Title	I know who you are wearable assistance for human memory augmentation
Sub Title	
Author	Samarawickrame, Kalani(Kato, Akira) 加藤, 朗
Publisher	慶應義塾大学大学院メディアデザイン研究科
Publication year	2016
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2016年度メディアデザイン学 第487号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40001001-00002016-0487

Master's Thesis
Academic Year 2016

I Know Who You Are

Wearable Assistance for Human Memory Augmentation

Keio University
Graduate School of Media Design

Kalani Samarawickrame

A Master's Thesis
submitted to Keio University Graduate School of Media Design
in partial fulfillment of the requirements for the degree of
MASTER of Media Design

Kalani Samarawickrame

Thesis Committee:

Professor Akira Kato	(Supervisor)
Associate Professor Kai Kunze	(Co-supervisor)
Associate Professor Kouta Minamizawa	(Co-supervisor)

Abstract of Master's Thesis of Academic Year 2016

I Know Who You Are

Wearable Assistance for Human Memory Augmentation

Category: Science / Engineering

Summary

The research explores using Lifelogging for Human Memory Enhancement by Synthetic Recalling of past memories. The system is developed after thorough investigation on two research questions; Selective Recall and Unobtrusive Feedback to the user which were identified as limitations in existing systems. The Voice Activity Detection Based System introduced in the research reduced the data capture by over 80% and increased the no of usable images. The mobile based implementation was accepted by 60% of the users during Usability Test compared to the heavy displays and extra wearable devices the existing research has introduced. It was identified presenting results in the notification bar is a better solution for further eliminating obtrusiveness but user's preference on presentation method would depend on their background. While it was revealed that under standard lighting conditions and slight changes in orientation of the device, the dominant factor that determines if an image can be recognized is detection of the face and it's landmark positions. Testing proved recognition system adheres to wearing and removal of spectacles.

Keywords:

Lifelogging, Human Memory Augmentation, Face Recognition, Memory Cues, Wearable System, Selective Capture, Voice Activity Detection

Keio University Graduate School of Media Design

Kalani Samarawickrame

Acknowledgements

I would like to thank my main supervisor Prof. Akira Kato, Sub Supervisor Associate Prof. Kai Kunze, Prof. Hideki Sunhara for the valuable guidance, support and patience throughout the research and my second sub supervisor Associate Prof. Kouta Minamizawa for adding new thoughts on the research direction. I would also like to thank Dr. Suresh Hettiarachchi from The University of Tokyo who is my former lecturer in Undergraduate studies for guiding me to achieve the MEXT Scholarship without which I would not have been able to join KMD for my Masters and my Undergraduate Research Supervisor Dr.Ranga Rodrigo from University of Moratuwa for inspiring me to become a researcher. I would also like to thank my friend Lahiru Lakmal for giving me valuable technical insights and Ethan Pitt for helping me with testing the application during busy times. Last but not least my three mothers back home, my mother, grandmother and aunt, the three pillars in my life for all the love and care they gave me since I was born to this date and helping me achieve my targets. I would not have come this far without them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Background

Lifelogging has received a high importance with the technological advancements. However, the first mentioning of Lifelogging dates back to 1945. Research paper "As We May Think"by Vannevar Bush [1] makes the very first prediction about a futurist lifelogging device, introducing Memex which is a "hypermedia device in which an individual stores all his books, records and communications which is mechanized so that it may be consulted with exceeding speed an flexibility."Bush described it is an "enlarged intimate supplement to one's memory" marking the first hint at what we now refer as Lifelogging.

In 1994, Lamming and Flynn [2] discusses about using mobile and ubiquitous computing for everyday memory problems such as finding lost document, operating new device or remembering people. They used a portable device called ParcTab that can be clipped to the belt or carried in a pocket which can communicate with other ParcTabs and connect with other user defined devices. The user can provide it with a list of the devices from which data are to be collected. As the user encounters and interacts with each of these devices, Forget-me-not automatically gathers up information describing the devices name and location. Then it appends a time-stamp and stores the records away in its own memory. The experimental setup was carried inside a small four-storey building where all the devices were connected with WiFi. This is also marks a trivial standpoint in the road to today's Lifelogging.

Although Lifelogging is interpreted as a method for replacing memory it is a mechanism to provide cues and work in synergy with organic memory. Total capture or capturing every single moment in life will be hard to be accepted by the general public as of yet. But it is interesting to see how a scaled down version of Lifelogging for certain application area can be used.

Have you ever experienced a situation where you meet a person and you know you have met before but you are unable to recall when and where you have met. Sometimes memory is unable to provide us with necessary cues to remember a past situation better. This is called Episodic Memory Impairment. Episodic memory is the memory of specific experiences that you can replay in your head. Lifelogging is a viable solution to address this issue by recording meetings with people and retrieving memory cues to remember them better.

1.2. Motivation

My initial interest in Lifelogging started with the habit of reminiscing. When I visit a certain place or when I remember an important event I would like to reminisce and live in that past moment. Specially when I'm studying abroad I would always do this and always thought it would be interesting to capture the moments rather than having them only in memory. When I go to dancing class after a long period of absence I would reminisce my past moments there and think it would be nice if I can visually relive the moment with the use of technology instead of having them in memory. Reminiscing is one aspect of Lifelogging and when researching I came across other uses such as Recollecting, Reflecting, Retrieving and Remembering intentions and read more about these four categories. When further reading recollection based Memory Augmentation was found to be more in line with day to day use. Being a foreign students I always experienced meeting new people and forgetting their names or where I met them before. The idea for developing A memory assistance system for remembering people was commended by the faculty at the Interim Presentation and the audience commented they require a system like this.

1.3. Background of the Research Questions

Lifelogging can be categorized in to two broad categories Total Capture and Selective Capture. According to Bell and Gammel in their 2009 book on Total Recall [3] it is discussed how every bit of information we capture everyday can be saved for Total Recall later. However total capture is not technically feasible since it requires different kind of sensors and capture devices embedded in the devices. It also requires vast amount of memory storage and careful design of processing and retrieving information for the users without overwhelming them. This is further proven by Sellen and Whittaker in [4]. They mention "capturing vast arrays of data might overwhelm end users maintaining and retrieving valuable information from large archives; it also ignores the burden huge amounts of data impose on system designers and developers."Hence it is clear that selective capturing is more useful for day to day use devices.

Recent lifelogging research has been primarily focused on developing new technology to support the capture of everyday events and experiences. Moreover Kalnikaite and Whittaker [5] also agrees that less attention has been spent on understanding what types of lifelogging tools have been built so far and how these different tools help reconstruct memories of the past.

1.3.1 Definition of Research Questions

As explained in Chapter 2 it can be observed users are reluctant to accept the Presentation method in current reserach based on Head Mounted Display. Hence a novel presentation method will be explored in a device being used in a day to day environment. Analyzing the existing literature explained in 2.2 two important factors were identified which will be explored during the course of this research.

