Doorbell Using Facial Recognition System

Anas Hamid Ali
(1152156)
ahamida@lakeheadu.ca

Karan Patel
(1148274)
kpatel77@lakeheadu.ca

Tirth Patel
(1143729)
tpatel16@lakeheadu.ca

Abstract

Bio-metric screening using fingerprint or retina scans have been in circulation for a while now. Since the advancements in high-resolution cameras and development of 3D facial recognition algorithms in recent times, facial recognition as a means of bio-metrics has attracted vast majority of enthusiasts. A facial recognition system is a technology capable of matching a human face from a digital image or a video frame against a database of faces, typically employed to authenticate users through ID verification services, works by pinpointing and measuring facial features from a given image. The objective of this project is to implement door-bot which provide security to home. This will help house owners keep track of visitors, even remotely. Moreover, as Robberies and Kidnapping have been a major problem these days when considering our peer safety, This system would also help deal with such scenarios and warn owner before-hand for suspicious activities.

1 Introduction

Security is a basic requirement of every living being. In this modern world crime has become ultramodern tool in this current time a lot incident occurs like robbery, stealing unwanted entrance happens abruptly. In such cases, Security does matters does matter the most in daily life. Be it a safety for own belongings or own life. People always remain busy in their day-to-day work also wants to ensure safety of their beloved ones and to keep their things secured and protected from any unwanted incidence and thus this project is implemented. Through this system user can keep an eye on their system from anywhere in the world who is at their door steps, the system keeps a picture of the visitor as evidence that would be needed if any unwanted situation occurs like stealing, robbery etc.

A face detection system is designed to answer the question: is there a face in this picture? A face detection system determines the presence, location, scale, and orientation of any face present in a still image or video frame. A face recognition system is designed to answer the question: does the face in an image match the face in another image? A face recognition system takes an image of a face and makes a prediction about whether the face matches other faces in a provided database. Both face detection and face recognition systems can provide an estimate of the confidence level of the prediction in the form of a probability or confidence score. If a facial detection system incorrectly predicts the presence of a face at a high confidence level, this is a false alarm or false positives. Similarly, a facial recognition system may not match two faces belonging to the same person. However, for use cases where the risk of missed detection or false alarm is higher, the system should use a higher confidence level. You should use a 99% confidence/similarity threshold in scenarios where highly accurate facial matches are important.

Modern buildings are typically equipped with wireless door-bell systems that employ radio technology to signal them and answer the doors remotely. Although these doorbells are much more convenient than wired ones, they do not always satisfy the demands of modern homes for the following three reasons :

1. Answering machines are normally located at a fixed place (often near to the door), if an occupant wants to answer the doorbell, he/she has to go to the answering machine.

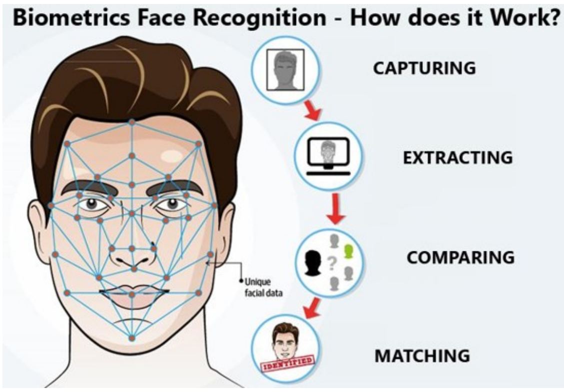

Figure 1: How does Facial Recognition work?

2. If the occupant would like to see and recognize the visitors, he/she has to go to door.
3. The occupant has no way to answer or admit guests when he/she is not at home, nor to keep a record of guests.

As smart home technology matures, smart doorbells can solve this problem greatly by connecting the doorbells to the Internet (or a local network) and allowing users to answer the bell through a smart device such as a smartphone or tablet. This enables a homeowner to answer and admit a visitor anywhere a smart device connecting to the Internet is available. However, such smart doorbells are quite expensive (usually more than 190 US dollars) due to technical and manufacturing difficulties. The high prices make these products unavailable to most home users with limited budgets, hindering the pervasiveness of smart doorbells. This is confirmed by a research shows that less than 4% of U.S. households have a smart doorbell system to protect the security of the homes.

