepancies with previous test results.

6.2 Objective Two

Assist in the development of critical projects applicable to the Canadian COVID-19 response, through research, authoring, brainstorming, and programming.

During my time with the National Research Council, I had the fortune of assisting in the proposal process for potential COVID-19 related tasks. This work consisted of being presented with the potential project's main idea, and then conducting a deep dive of the related field of current research to check the validity of the theory.

-

This brings with it a few challenges, namely that the research being released on COVID-19 is incredibly early in its review stages, and the science is constantly changing. With this comes the challenge of coming up with any theory around a recently released paper, out of a worry that the science will be rapidly undermined by a more modern study. One needs to look no further than modern CDC & WHO guidelines for stopping the spread of the virus to see this in action. With the rapidly evolving nature of our understanding of COVID-19, firmly making a prediction of the most valid, pre-peer review, research becomes very difficult to do; a majority of the papers which were being released we're posted without peer review or were awaiting board review, causing consistent disclaimers throughout the proposal of potentially invalid proof.

6.3 Objective Three

Conduct the collection of data from human trials, using this data to validate leading edge cognitive research theories.

Initially during this semester there was potential for me to gain access to the lab in order to conduct professional collection of data with volunteers. The application which I developed over the term as designed to conduct longitudinal studies of patients over the course of many months, across a wide range of volunteers. Unfortunately, with the realities of this term, I was unable to achieve this goal fully, as lab access was officially taken off the table entering the half-way point of the summer. With this the opportunity for lab-condition testing and prototyping was also removed alongside it.

-

In light of this, the focus of the project began to shift towards that of in-home-testing, hence the reliance on the standard and easily set-up RGB laptop USB cameras. The program is designed to be interchangeable between these two settings, both in a laboratory and in home, in an effort to keep the lab-based route open for future studies.

-

While not having lab access this summer is unfortunate, I have been invited to continue to participate in the collection of data and allowed to assist in the overseeing of the study's execution with volunteers. This is a great opportunity to experience an in-field psychological study, the proper storage of confidential information, the application process for a study, as well as the ethics associated with looking at the collected results. Being included in this study will work as a great glimpse into the combinations which computer science and psychology can create, and the interesting outcomes of being well versed in the overlap of both fields.

7.0 Learning Objectives

Entering my first work term, I was not certain of what to anticipate, especially with the current international climate. With regards to support, training, and opportunity I found myself concerned of a potential lack. While networking within a zoom-call based position is incredibly difficult, it was fortunate that both training and support were both abundant within the National Research Council, and my supervisor was very willing to meet my learning objectives with me.

7.1 Objective One

To gain a thorough understanding of computer-vision field, including the Ai analysis of images for object detection and recognition, as well as to learn the foundations of advanced automated image processing techniques.

Prior to this position I had not considered computer vision as a potential path of career. I often found myself drawn to the more over-encompassing umbrella term of AI (Artificial Intelligence), without specifying which field I would prefer to apply it to. The computer visions team at the National Research Council has allowed me to take a real look into some realities of Artificial Intelligence, namely in the Convolution Neural Network area. While it was a quick and efficient approach to implement, I enjoyed to manual manipulation of images far more, as it was incredibly interesting to approach vision from a purely mathematically perspective. With the job came a new consideration for the amount of effort that the human eyes require to do what we take full advantage of, and how rudimentary most computer's applications are with regards to vison.

7.2 Objective Two

To plan, propose, and execute human trials using the software developed, to further my training and gain real world experience in the field of psychology, and moreover gain experience in the collection and analysis of human data.

To break this objective down into its core components, I have either completed the intended goal or have made plans to do so in the near future. Executing human trails

wasn't an option for the priorly mentioned reasons, and as such I was made scheduled plan to assist in the planning, proposing, and completion of the initial stages for the longitudinal study associated with y application. I am eager to see how the process of psychological studies works from a real-world setting, outside of academia. My time at the National Research Council has shed light on the fact that not all research is conducted within universities, and that there are many public-sector research opportunities available, with lots of funding to be had for the right ideas.

-

Additionally, this position has provided me with real experience in the field of psychology, from the research process of new findings, to the inference of the results into potentially large real-world applications, to the execution and comparison of the data collected through those applications.

8.0 Future Improvements

Going forward with the project, I have been invited to continue participation with the lab collection of data with volunteers. This will allow me to explore the more lab-based side of psychology, as well as allow me to continue improving and debugging my produced application. I have several intended improvements to produce better, more accurate data, as well as to analyze that data more efficiently.

8.1 Method of Gaze Estimation

The current method of gaze-location prediction has a fundamental flaw where if the head begins to move, the estimation becomes highly inaccurate. With the intention of data collection to be in-home, this is not ideal, as it would require some type of head constraint before the program is run.