- Effectiveness of Selective Capturing vs Total Capturing and Voice Activity Detection as a Trigger
- Presentation of memory cues to the user in an unobtrusive manner

1.3.2 Research Method

Generic Framework of a similar system would be

- Capturing the Lifelog Data
- Process
- Presentation

In the Capture Module it will be looked at Capturing Lifelog data. It will be evaluated the effectiveness of Total Capture vs Selective Capture for situation specific scenarios like the current research. This module will look at using available Lifeogging devices and their effectiveness in the proposed research. Based on the performance it will look at selective capture for improvement. A novel trigger based on Voice Activity Detection will be introduced which will then be compared with the existing methods.

In the Processing Module focus will be given for analyzing the captured images for Face Recognition. In this module new faces will be used for training the Face Recognition Engine which will then be used for recognizing a person against a test image.

Presentation Module is focused on presenting the processed data to the user in an unobtrusive manner. This module will look at what are the most important memory cues and how they should be presented eg. Audio, Textual, Pictorial depiction. Presentation module will have two iterations of implementation. The initial prototype will be implemented and based on user feedback, second version will be implemented as an improved version.

1.4. Thesis Structure

The three modules mentioned in Section 1.3.2 will be carefully examined in Chapter 2 Related Works. Based on the existing research and commercial applications the limitations will be identified in the current approaches. Then a new approach will be designed in Chapter 3 and a prototype will be implemented in Chapter 4 which will be evaluated in Chapter 5 in order to find how well the system adheres to solving the research questions. The system will be evaluated quantitatively and qualitatively. Finally Chapter 6 will summarize the research findings and conclude the thesis.

Chapter 2

Related Works

2.1. Lifelogging Overview

2.1.1 Human Memory Overview

As humans we highly rely on our memory for our day to day activities. But due to it's limitations we need external support such as photographs to remember events such as holidays, birthdays, weddings and cherish the times spent with our loved ones. For day to day activities we use Calendars, Alarms, Sticky Notes, etc. Technology has always created new ways to remember things better. Now Lifelogging is paving it's way as the latest method for memory enhancement.

Human memory can be categorized into two main areas. Episodic Memory relates to the memory of specific experiences that you are able to recall and semantic memory is memory of facts about the world. According to the Human Memory Website [6] Episodic memory represents our memory of experiences and specific events in time in a serial form, from which we can reconstruct the actual events that took place at any given point in our lives. It is the memory of autobiographical events (times, places, associated emotions and other contextual knowledge) that can be explicitly stated. Individuals tend to see themselves as actors in these events, and the emotional charge and the entire context surrounding an event is usually part of the memory, not just the bare facts of the event itself. Surprenant and Neath [7] states that Episodic Memory is often thought to be synonymous with conscious awareness and it's evolutionary and ontologically the highest form

of memory.

Semantic memory, on the other hand, is a more structured record of facts, meanings, concepts and knowledge about the external world that we have acquired. It refers to general factual knowledge, shared with others and independent of personal experience and of the spatial/temporal context in which it was acquired. According to Surprenant and Neath [7] a special feature of the Semantic Memory is the combination of conscious knowledge about the content but lack of conscious awareness about it's learning episode. For example those who know the Capital of France is Paris would not be able to remember the situation where they learnt it.

Humans tend to associate memories with cues. Surprenant and Neath [7] also states that human memory is cue driven and these cues can be verbal, images, non verbal sounds, emotions, mood, locations and study environment. These cues help them mentally relive a past experience. For example a meeting with a friend would be cued as a meeting an old friend on a rainy day in a coffee shop. In this case, rainy day and coffee shop are the memory cues used to retrieve this memory.

2.1.2 Lifelogging for Human Memory Augmentation

Sellen et al [8] defines "The vision of Lifelogging"as that technology will allow us to capture everything that ever happened to us, to record every event we ever experienced and to save every bit of information we have ever touched.

Lifelogging technology can help capture memory cues for augmenting human memory. It is important to capture cues that help retrieve memory. These cues can vary depending on situations. But images is the most vital cue in recollecting a past memory. It is also said the more you try to remember you would remember something better. According to Lee and Dey [9] Specially with relation to patients with Alzheimer's disease it is said that as the cueing process continues with the caregiver providing cues until the memory is recalled at an adequate level of detail the patients find it easy to recall the information themselves. Wilson et al [10] also confirms that Lifelogging is a viable mechanism for increasing patient's memory by stating there is clinical evidence that engaging in such cognitively stimulating mental exercise can slow the progression of cognitive decline.

Lifelogging creates a vast volume of data. But data only is not helpful for the

users. If this data is processed to provide some analysis for the users it would help them better.

According to Kalnikaite et al [11] their research also confirms that we lack theoretical insights into exactly how such tools might support everyday memory processes.

2.2. Selective vs Total Capture

Total Capture is capturing every moment in life, everything we touch, every object we interact with, every meal we have, every person we meet and talk, every website we browse, every email, every telephone call, every place we visit all these information will be recorded in Total Capture. Gordon Bell [3] and Steve Mann [12] have experimented the concept of Total Capture and lived up to the true meaning of the concept.

But total capture comes with a cost to pay. It overwhelms users with vast amount of data while storage will also has a limitation. With the advancement of technology companies are focusing more on increasing the storage capacity and look and feel of the devices but less attention has been given for analyzing the information and proving useful information. According to [5] Lifelogging research has so far focused on technical innovation rather than understanding when, why or how these tools are used in practice. This is further confirmed by Vaiva and Steve in [5] "Recent lifelogging research has been primarily focused on developing new technology to support the capture of everyday events and experiences. Less attention has been spent on understanding what types of lifelogging tools have been built so far and how these different tools help reconstruct memories of the past." Selective capture can be situation specific, carefully designed to capture data only related to a certain event or a situation which can be processed and provide an analysis for the user in a more useful manner than just recording every moment.

However, existing Lifelogging systems fail to describe explicit benefits to the users. Sellen and Whittaker [4] has identified 5 activities called "The 5Rs" where human memory augmentation systems can support and provide potential benefit for memory. These activities are described below.

- Recollecting. Technology could help us mentally re-live specific life experiences, thinking back in detail to past personal experiences or our episodic memories. Remembering aspects of a past experience can serve many practical purposes; examples including locating lost physical objects by mentally retracing our steps, recollecting faces and names by recalling when and where we met someone, or remembering the details of what was discussed in a meeting or remembering what was taught at a lesson.

- Reminiscing. As a special case of recollection, lifelogs could also help users re-live past experiences for emotional or sentimental reasons. This can be done by individuals or socially in groups; examples are watching home videos and flipping through photo albums with friends and family or viewing photographs of someone far away and reliving the moments spent with them. This could also be interesting for pet lovers who have lost their pets to go down the memory lane to the good times they spent with their pets.

- Retrieving. Lifelogs promise to help us retrieve specific digital information we've encountered over the years (such as documents, email, and Web pages). Retrieval can include elements of recollection; for example, retrieving a document might require remembering the details of when we wrote it, when we last used it, or where we put it. Alternatively, retrieval might depend on inferential reasoning (such as trying to deduce keywords in a document or thinking about the document's other likely properties, like type and size).