2 Related Work

The study by Thabet and Amor[1] influenced us the most to create this project. It depicted a system that can identify and recognize faces, which can be used for a variety of purposes such as crowd and airport monitoring, private security, and improved human-computer connection. An automatic face recognition system is ideal for addressing security concerns while also providing smart house control. The goal of their paper was to use a Raspberry Pi board with an ARMv7 Cortex-A7 processor as the core and the Opencv library to replace expensive image processing devices. Their study was primarily focused on image processing, with the Opencv library being ported to the Raspberry Pi. However, we decided to maintain the webcam instead of the Raspberry Pi. Another issue that doesn't fit is that we intended to use AWS services in our project.

Viola, Paul and Michael[2] combined novel algorithms and insights to create a framework for object detection that is both reliable and fast. The job of face detection is used to explain and motivate this architecture. In order to do this, they managed a frontal face detection system that achieves detection and false positive rates that are comparable to the best reported results.

A face detector that is incredibly fast will have a wide range of uses. User interfaces, picture databases, and teleconferencing are examples of these. Real-time face detection applications will be possible on systems that were previously unfeasible due to the improvement in performance. This solution will allow for significant additional post-processing and analysis in areas where high frame rates are not required. Furthermore, their technology can be used on a variety of low-power devices, such as handhelds and embedded CPUs. This was the paper we used to get the face detection.

Another study[3] we came across expands their rapid object detection technique in two key-ways: To begin, an efficient collection of 45° rotated features is added to their fundamental and over-complete set of Haar-like features. Second, for a given boosted classifier, we derive a new post optimization approach that dramatically enhances its performance. We studied their efforts and were intrigued enough to incorporate it into our project, but it appears that the previous classifier detection is sufficient for us to detect on the doorbell camera.

3 Proposed Method

The system's central idea is that if a visitor attempts to visit when the owner is not at home, a camera (in our case, a webcam) takes a photo of the visitor. We used the Open Source Computer Vision Library and Haar feature-based cascade classifiers to both capture and analyze the image. As the technology is emerging and has broad usability, we plan to discover a variety of AWS services and incorporate them into our project. One of the most important services we will employ is AWS Rekognition. One can also use Amazon Rekognition to recognize, analyze, and compare faces for a range of user verification, people counting, and public safety use cases. We utilize AWS CLI to publish the collected image to AWS S3 buckets after we've detected the face. Following that, AWS Rekognition was used to search through a default faces photo collection, exhibiting AWS' machine vision capabilities. The visitor's name is extracted from the database table if a match is found; if no match is found, the image is indexed as a new face, and the image is delivered to the owner's phone using AWS Simple Notification Service. Some of the Proposed System's Benefits

1. This is a simple and useful security system that is simple to install.

2. Crowd and airport monitoring, private security, and improved human-computer interaction are all possible uses for a system that can differentiate and recognize faces.

3. Faces of house members can be saved in Amazon's S3 buckets, and comparisons with visitors' faces can aid distinguish between house members and strangers.

4. Two-way communication

5. Can see the visitor's face, as well as the date and time, on any of our smart devices, which can be accessed from anywhere.

6. The message protocol can be used for both SMS and email.

3.1 Architecture

The architecture of our project is shown in the diagram; when the doorbell is pressed, the camera detects the face and clicks it through the camera, then stores it in the s3 bucket by recognition.

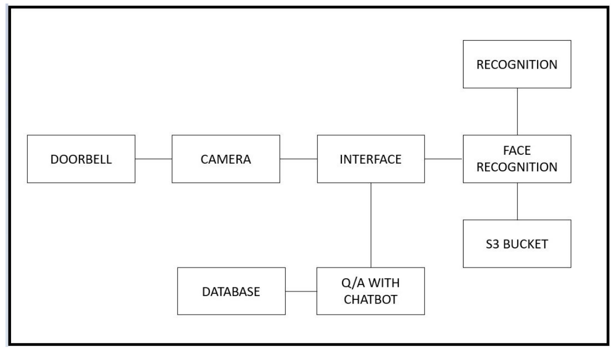

Figure 2: Architecture

3.2 Use Case

The diagram depicts a use case with two users. The first is the visitor, who can use the doorbell, Chat bot,and camera. The owner, who may examine the user's interaction and details in his phone through a message, is the other main user.

