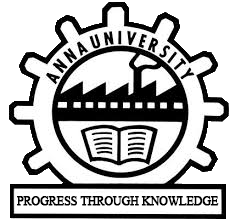
##### OBJ-RECO

###### ***Submitted by***

**S SASHANK (2016103060)**

**APARAJIT K R (2016103505)**

**COLLEGE OF ENGINEERING GUINDY**



**ANNA UNIVERSITY: CHENNAI 600 025**

**MAY-JUNE 2018**

**CHAPTER NO. TITLE PAGE NO.**

**LIST OF FIGURES**

**LIST OF TABLES**

**1. PROBLEM STATEMENT 5**

**2. INTRODUCTION 5**

**2.1 OBJECT DETECTION 5**

**2.2 OBJECT RECOGNITION 5**

**2.3 REVERSE IMAGE SEARCH 6**

**2.4 TEXT TO SPEECH 6**

**2.5 AUGMENTED REALITY 6 3. PROBLEM DESCRIPTION 7**

**4. WORKFLOW 7**

**5. IMPLEMENTATION 9**

**5.1 IMPLEMENTATION 9**

**5.1.1 TAKE PICTURE 9**

**5.1.2 SEND IMAGE TO CLOUD 10**

**5.1.3 SEND IMAGE URL TO GOOGLE 10**

**REVERSE IMAGE SEARCH**

**5.1.4 OXFORD DICTIONARY SEARCH 10**

**5.1.5 USER DEFINED TARGETS 11**

**5.1.6 TEXT TO SPEECH CONVERSION 13**

**5.2 TOOLS 13**

**6. ANALYSIS 14**

**7. CONCLUSION 15**

**8. REFERENCES 16**

**8.1 ONLINE REFERENCES 16**

**9. RESULTS AND SCREENSHOTS 17**

**LIST OF FIGURES**

**FIGURE NO. TITLE PAGE NO.**

**4.1 GENERAL OUTLINE OF WORK FLOW 7**

**4.2 GENERAL OUTLINE OF USER DEFINED 8**

**TARGETS**

**5.1 START CAMERA 9**

**5.2 SEND IMAGE TO CLOUD 10**

**9 SCREENSHOTS 17**

**LIST OF TABLES**

**TABLE No. TITLE PAGE NO**

**5.1 TABULATION OF TOOLS 13**

**6.1 TYPES OF ANALYSIS 14**

**1 PROBLEM STATEMENT:**

To capture an object and recognize what it is and then display the guess with the object serving as an user defined target.

**2.INTRODUCTION:**

**2.1 OBJECT DETECTION:**

**Object detection** is a computer technology related to [computer vision](https://en.wikipedia.org/wiki/Computer_vision) and [image processing](https://en.wikipedia.org/wiki/Image_processing) that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos. Well-researched domains of object detection include [face detection](https://en.wikipedia.org/wiki/Face_detection) and [pedestrian detection](https://en.wikipedia.org/wiki/Pedestrian_detection). Object detection has applications in many areas of computer vision, including [image retrieval](https://en.wikipedia.org/wiki/Image_retrieval) and [video surveillance](https://en.wikipedia.org/wiki/Video_surveillance).

**2.2OBJECT RECOGNITION:**

**Object recognition** – technology in the field of [computer vision](https://en.wikipedia.org/wiki/Computer_vision) for finding and identifying objects in an image or video sequence. Humans recognize

a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different view points, in many different sizes and scales or even when they are translated or rotated.

Objects can even be recognized when they are partially obstructed from view. This task is still a challenge for computer vision systems. Many approaches to the task have been implemented over multiple decades.

**2.3REVERSE IMAGE SEARCH:**

Google Reverse Search, available at **reverse.photos**, lets you search by images instead of keywords. Upload a picture from your desktop, tablet or mobile phone, and Google will show all the other web pages on the Internet that have similar images.

The [images are hosted](https://ctrlq.org/images/) anonymously on the Internet and cannot be discovered by other users.