- Reflecting. Lifelogs might support a more abstract representation of personal data to facilitate reflection on, and reviewing of, past experience. Reflection might include examining patterns of past experiences (such as about ones behavior over time). Such patterns might provide useful information about our general level of physical activity or emotional states in different situations, allowing us to relate it to other data about, say, our health. Alternatively, reflection might involve looking at ones past experiences from different angles or perspectives. Here, the value is not in re-living past events (as in recollecting) but in seeing things anew and framing the past differently. The value is less about memory per se than it is about learning and

self-identity.

- Remembering intentions. Another type of activity concerns remembering prospective events in ones life (prospective memory), as opposed to the things that have happened in the past. Our everyday activities require that we constantly defer actions and plan future activities; examples include remembering to run errands, take medication, and show up for appointments.

During this research it is intended to look at Recollection Lifelogging tools which would help users remember the faces of people they meet. It is a common to forget names and faces of people specially if you are

- Visiting back a foreign country after being away for vacation
- Meeting people you have met at previous conferences or alumni parties
- Meeting lot of new students in the beginning of a new semester

Unfortunately this can lead to an embarrassing situation and have an adverse effect on human relationships. However if you can at least remember where you met the person before this would help you initiate a conversation and find out the name during the conversation.

An interesting research in the context of Face Recognition using Wearable Systems is done by Sreekar et al [13]. In this system they try to assist visually impaired people to recognize the person in front of them. An analog CCD camera embedded in a glass was used for acquiring the video for the assistive device. Since the camera provides an analog video output, they have used an Adaptec video digitizer to convert the composite video into a digital video format and transmit the AVI stream over a USB cable that can be used inside a computer for analysis. A laptop was used to execute the face recognition algorithm. When the system produces a guess for the person in the video frame, the user WAS notified with an audio signal. Microsoft Speech Engine was used to convert the name of the identified individual from text to speech. This was fed to the headphones that the user wears. One aspect that can be improved in this system is identifying the best moment to perform face recognition instead of performing Face Detection on every video frame which would save the computational cost drastically.

Looking at the existing research by Iwamura et al [14] their proposed system consists of a wearable camera and a head-mounted display installed in a laptop computer. The wearable camera is set up to capture the frontal view of the user. The largest face in the captured image is detected by a face detection method. This starts both indexing and retrieval processes. The researchers are taking a video when a face is detected and tracks the face until it disappears and the frames captured in between this period are stored as a video. It can be seen the researchers are using selective lifelogging based on face detection rather than storing unrelated data but during the evaluation phase the user acceptance was low due to the fact having to wear a head mounted display. This is the closet research implementation available. Nevertheless, it can also be argued that storing an image vs the video would help low storage space and speed up the accuracy than processing a video. Several example devices with sensor based triggers in the current market are described in the following subsections.

2.2.1 Sensecam

Sensecam is a small wearable camera developed by Microsoft Research which can take photographs with no user intervention. It has a fisheye lense and a number of sensors built into it. According to [15] These sensors are moni-tored by the cameras microprocessor, and changes in sensor readings can be used to automatically trigger a photograph to be taken. For example, a significant change in light level or the detection of a person in front of the camera can be used as triggers. In addition, by default an internal timer is used to automatically trigger a photograph every 30 seconds. SenseCam also has a manual trigger button that lets the wearer take pictures in the more traditional fashion, a privacy button which causes the camera to stop taking photos for four minutes at a time, and an on-off button. Three LED lights and an internal sounder are used to give the wearer feedback. The device is called SenseCam because two of the main components of its operation are sensing the environment and using its camera to record images. The images captured and stored by SenseCam during the course of an event may subsequently be uploaded onto a desktop or laptop computer via a USB connection.

Source: SeniorTechDaily Website [16]

Figure 2.1: Sensecam

Sensecam weights 94g and has 65×70×017mm dimensions. This technology is now licensed to UK-based Vicon Motion Systems, so that the company can manufacture the device as a memory aid. In 2010 it was sold at US$775 but the current pricing information is not available.

2.2.2 Autographer

Autographer is also based on sensecam with 8GB of memory and embedded with 6 sensors. Magnetomer is the Autographer's compass which determines which direction the camera is facing. Color Sensor checks the brightness and light and adjust the image accordingly. It also has a motion sensor which can determine moving objects. It also has an accelerometer which can determine how fast it is moving. It's Thermometer determines the temerature of the ambient and last but not lesat it has a GPS sensor which can pinpoint autopgapher's location on earth.

Source: The Verge Website [17]

Figure 2.2: Autographer

Autographer weights 58g with width 37.4mm (with side buttons), length 90mm; 95.5mm (with lanyard ring) and thickness 22.93mm (with clip and lens). Autographer can be bought for around US$180 from Amazon United States website.

2.2.3 Narrative Clip

This is also another lifelogging camera in the market with the ability to take first person view images every thirty seconds. It has 8GB of in built memory with Accelerometer, Magnetometer, GPS, Gyroscope and 5MP pixel images.

Source: Slashgear Website [18]

Figure 2.3: Narrative Clip

Narrative Clip 2, the latest release weights 19g and has dimensions 36×36×12mm. This is priced at US$199.

2.2.4 Evaluation of the devices discussed

The above devices are technologically advanced with increased memory storage and different sensors. But however when they are practically used, it is cumbersome to browse through still images and review them at the end of a tiring day. Later on theses images may even be forgotten by the users. Plus after recording around 7000 images with Autographer only about 30 images were of good quality which can be used for any kind of analyzing. Lots of images contained blurred images and images taken inside private places such as bathrooms were of privacy concern. Hence we can see these sensors are good but when designing a situation specific lifelogging approach it would require different unexplored triggers which are currently unavailable.

2.3. Presentation of Data

Not only capturing but presentation also requires careful design. While Lifelogging allows automatic capture of a human's day, this still does not resolve the problem of trying to augment human memory. For this raw data to be useful, even after it has been well processed and tagged, there must be a useful and intuitive way to allow the user to access the potentially very small part of the whole collection that is needed to jog the memory of a particular event. As described by Harvey et al [19] we can clearly see that careful design is important in the presentation module as well.

Looking at the existing research one of the the most interesting work was created by Steve Mann in [12] that does Face Recognition on the fly. The user will be presented matched images from the database and once the wearer confirms the match the system inserts a virtual image into the wearer's field of view creating an illusion that the person is wearing a name tag.