-

As such, a new method can be implemented, where an abstract 3D space is produced. The camera will be located at the origin of said 3d space, with the X-Y plane being the location of the abstract screen (Figure X). The user is then asked to look towards each corner of their laptop screen during the calibration sequence (Figure X), which then sections off a portion of the X-Y Plane (Figure X), which will now act as the representation for the laptop monitor within this space. Using the location of the user's nose in the camera's vision, and projecting the gaze position from this point towards the X-Y Plane (Figure X), the system will be able to tell if the intersection of this line and the plane falls within the marked monitor frame from the calibration (Figure X).

8.2 Method of Researching

Throughout the term, the methods which I used to research improved drastically. AN important aspect of research which I was unaware of entering the semester was the usage of key terms while searching. It is tempting to research a phrase which you are seeking to prove or back up, however the proper way to search seems to be heavily dependent of picking up the correct key words and transferring them into the search.

-

Initially I had difficulties transferring a found paper into a physical example and took the approach of attempting to implement the algorithms myself through coding. As a result, I lost a lot of time to implementing other's previous attempts, where I could have easily discovered their already coded and functioning versions.

8.3 Scientific Approach

While searching for implementations of algorithms to be used in the application, I often found myself struggling to locate code from relevant papers. I would have been able to save a large portion of my overall time if I had placed more effort into reading all previous work done in the related field. It was common for me to find myself exploring a path which had already been explored in priorly-published works and discovering the same conclusions that they had already discovered. As such, as I explore further into the computer science field, I can improve greatly by thoroughly researching topics before attempting to approach them.

-

This issue appeared while researching algorithms to locate the pupil of the eye in low-resolution RGB images. There were some papers existing on the topic of doing such a task, which if I had read more thoroughly earlier in the process, hint towards using a Convolutional Neural Network, which was ultimately the decision we had made.

9.0 COVID-19

At the time of accepting this position, COVID had not yet become a global health emergency. As a result, the realities of my work term had changed drastically between the time of accepting and the time of beginning my employment. The most drastic effect which this had on my work term was that it removed most all of the marketing opportunities which would have been available to me otherwise. I was not able to meet with the term I was assigned to, nor did the opportunity arise to form close relationships with my colleagues on the team nor within the organization. The National Research Council has a reputation for having a great community around it, with amenities and activities for staff to bond and relax. Due to the shift to at-home working these were removed, and the job took a very professionalism-only tone. This felt odd as the team still seemed to be approaching things in a way of comradery, likely due to them already being well acquainted entering the work-from-home order. This, however, made it difficult to 'get one's foot in the door',

and leaving this term I do not feel I have made many contacts. While my supervisor was phenomenal, and was generous enough to offer me a future term at the National Research Council, this coop term has an aura of throw-away to it, where the main thing I have gained is 4 months infield experience, and a great summer job.

-

It may be due to my intentions for co-op, but I heavily weight the human aspects of my future career, with who you know being more important than what you know more often than not. Opportunities to network allow for one to meet potential future employers, future contacts in the field, and moreover future business partners if one were to explore the entrepreneurship route. With such a strong weight being put on this as I personally have placed upon it, these feelings of being jibbed may be inevitable for the COVID-19 pandemic. I am, however, profoundly thankful for the opportunity which I have been given considering the same opportunity has been removed from many of my peers during their first coop term.

References

1. 100 Computer Science Careers to Consider. (2018, September 14). Retrieved September 12, 2020, from https://www.careermatch.com/job-prep/career-insights/articles/majoring-in-computer-science-100-careers-to-consider/

2. Appearance-based Gaze Estimation in the Wild (MPIIGaze). (n.d.). Retrieved September 12, 2020, from https://www.mpi-inf.mpg.de/departments/computer-vision-and-machine-learning/research/gaze-based-human-computer-interaction/appearance-based-gaze-estimation-in-the-wild

3. Bulling, A. (n.d.). Perceptual User Interfaces. Retrieved September 12, 2020, from https://www.perceptualui.org/research/datasets/MPIIFaceGaze/

4. Computer Science jobs. (n.d.). Retrieved September 12, 2020, from https://ca.indeed.com/Computer-Science-jobs

5. Hysts. (n.d.). Hysts/pytorch_mpiigaze. Retrieved September 12, 2020, from https://github.com/hysts/pytorch_mpiigaze

6. Webcam-Based Eye Tracking vs. an Eye Tracker [Pros & Cons]. (n.d.). Retrieved September 13, 2020, from https://imotions.com/blog/webcam-eye-tracking-vs-an-eye-tracker/

Figures

1. Representation of Industry Standard Quality as compared to RGB Webcam images [6]

2. Tradeoff representation between quality and frequency [6]

3. Gaze-Vector Visual Representation [5]

4. Visual representation comparison of reading time VS speaking time

5. Heatmap representation of a subject's gaze path