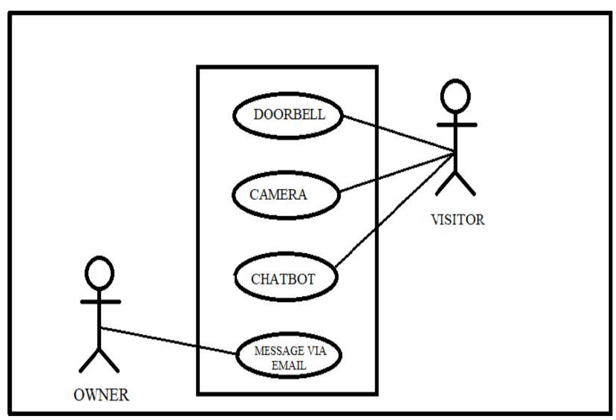

Figure 3: Use Case Diagram

3.3 Workflow Diagram

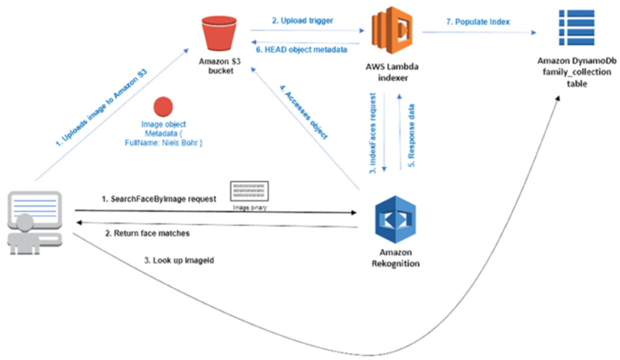

Figure 4: Workflow Diagram

4 Experimental setup

4.1 Dataset

For Dataset and to get these classifiers, we tried various trained classifier XML files from the open CV GitHub. After experimenting out various scenarios, we used haarcascade-frontal face-default.xml; which perfectly fitted our project requirements and scope.

We learnt about effective object identification approach based on Haar feature based Cascade classifier. It is a machine-learning approach in which a cascade function is taught for a large number of positive and negative images. So, we initially trained the classifier using a large number of photos that contain faces, which are referred to as positive images because they contain the target item as per our project scope. Similarly, we also trained the classifier on negative photos, or images that don't include our target object. After the classifier was perfectly trained, we applied it to a region of interest in an input picture; with the classifier returning a '1' if the region is likely to exhibit the object or a '0' if the region is unlikely to show the object.

4.2 System Outlook

The basic outlook of the system is, if a visitor tries to visit while owner is not at home, a camera snaps a picture of the visitor (in our instance, we're using a webcam). To capture and analyze the image, we have used the Open Source Computer Vision Library and Haar feature-based cascade classifiers. After detecting the face, Python software will use AWS CLI to publish the pictures to AWS S3 buckets. Post that, AWS Rekognition service has been used to search from an default faces photo collection, demonstrating the capability of AWS in machine vision. If a match is found, the visitor's name is retrieved from the database table; if no match is discovered, the image is indexed as a new face, and the image will be sent to the owner's phone using AWS Simple Notification Service.

4.3 Implementation Stack

4.3.1 AWS S3 Buckets

Through a web interface, Amazon S3 enables object storage. It's designed to store, protect, and retrieve data from "buckets" on any device, at any time. A bucket is created by a user, and the bucket is used to store objects in the cloud. All our default family members photographs have been included in this. As a result, it assists AWS rekognition in comparing the real time photo to the default photo stored in the AWS S3 bucket.

4.3.2 Amazon Rekognition

Cloud-based software used to extract metadata from image and video files. This is used to add image and video analysis to our application. It just requires uploading a photo or video to the Amazon Rekognition API, and the service will recognize objects, people, text, scenes, and activities. Compared to other applications, Amazon Rekognition can also perform highly accurate facial analysis, face comparison, and face search. Considering our project scope, we're only interested in getting the captured and detected image from the webcam.

4.3.3 AWS Lambda

Event-driven and serverless computing platform that runs code in response to events and automatically manages computing resources.

4.3.4 Amazon SageMaker

Cloud machine-learning platform, enables developers to create, train and deploy machine learning models in the cloud.

Cloud-based software used to extract metadata from image and video files. This is used to add image and video analysis to our application. It just requires uploading a photo or video to the Amazon Rekognition API, and the service will recognize objects, people, text, scenes, and activities. Compared to other applications, Amazon Rekognition can also perform highly accurate facial analysis, face comparison, and face search. Considering our project scope, we're only interested in getting the captured and detected image from the webcam.