One of the important research that focuses on the Presentation of information is research by Lee and Dey [20] on identifying important memory cues to present to the user who is suffering from Alzheimer's disease. Their system has three main

modules Capture, Selection and Review. The proof-of-concept capture system records photographs, ambient voices and sounds, and location information from the users experiences using three devices the Microsoft SenseCam, an offthe-shelf voice recorder, and a GPS logger. Their approach leverages automated content and context analysis (e.g., face detection with photos for people based experiences and localization using GPS for location-based experiences) to extract potentially helpful memory cues from the lifelog and filter it down to a more manageable size. Since this is specially targeting a certain user group "Alzheimer patients" they also leveraged the expertise of the caregiver of the person with EMI (Episodic Memory Impairment) to hand select meaningful memory cues from the filtered content to present to the person with EMI. The caregiver constructed a slide show narrative of photos, sounds, and annotations. Final application mimics a picture frame that the person with memory impairment can pick up and use without having to worry about operating a computer or bothering their caregiver. MemExerciser reveals each photo, the associated recorded sounds, and the caregivers annotations one at a time so that it follows a similar cueing process used by the caregiver in the absence of any technological support. The system not only was successful at capturing a personal experience, but it also provided a way to structure the information in such a way that the person with EMI could best utilize it to reminisce about recent events and reduce the burden on their caregiver.

2.4. Summary

Scrutinizing the key points of Literature it can be deduced it is important to analyse the captured data and provide some valuable feedback to the user rather than presenting vast volume of data. A feasible approach for this is using Lifelogging for Selective Capture which would not overwhelm the user with data. Nonewithstanding this, it was also identified that trigger based capture is more useful to achieve this. The Design Chapter would discuss how to use the key findings of the Literature Review to design the proposed system.

Chapter 3

Design Approach

3.1. Addressing the Research Questions

Two main research questions were identified in the Introduction chapter,

- Using Trigger based Selective Capturing vs Total Capture
- Presentation of memory cues to the user in an unobtrusive manner

This chapter will look at how to design the system to evaluate the two research questions. The three main modules Capture, Analysis and Presentation will be analysed and the structure of implementation will be designed.

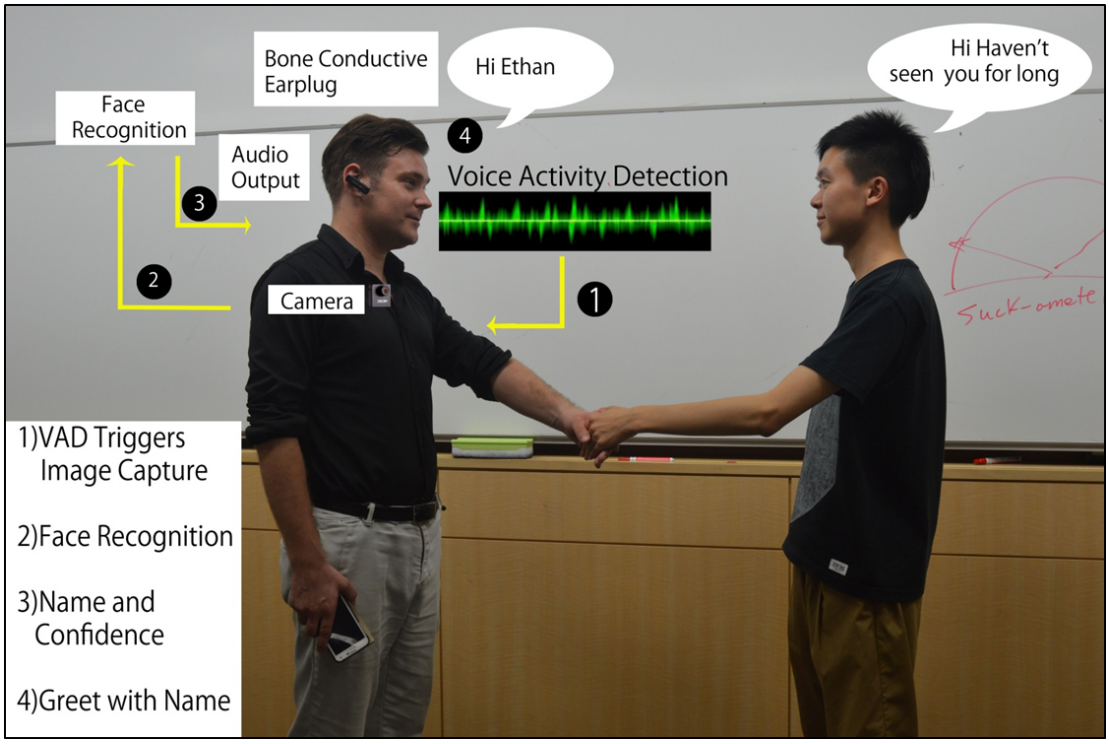


Figure 3.1: Expected User Experience

Figure 3.1 demonstrates the ultimate user experience expected to generate by a fully functional system. As displayed, someone greets you by the name but you are unable to remember the name. You start talking and the trigger based system triggers the camera to capture an image which will then be sent to the Face Recognition Engine. The system will give audio output and let you know the name of the person in front of you and how confident it is this person. Mobile phone notification, Display on smart watch, Display on smart glasses are couple of options that can be used in a fully functional system depending on the user's preference and background. Initial prototype implementation will focus on presenting audio output and visual output on mobile phone.

3.2. Designing Trigger based Selective Capture

Conventional lifelog cameras are configured to capture images on a timed basis. In some devices, for example, an image is captured every 30 seconds. If left

to take pictures over the course of several hours or an entire day, the lifelog camera could take hundreds or thousands of pictures at the predetermined time intervals. Under this approach, many of the images captured by conventional lifelog cameras are not very interesting. Therefore, a lifelog camera's memory may become filled with photos that are not of interest to the user or any potential benefit for memory. More compelling moments may occur rather quickly and between the timed increments for taking a photo. However, it is difficult to determine when those compelling moments are occurring.

Figure 3.2 illustrates a sample set of images captured by Autographer over one hour.

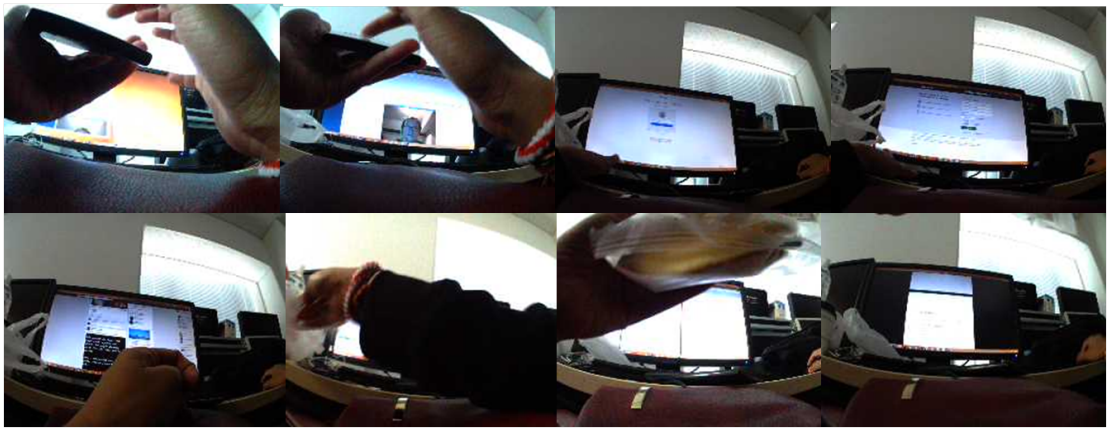


Figure 3.2: Sample Images Captured by Autographer

As it can be seen in above image 3.2 it can be observed these images are not effective in Face Recognition. The existing Lifelogging devices fail to capture good images that can be used for Face Recognition because they are captured on a timely basis. Hence they tend to capture more unuseful data.