**2.4TEXT TO SPEECH:**

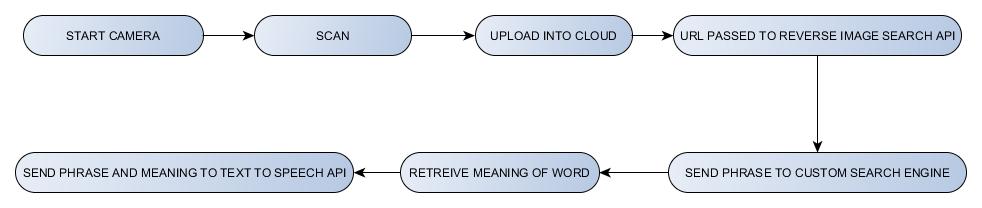
**Text-to-Speech** (**TTS**) refers to the ability of computers to read text aloud. A **TTS Engine** converts written text to a phonemic representation, then converts the phonemic representation to waveforms that can be output as sound. TTS engines with different languages, dialects and specialized vocabularies are available through third-party publishers.

**2.5AUGMENTED REALITY:**

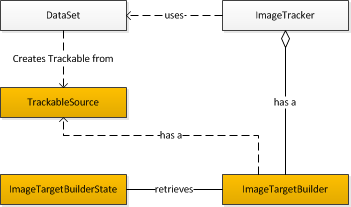
Augmented reality is an enhanced version of the real physical world through the use of visual elements, sound or other sensory stimuli. It is a growing trend among companies involved in mobile computing and business applications in particular. Amid the rise of [data collection](https://www.investopedia.com/terms/d/datamining.asp) and analysis, one of augmented reality’s primary goals is to highlight specific features of the physical world, increase understanding of those features and derive smart and accessible insight that can be applied to real-world applications.

**3.PROBLEM DESCRIPTION:**

The user can specify the area to be captured by boxing it or the entire scene can be captured as an image. The image is uploaded to a cloud service and then we pass the image url to google reverse search API. The API will return a key-phrase which we can then pass to the google custom search engine (or) Oxford dictionary API to retrieve the meaning. The image captured initially serves as an user defined target for the text and the text is super imposed on the object. Whenever the object is in focus the text will be displayed. If the object goes out of focus the text disappears. The meaning of the word will be read out,

**4.WORKFLOW:**

**4.1 GENERAL OUTLINE OF WORKFLOW**



**4.2** **GENERAL OUTLINE OF USER DEFINED TARGETS**

**5 IMPLEMENTATION**

**5.1.1 TAKE PICTURE:**

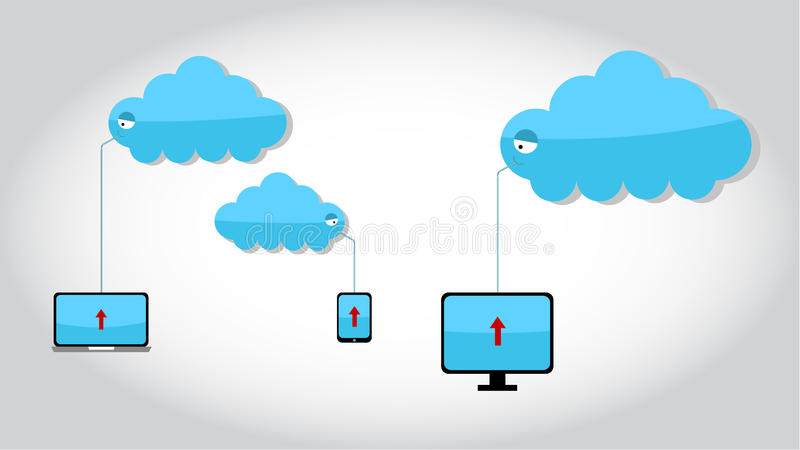
Draw a box around the object you want to capture. The coordinates and dimensions of the box is saved. The entire scene is captured and then the image is cropped as per the dimensions of the box previously saved.