5 Results

5.1 Solo

The model was tested on over 100 different looks of each celebrity with different hairstyles, makeup, etc. It gave a similarity accuracy of more than 99% and performed best when the threshold was set to 70. Figure 5 shows the matching similarity percentage of among the source image and target image. Furthermore, Figure 6 shows the in-depth analysis of the person identified.

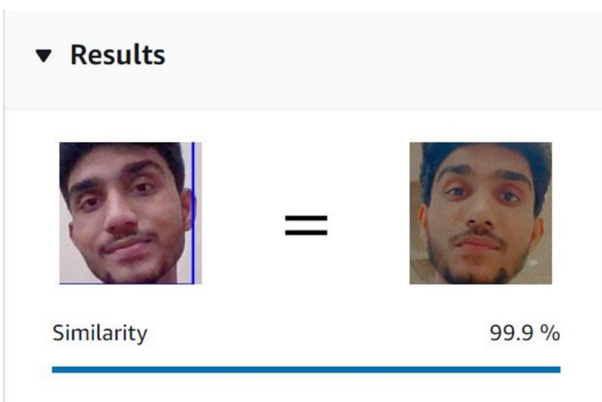

Figure 5: Solo guest results

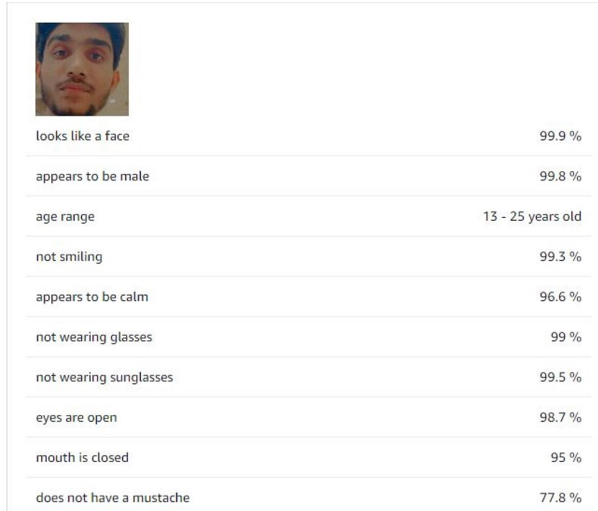

Figure 6: Solo guest result analytics

5.2 Group

The model was clearly able to identify the target person among the group of people(Figure 7) when arriving altogether at the door step. The person resulting in the highest similarity percentage with the target image, is the output. As shown in Figure 8, the matching image gave a similarity of more than 99% and for the non-matching images the similarity will be less than 1%.

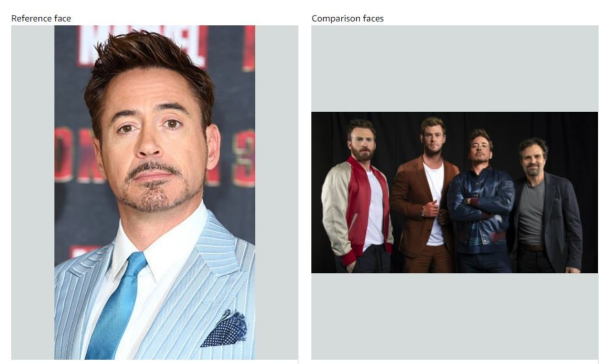

Figure 7: Group of guests scenario

6 Discussion

Modern residences are usually outfitted with Wi-Fi routers and Internet connectivity. Smartphones are also highly available to the majority of population. To build