3.2.1 Voice Activity Detection As the Trigger

Talking to someone can be considered as the starting point of a communication. Hence Voice Activity Detection denotes that the user is talking to someone. This could be an interesting moment to capture an image vs storing large number of images which are not useful to the user. The designed approach is expected

be more user oriented than existing methods. Figure 3.3 describes the steps in Capturing an Image based on Voice Activity Detection as the trigger.

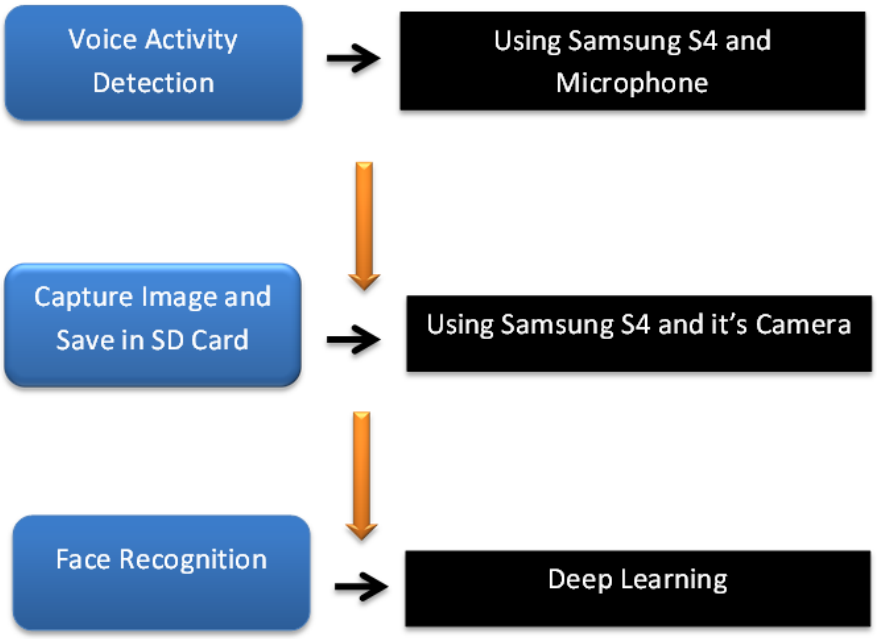


Figure 3.3: Flow Diagram of the Proposed Capture Module

3.3. Designing analysis of data for Memory Augmentation

A typical scenario where Lifelogging can be used for enhancing memory, and a device that can be used in a day to day environment was designed as explained in Introduction Chapter. It was identified in the Literature Review that synthetic memory can be volatile despite age. Based on personal experience having faced situations where forgetting names and where I met the new Japanese friends I met in Japan and also based on existing Literature it was decided to design an application which would remind you of the person infront of you. This involves capturing the images of people someone interact with and performing Face Recognition on the images to inform the user when they meet for the second time.

Any Face Recognition System has a common flow of identification. The pic-

torial depiction of this process is given in Figure 3.4.

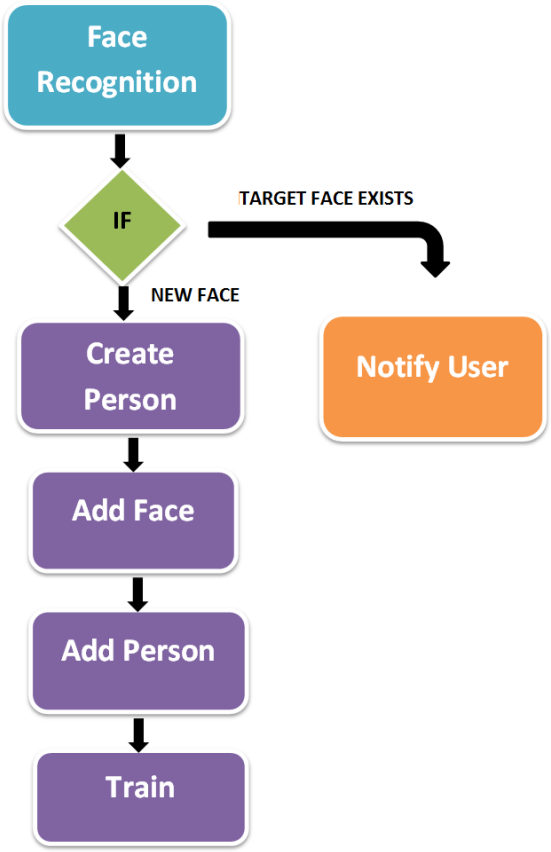


Figure 3.4: Flow Diagram of the Recognition Process

3.4. Designing Presentation of Memory Cues Unobtrusively

Richard Feynman says "You can know the name of that bird in all the languages of the world, but when youre finished, youll know absolutely nothing whatever about the bird. Youll only know about humans in different places, and what they call the bird. So lets look at the bird and see what its doing thats what counts." This statement by Richard Feynman clearly distinguishes the difference between knowing the name of something and knowing something. But in our

scenario it's not easy to implement scene analysis or activity recognition. Hence this can be suggested as a future implementation which can be made feasible with the advancement of technology.

As discussed in the Literature Review it is important to identify the important memory cues to remember a person better. For this purpose we need to look at what are the features we remember first when we try to think of someone. The number one fact that comes to our mind is the face, proceeded by any special features about the face such as hair color, length, eyes, shape of the face. Next we remember our past encounters such as location and ambient, if we met the person at a birthday party, conference, office meeting, on-site travel to a client location in another country etc.

Hence we can understand that just telling the name is not enough to remember a person. Presenting the name is influencing the memory but more information about that person such as the conference you met will be important to start a conversation and really remember the person.

Another fact to consider is that results can sometimes be misleading. Therefore it is important to find a mechanism to tell the user an indication how true the system could be. Hence the current system is planned to give a confidence level to the user of the recognition process's prediction accuracy.

Chapter 4

Implementation

4.1. Overall Architecture

The proposed system can be decomposed into three main sub sections; Capture, Processing Data and Presentation.