**5.1 START CAMERA**

**5.1.2 UPLOAD IMAGE INTO CLOUD:**

Uploading image into cloudinary cloud service



**5.2** SENDING IMAGE TO CLOUD

**5.1.3.SENDING CLOUD URL TO GOOGLE REVERSE IMAGE SEARCH API:**

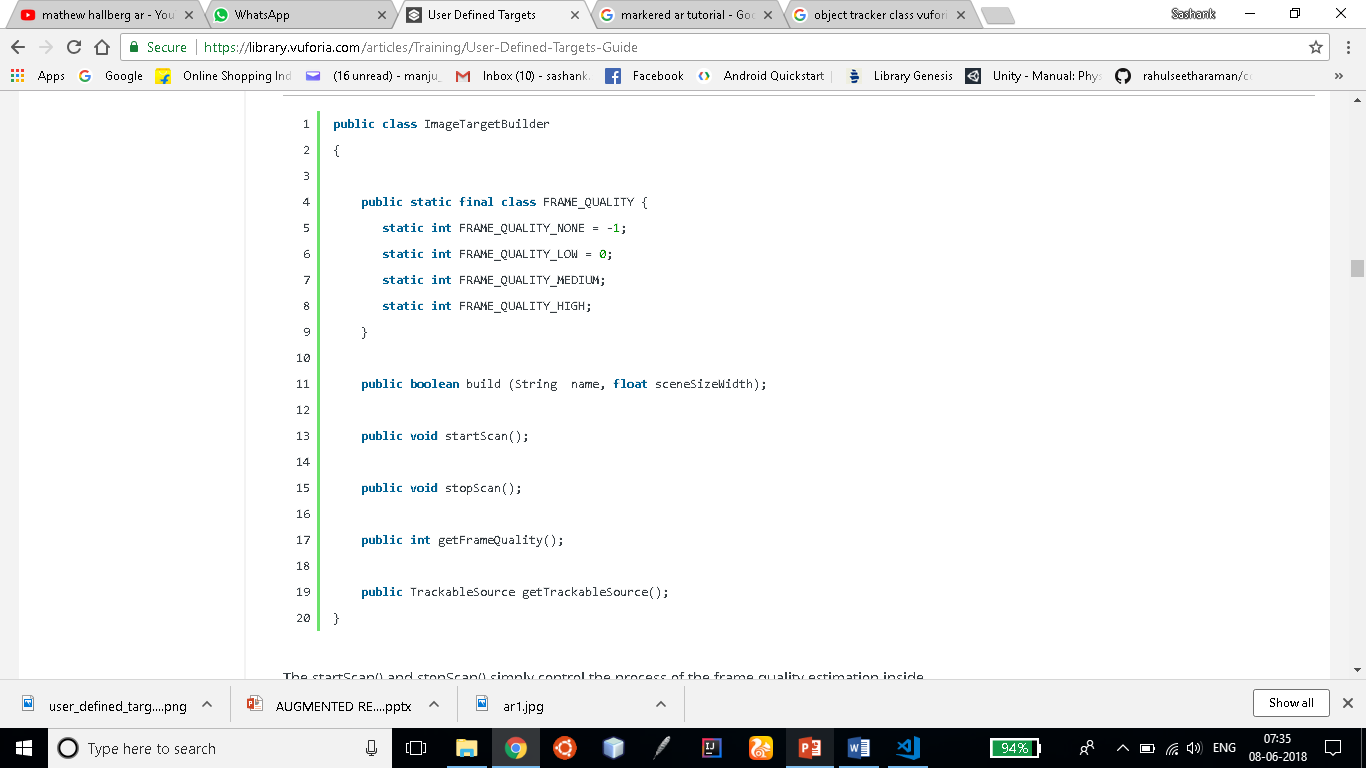
The URL of the image from the cloud service is sent to the reverse image search

Engine and the word is acquired.

**5.1.4.OXFORD DICTIONARY SEARCH:**

Send the word or phrase to oxford dictionary API to find the description (or) meaning of the object :

**5.1.5 USER DEFINED TARGETS:**



**5.3 Image Target Builder Class**

User-defined targets are Image Targets that are created at runtime from camera frames selected by the user. Just as Image Targets allow a developer to choose an image ahead of time that the app recognizes, User-Defined Targets allow an end user to pick an image at runtime. Thus, the user experiences AR “any time, anywhere” by selecting an image – like a photograph, book cover or poster – from his or her immediate environment without having to carry around pre-defined targets.

**5.1.5.1Working with User Defined Targets**

The process of capturing, building and tracking a User Defined Target is managed using the Vuforia API.