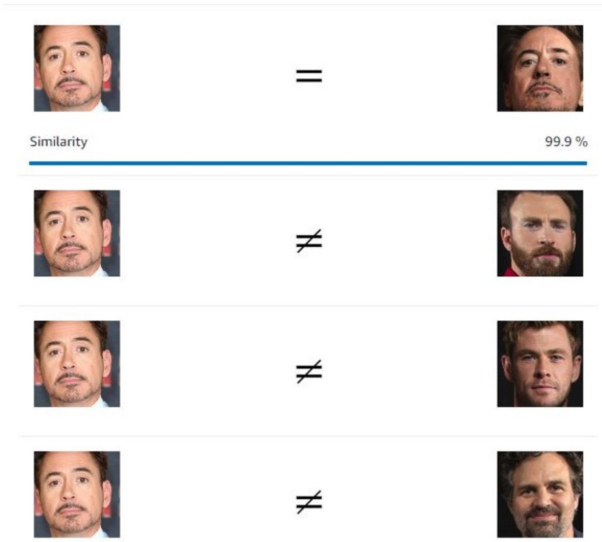

Figure 8: Group of guests similarity results

a low-cost smart doorbell system like this, the user only has to buy Amazon services, a Raspberry Pi if needed, a webcam, and a buzzer, with the rest depending on the user's demands and usage. Our system, we believe, distinguishes itself from other smart doorbell systems in the following ways. To begin with, it is significantly less expensive than existing smart doorbells. Second, rather than being an integrated system, it is a dispersed system, making failure detection and diagnosis faster. If some of the components fail to function, for example, instead of disassembling or replacing the entire doorbell system, one can simply identify and repair or replace the parts by verifying each individual device. Third, because the majority of smart doorbell gadgets are pricey, they can be easily removed and stolen. However, in our instance, it is both inexpensive and changeable, making it a far superior option in terms of device security.
While our system has some advantages over a traditional doorbell in terms of features and security, it does have a few security and privacy concerns. Because the device is connected to a home WiFi network, it is possible to compromise the network and use the device, allow access to uninvited visitors, or gather data without the knowledge of the owner. Our project also takes and keeps photographs of visitors without their permission.

7 Limitations

The system limitation is that it can even detect 2D images or photos if they are provided. So, in order to solve this problem, we email the owner a real-time photo so that it can be identified if it was person holding an image or an actually present person.
It's a subscription-based project, which means you'll have to pay for the cost of running resources like SNS, S3 Bucket, and Lambda services on a regular basis.
Due to extreme weather or other unfavorable situations, it is occasionally hard to distinguish faces, and it may also be unable to portray some moving objects.

8 Challenges and Future Work

The major challenges we faced when dealing with extreme scenarios were :

- Major challenge we tracked upon was bifurcating between 2D and 3D images. As we had to handle the scenario that, if a stranger comes by and shows the image of family member and accesses the entrance.
- Another challenge we faced was to maintain the recognition accuracy across bright day light that is to manage light exposure, etc. and night vision that is to manage Ambient light, illumination, etc.
- The last and usual challenge was to accurately detect and identify each face when a large group of people show up all together.

Below are the modules to be considered for future enhancements :

- Include Raspberry Pi as an add-on attachment to improve results.
- We intend to implement two-way audio and video communication between the visitor and the homeowner in the future. We'll also incorporate face recognition, natural language processing, and machine learning techniques (e.g., Support Vector Machines, Neural Networks) to automatically answer the door without the need for human participation.

- Add Chat-bot to improve communication.
- Identifying accessories, the visitor is carrying/wearing and determining his occupation. E.g. Visitor wearing a cap and a bag can be considered as delivery boy.
- Identifying sharp objects in the frame and warning the owner before-hand.
- Introduce AWS Polly to greet and welcome guests.

9 Conclusion

This project shows how to remotely monitor and control a door. We were able to successfully combine facial detection with the Haar Cascade Classifier, a pre-trained model. As a result, we acquired a thorough understanding of the OpenCV library and its significance in computer vision. We were able to test out a variety of AWS services, with some of them assisting us with facial analysis. We also came upon the Boto3 package, which allows you to use the Python SDK on AWS. Furthermore, we gained valuable experience working with the Sagemaker, a notebook instance that assisted us in running and deploying our code and model.
A distributed technique is used in the Doorbell Recognition system, allowing for quick defect identification and diagnostics. By doing so, we help to extend the benefits of smart doorbells to home users and contribute to making the world a smarter and safer place through technology.

References

[1] A. Ben Thabet and N. Ben Amor, ”Enhanced smart doorbell system based on face recognition,” 2015 16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2015, pp. 373-377, doi: 10.1109/STA.2015.7505106

[2] Viola, Paul, and Michael Jones. ”Robust real-time object detection.” International journal of computer vision 4.34- 47 (2001): 4

[3] R. Lienhart and J. Maydt, ”An extended set of Haar-like features for rapid object detection,” Proceedings. International Conference on Image Processing, 2002, pp. II, doi: 10.1109/ICIP.2002.1038171.