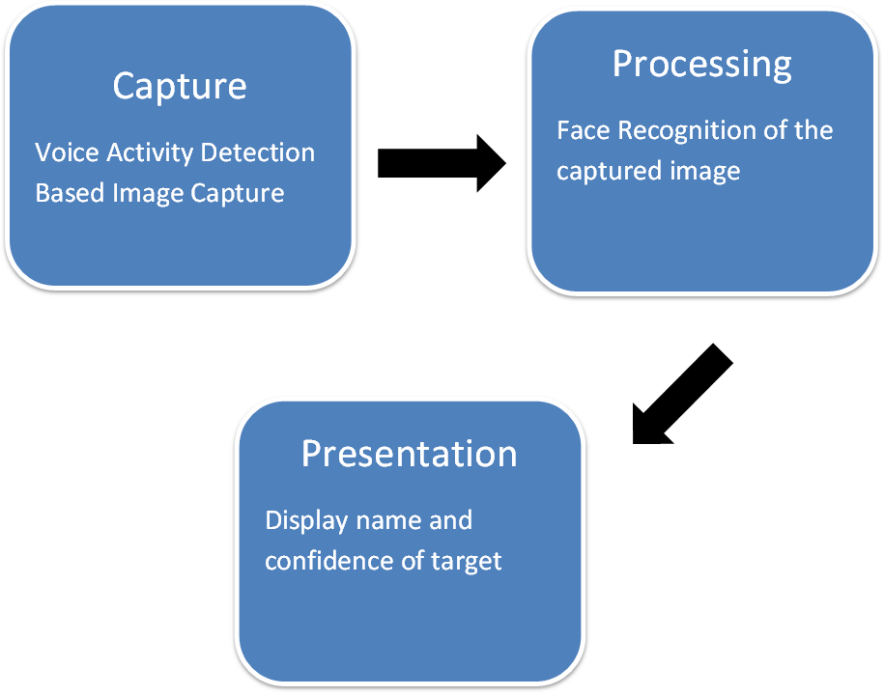


Figure 4.1: High Level Architecture of the proposed system

Table4.1 explains the Tools and Technologies used for developing the three modules.

Table 4.1: Tools and Technologies.

	Tools	Technologies
Capture Module	Samsung s4 (GT-19505) Samsung s4 Camera Module Sony Headset with Mic	Eclipse Mars Android API 4.0.3 Java 1.8
Process Module	Face++ API	Android API 4.0.3 FaceppSDK.jar
Presentation Module	Samsung s4(GT-19505)	Eclipse Mars Android API 4.0.3 Java 1.8

4.2. Capture Module

Figure 4.2 shows the low level decomposition of the Capture Module. The first step is detecting voice using Android Speech Intent and capturing the image once the voice activity detection trigger is fired. Then the captured image needs to be rotated and resized in order to be able to be used by the Processing Module. While capturing the image, some context information is also acquired such as accelerometer data, magnetometer data and illumination data.

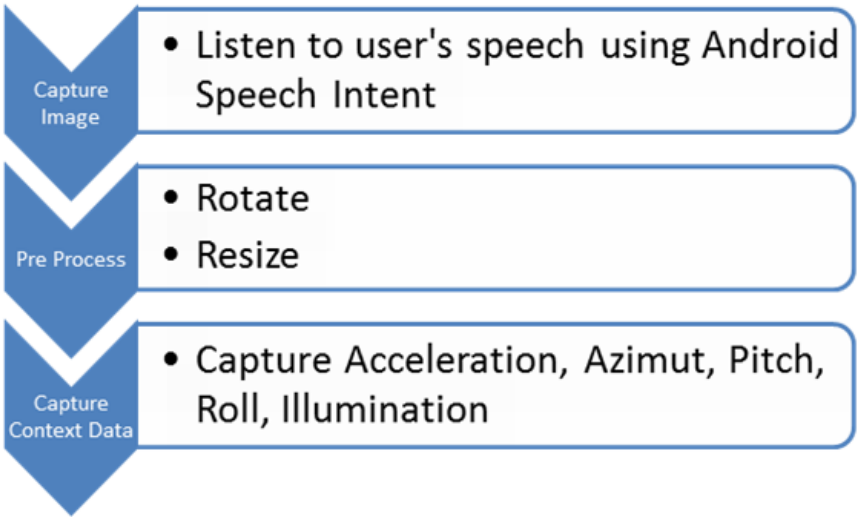


Figure 4.2: Low Level Decomposition of the Capture Module

4.2.1 Voice Activity Detection Based Capture

A typical Voice Activity Detection involves a comprehensive algorithmic implementation as depicted in Figure 4.3.

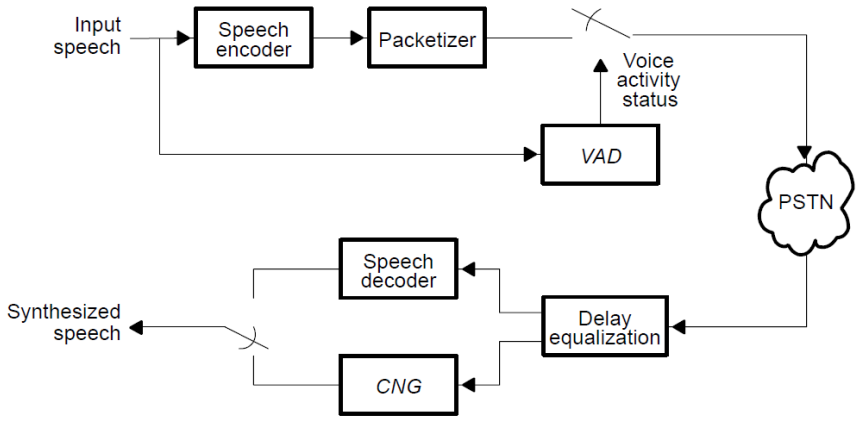


Figure 4.3: Typical VAD Integration by Sprit Corp [21]

According to Spirit Corp [21] the VAD algorithm developed by their company for commercial purposes detects the presence of speech in the signal. It has special adaptive algorithm to automatically adjust to the level of the noise in

the signal, in order to provide robust operation even in the noisy speech. It has many user configurable parameters, allowing the algorithm to optimally tune itself for a specific application. VAD also outputs several coefficients (up to 10) that characterize spectral envelope of the noise (when no speech is detected), so that the regenerated noise would be similar to the original noise. Hence it's clear that developing an algorithm in this capacity is a research project on it's own.

4.2.2 Android Speech Recognition Intent for Detecting Voice

Due to the reason implementing a Voice Activity Detection System itself is a separate research on it's own it was decided to use Android Speech Recognition Intent to identify when the user starts talking. If Android Speech Recognition Intent recognizes words in the speech this moment is being considered as the point when the user starts talking. This has the advantage of misinterpreting background noise as speech. This came with 1 to 3 seconds of delay for the Intent to send an output. The delay was based on the surrounding noise in the environment making it difficult for the intent to identify the words and loudness of voice not being enough to be detected by the microphone.

4.2.3 Capturing Context Information

Android mobile phones are equipped with sensors such as Accelerator, Magnetometer and Ambient light sensor. Samsung S4 was used for testing purposes and Table4.2 describes the Sensor, Vendor and version number of the sensor used along with the unit these values were measured.

Table 4.2: Sensors used to acquire context information.