**5.1.5.2Attributes of an ideal User Defined Target,**

* Rich in detail, e.g., street scene, group of people, collages and mixtures of items, sport scenes
* Good contrast, i.e., bright and dark regions, well-lit
* No repetitive patterns, i.e., a grassy field, the façade of modern house with identical windows, a checkerboard
* Ease of availability, e.g., business cards, magazines, memos

**5.1.5.3Framing the Image**

The user should be instructed to capture the image when their device is parallel to the plane of the image's surface to minimize perspective distortion. This will provide a good rectilinear reference of the image to the ObjectTracker.

**5.1.5.4Using Extended Tracking with User Defined Targets**

Extended Tracking improves tracking robustness by using features of the environment surrounding the target. Target will be staged in a stable environment and won't be moved by the user. Moving the target or changing its environment when Extended Tracking is activated can corrupt tracking.

**5.1.5.5For user-defined targets, the application is responsible for the following tasks**:

* Starting the process of scanning a target – When the app begins to run
* Triggering the process for building the target – When the Frame around the object to be captured is released
* Adding the newly acquired target into a database for tracking

**5.1.6. TEXT TO SPEECH CONVERSION:**

SEND THE WORD AND ITS MEANING TO WATSON TEXT TO SPEECH API:

**5.2Tools:**

Since we use three APIs we will describe about them.

|  |  |  |
| --- | --- | --- |
| S.No | TITLE | DESCRIPTION |
| 1 | **Google reverse image search API** | This app lets you search by images using ( Google Reverse Search engine ) instead of keywords. |
| 2 | **Oxford Dictionary API** | The Oxford Dictionaries API allows easy access to our world-renowned dictionary content. We can pass the words to it and we can extract the meaning of those words. |
| 3 | **Google Text to speech API:** | We can say that this helps in speech synthesis by the name of the API.We recognise the word or definition using the other APIs and passing them to this API we can spell it out |
| 4 | **Cloudinary** | Cloudinary is the media management platform for web and mobile developers.  An end-to-end solution for all your image and video needs. |
| 5 | **Vuforia** | **Vuforia** is an Augmented Reality [Software Development Kit](https://en.wikipedia.org/wiki/Software_Development_Kit) (SDK) for mobile devices that enables the creation of [Augmented Reality](https://en.wikipedia.org/wiki/Augmented_Reality) applications. |
| 6 | **Unity** | Unity offers everything you need to build beautiful and engaging content, boost your productivity, and connect with your audience. |

5.1. TABULATION OF TOOLS

**6.ANALYSIS:**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **TITLE** | **DESCRIPTION** |
| **1** | *REQUIREMENTS* | * The user’s device needs to have a standard camera * The environment should have proper lighting conditions * User’s device should have internet services enabled * The minimum API level required is API 24 which is the Android version Nougat(7.0) |
| **2** | *USE CASES* | * We can use this app to help kids improve their vocabulary and help them things about all the things they see at their hands’ disposal * It helps visually-challenged people to know what is before using the speech which comes out from the device * Has a high educational value and this app can be implemented as a futuristic learning way in schools and children will enjoy their learning |
| **3** | *FURTHER IMPROVEMENTS* | * This app can be integrated for educational purposes * As of now, the app kind of struggles to differentiate between two similarly-structured objects placed close to each other which will be done in the future |

**6.1 TYPES OF ANALYSIS**

**7.Conclusion:**

A project well started is half solved. The better the clarity around what the team is attempting to fix, the more efficient they'll be in solving the problem, the solution will better 'fix' the issues, and the team can get back to executing the business versus fixing it.

We aimed at making a useful software that makes learning enjoyable.

This is not to say that the app is perfect and needs no improvement , we still feel the need to improve the project on its object recognition, say to recognize many objects at once.

We would like to improve the working of the app under low light and artificial lighting conditions

**8.REFERENCES:**

**8.1** **ONLINE REFERENCES:**

**Unity documentation:**

https://docs.unity3d.com/Manual/index.html

**Vuforia documentation:**

https://library.vuforia.com/api

**Wikipedia :**

https://en.wikipedia.org/wiki/Augmented\_reality

**Google reverse image search:**

http://in.pcmag.com/smartphones/96330/feature/how-to-do-a-reverse-image-search-from-your-phone

**Object recognition:**

https://virtualrealitypop.com/object-recognition-in-augmented-reality-8f7f17127a7a

**9.RESULTS AND SCREENSHOTS:**