[4] Quadros, B.,Kadam, R., Saxena, K.,Shen, W., Kobsa, A.(2017). Dashbell: A Low-cost Smart Doorbell System for Home Use. UC Irvine. Retrieved from https://escholarship.org/uc/item/4v8290zh

[5] Chandrappa D.N., Akshay G., and Ravishankar M. Face detection using a boosted cascade of features using opencv. In K. R. Venugopal and L. M. Patnaik, editors, Wireless Networks and Computational Intelligence, pages 399–404, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[6] Mane, S., Shah, G. (2019), “Facial recognition, expression recognition, and gender identification” Data Management, Analytics and Innovation (pp. 275-290). Springer, Singapore

[7] V. Attina, G. Gibert, E. Vatikiotis-Bateson, D. Burnham Production of Mandarin lexical tones: auditory and visual components Auditory-Visual Speech Processing (2010), p. 4

[8] D.L. Blicher, R.L. Diehl, L.B. Cohen Effects of syllable duration on the perception of the Mandarin tone 2/tone 3 distinction: evidence of auditory enhancement

[9] D. Burnham, J. Reynolds, G. Vignali, S. Bollwerk, C. Jones Rigid vs non-rigid face and head motion in phone and tone perception INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association (2007), pp. 698-701

[10] Computer-vision analysis reveals facial movements made during Mandarin tone production align with pitch trajectories Garg, Saurabh ; Hamarneh, Ghassan ; Jongman, Allard ; Sereno, Joan A ; Wang, Yue Speech communication, 2019-10, Vol.113, p.47-62

[11] Computer Vision in Human-Computer Interaction ICCV 2005 Workshop on HCI, Beijing, China, October 21, 2005, Proceedings International Conference on Computer Vision Corporate Author; International Workshop on Human-Computer Interaction

(2005); Sebe, Nicu. editor.; Lew, Michael S. editor.; Huang, Thomas S. editor.

10 Appendix A.

1. What is the project?
   Ans: Our project is a smart doorbell recognition system that demonstrates how to monitor and control doors remotely.

2. What has already been done or claimed to have been done for this project?
   Ans: In this field lots of work has been done with different approaches. Most of the case were found that researchers are using raspberry-pi for the implementation part for higher accuracy.

3. How will I know that my program is working as intended? How can I verify the correct operation?
   Ans: Initially, we employed a pre-trained model to assist us in retrieving the facial detection findings. Later in the project we conducted real-time experiments by altering them, up-scaling and down-scaling them, changing the resolution, adjusting the similarity threshold, and so on, all in order to test the system's performance.

4. How can I demonstrate the concept of my advanced algorithm?
   Ans: We discovered a Haar-based cascade classifier for face detection, which is utilised to compare and analyse the face between the source and target images using AWS rekognition.

5. How will I demonstrate that my claimed algorithm is an improvement over standard naive method?
   Ans: We chose AWS services over Raspberry Pie and other processors since they are inexpensive and easily replaceable. That was ideal for our project, which aimed to make smart doorbells more affordable.

6. What data set will I use to evaluate the algorithm (training set and test set)?
   Ans: We used a Haar feature-based cascade classifier, which is a pre-trained model, for accurate object detection. It's a machine-learning technique that involves learning a cascade function from a large number of positive and negative images. The detecting stage use models based on HAAR over LBP as it was chosen because it provides better accuracy. All of the work has been completed, as evidenced by the OpenCV documentation in the cascade classifier.

7. How will this data set be documented (have the correct outcomes been pre established)?
   Ans:

8. What is the evaluation metric? How will I know that one algorithm is better than another? a single number.
   Ans: When we are looking for evaluating the machine learning model, we should go for F1-score which is very useful evaluation metric.

9. What is the evaluation experiment? How will the results be presented (graphs rather than pictures)?
   Ans: We previously did not use the AWS CLI, which resulted in a downside of requiring more time to complete the operation that was about 2 minutes. We witnessed how much easier it was to interface with the cloud and do tasks more quickly and efficiently when AWS CLI was added to our project. We saw a significant difference in completing the work in 7-8 seconds.

10. How will I optimize the algorithm's settings and parameters through multiple executions of the experiment?
    Ans: For example, we chose a celebrity and obtained various viewpoints and photographs from various films for each of them. The image was then compared using a similarity threshold, with a 90 percent threshold value chosen after much trial and error. We discovered that keeping the similarity threshold low resulted in poor results.