Sensor	Vendor	Version	Unit Measured
Acceleration Sensor	Android Open Source Project (AOSP)	V3	SI units (m/s2)
YAS532 Magnetic Sensor	Yamaha Corporation	V1	micro-Tesla (uT)
CM3323 RGB Sensor	Capella Microsystems, Inc.	V1	SI lux units

4.2.4 Image Pre Processioning

When the camera application was first developed Android Layout Page of the camera was fixed to Vertical to get a portrait image without having to rotate the mobile phone.

```
<LinearLayout xmlns:android="http://schemas.android.com/apk/res/android"
android:orientation="vertical">
```

Figure 4.4: Configuring Android Layout to be Vertical

But the Surface Preview was displaying as landscape. This was identified as a device-specific issue that mostly affects Samsung devices. The Google Android Developers included a setDisplayOrientation call in API 8 to work around the issue.

```
mCamera.setDisplayOrientation(90);
```

Figure 4.5: Changing Surface View to be Portrait

This set the Camera Surface View to portrait but when the images were captured they were not portrait although the mobile view was fixed to be portrait.

The issue was caused by camera orientation. When capturing an image the orientation behaves differently because OEMs do not adhere to the standard. HTC phones do things one way, Samsung phones do it a different way, the Nexus line seems to adhere no matter which vendor, CM7 based ROMs follow the standard no matter which hardware but since the application was tested in Samsung it needed a workaround.

The pseudo code of the preprocessing is listed below and full coding will be given in Appendix A and B.

```
//STEP 1: Get rotation degrees
 Camera.CameraInfo info = new Camera.CameraInfo();
 Camera.getCameraInfo(Camera.CameraInfo.CAMERA_FACING_BACK, info);
 int rotation = this.getWindowManager().getDefaultDisplay().getRotation();
 int degrees = 0;
 switch (rotation)
 {
  case Surface.ROTATION_0:
   degrees = 0;
   break; //Natural orientation

  case Surface.ROTATION_90:
   degrees = 90;
   break; //Landscape left

  case Surface.ROTATION_180:
   degrees = 180;
   break;//Upside down

  case Surface.ROTATION_270:
   degrees = 270;
   break;//Landscape right
  }
int rotate = (info.orientation - degrees + 360) % 360;

//STEP 2: Set the 'rotation' parameter
Camera.Parameters params = mCamera.getParameters();
params.setRotation(rotate);
mCamera.setParameters(params);
```

Figure 4.6: Pseudo Code of rotating actual image

This solved the issue of saving image as landscape in SD card. But when the images were fed into the training module of Face++, it was not accepting the

images since the images were over the size limitation of 2MB. This required the captured images to be resized before sending to Face Recognition Module. The available methods caused the above rotation to be void. The resizing caused the image to be rotated back to it's original. Most of the available examples indicated the image needs to be first saved and then resized to overcome this but this had to be done at the exact moment of capturing the image since saving and resizing was an overhead in the application because the response time is a trivial factor. During resizing, it was important to maintain the scale of height and width in order to maintain the quality of the picture.

When resizing the image, Scaling the Image is very important to maintain the height to width ratio for the quality of the image. When the image is scaled it will be read as a byte array and decoded into Bitmap. Then original height and width will be calculated. When scaling the image it will be scaled in the widthToHeight ratio of

Scaling * originalWidth / originalHeight

where scaling of 1.0 is used. Then the image will be scaled based on the following logic. If the original width is larger than max width or oroginal height is larger than max height it will be checked if original width is larger than original height. If the condition matches, image will be scaled from width. If the Original Height is larger than original width image will be scaled from Height.

But this will rotate the image and Figure 4.7 shows pseudo code how the image will be read as a Matrix and the matrix will be rotated. A new bitmap will be created from the rotated matrix which will be compressed in JPEG format and saved.

This method saves pre process time by not having to save the resized image first and then rotating the saved image.

```
// createa matrix for the manipulation
Matrix matrix = new Matrix();

matrix.postRotate(90);

// recreate the new Bitmap
Bitmap resizedBitmap = Bitmap.createBitmap(matrix);

//create a byte array output stream to hold the photo's bytes
bytes = new ByteArrayOutputStream();
//compress the photo's bytes into the byte array output stream
resizedBitmap.compress(Bitmap.CompressFormat.JPEG, 40, bytes);
```

Figure 4.7: Pseudo Code of Rotating Image after Resizing

4.3. Analysis Module

4.3.1 Background of Face Recognition APIs - OpenCV

One of the most popular candidates for Face Recognition that is being widely used is OpenCV [22] which is an open source computer vision library started by Intel in 1999. It is free for commercial and research use under a BSD license. The library is cross-platform, and runs on Windows, Linux, Mac OS X, mobile Android and iOS.

OpenCV uses the Eigenfaces method which is described in [23] which takes a holistic approach to face recognition. A facial image is a point from a high-dimensional image space and a lower-dimensional representation is found, where classification becomes easy. The lower-dimensional subspace is found with Principal Component Analysis, which identifies the axes with maximum variance. While this kind of transformation is optimal from a reconstruction standpoint, it doesnt take any class labels into account. Imagine a situation where the variance is generated from external sources, let it be light. The axes with maximum variance do not necessarily contain any discriminative information at all, hence a classifica-

tion becomes impossible. So a class-specific projection with a Linear Discriminant Analysis was applied to face recognition. The basic idea is to minimize the variance within a class, while maximizing the variance between the classes at the same time.

The Face Detection is achieved through a "cascade" file. Paul Viola and Michael Jones introduced a "cascade" based Framework for Rapid Object Detection [24]. This work is distinguished by three key contributions. The rst is the introduction of a new image representation called the "Integral Image" which allows the features used by our detector to be computed very quickly. The second is a learning algorithm, based on AdaBoost, which selects a small number of critical visual features from a larger set and yields extremely efscient classiers following the research conducted by Freund et al [25]. The third contribution is a method for combining increasingly more complex classiers in a "cascade" which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions. The cascade can be viewed as an object specic focus-of-attention mechanism which unlike previous approaches, provides statistical guarantees that discarded regions are unlikely to contain the object of interest. In the domain of face detection the system yields detection rates comparable to the best previous systems. Used in real-time applications, the detector runs at 15 frames per second without resorting to image differencing or skin color detection.

4.3.2 Background of Face Recognition APIs - Face++

Another candidate for Face Recognition is Face++ [26]. Face++ is a new cloud platform providing comprehensive face recognition technology. It provides an API to help developers and researchers to easily use face-related vision techs as a tool. As opposed to Eigenface method employed by OpenCV, Face++ uses Facial Landmark Localization and Deep Convolutionary Neural Networks based Training for Recognition. Deep convolutional neural networks (DCNN) have been successfully utilized in facial landmark localization for two-fold advantages:

- Geometric constraints among facial points are implicitly utilized
- Huge amount of training data can be leveraged.

However, as discussed by Zhou et al [26] in the task of extensive facial landmark localization, a large number of facial landmarks (more than 50 points) are required to be located in a unified system, which poses great difculty in the structure, design and training process of traditional convolutional networks. According to Face++ [26] a four-level convolutional network cascade was designed, which tackles the problem in a coarse-to-fine manner. In this system, each network level is trained to locally refine a subset of facial landmarks generated by previous network levels. In addition, each level predicts explicit geometric constraints (the position and rotation angles of a specic facial component) to rectify the inputs of the current network level. The combination of coarse-to-fine cascade and geometric refinement enabled this system to locate extensive facial landmarks(68 points) accurately in the 300-W facial landmark localization challenge.

4.3.4 further discusses how the Facial Landmark Detection is performed in Face++. This section elaborates on how the face is classified into two parts, namely inner and contour where inner represents eyebrows, mouth, eyes and nose and contour represents the cheek area. 4.3.5 elaborates the Learning Architecture of the system. This section focuses on the three layers of the learning system; the first layer dividing the input face into inner and contour and after the second level how the facial components in inner part are further separated.

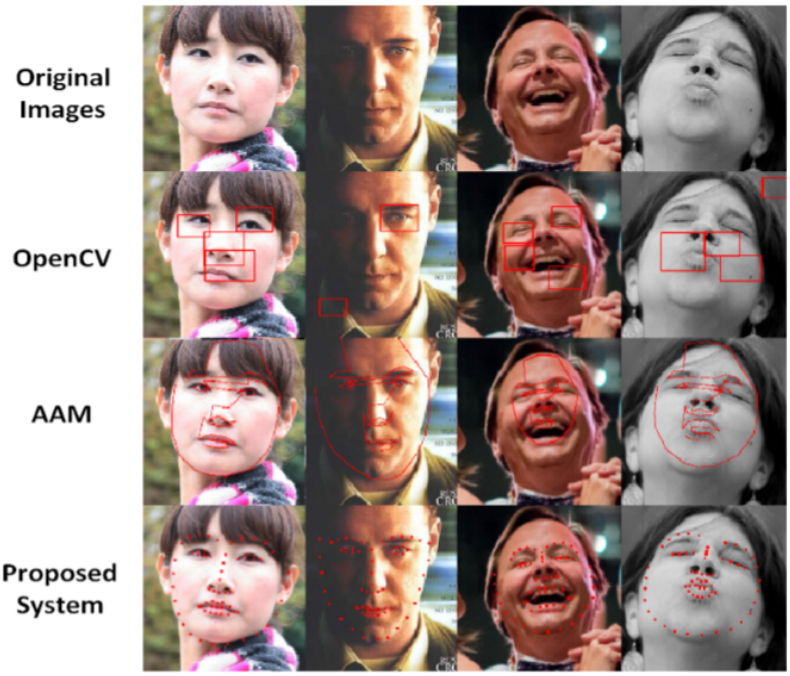
According to Face++ website [27] "Face++ APIs are currently provided free of charge. However, Face++ may charge fees for future use of or access to the Face++ APIs or the Face++ services according to its sole discretion. If Face++ decides to charge for the Face++ API's Services, Face++ will provide you prior notice of such charges. Face++ may also charge you when providing you a service different from the service under these Terms; for example, we may charge additional fees for excessive API use".

4.3.3 Comparison between OpenCV and Face++

OpenCV uses Eigenface which is a holistic approach, where information is extracted from face as a whole as opposed to Face++ which uses facial landmark points for face recognition. OpenCV is free under the GNU license where as Face++ is as of now free but the team states in future there might be a possibility of incurring charges for license. OpenCV and Face++ both provide an

Android SDK. In that case, both mechanisms allow the researcher to develop the Face Recognition in mobile phone itself. Nevertheless, both systems support Face Recognition in a server environment and connection from the mobile phone to the server environment via internet as well. Eigenface based method in OpenCV is easy to implement in terms of algorithmic complexity. But this has the drawback of adjusting to variations in facial expressions, illumination and pose. OpenCV Face Recognition Documentation states [28] that such a landmark detection method is robust against changes in illumination by its nature, but has a huge drawback: the accurate registration of the marker points is complicated, even with state of the art algorithms. But with the advancement of technology Face++ addresses this drawback in existing Facial Landmark Detection system as explained in 4.3.2.

Figure 4.8 shows the Comparison of Facial Landmark Localization Systems and the "Proposed Method" mentioned in this image refers to Face++ landmark detection implementation. From this Figure 4.8 it can be observed Face++ stands out from OpenCV and Active Appearance Model(AAM) based systems and manages to locate higher number of landmark points compared to OpenCV and AAM based methods.



Source: Zhou et al [26]

Figure 4.8: Comparison of Available Landmark Localization Systems

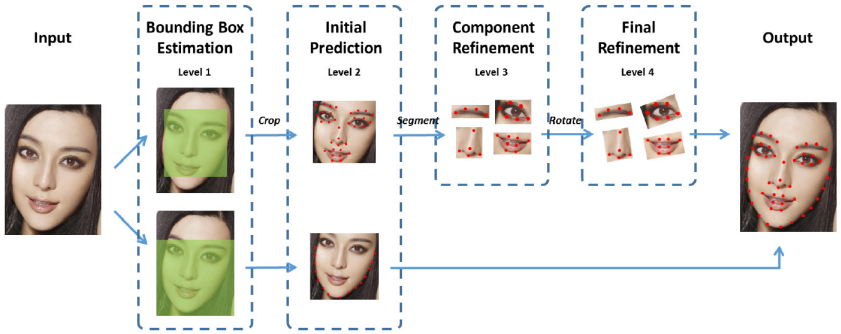
The system being implemented requires high accuracy and adhering to variations in expression, pose and illumination conditions. Even OpenCV Face Recognition Documentation [28] states Face recognition based on the geometric features of a face is probably the most intuitive approach to face recognition [28]. Face++ uses Facial Landmark detection for Face Recognition. Face++ facial point detection can locate the key components of faces, including eyebrows, eyes, nose and mouth. Face++ supports covered and multi-angle faces, which can handle complex facial expression. This leverages the system to recognize people even with different expressions on them. Hence it was decided as one of the dominating factors for choosing Face++ against OpenCV.

Face++ also supports a Cloud Based Recognition Engine. This requires the user to upload an image to the Cloud Based System and processing is performed in the Cloud. Hence, mobile phone does not require to do any extensive processing and the system can be installed in a mobile phone with average configuration. This

eases the use of high end processors on mobile phone or any explicit server maintenance OpenCV would cost. This was also a trivial factor in choosing Face++ vs OpenCV.

4.3.4 Facial Landmark Detection by the system

Figure 4.9 gives a brief illustration of the multi-level facial landmark localization system used by Face++ [26]. They have classified the face into two parts Inner and Contour. The the term inner points is used to denote the 51 points for eyes, eyebrows, mouth and nose, and contour points for other 17 points on the contour. In the first level, two neural networks are trained to estimate the bounding boxes (the maximum and minimum value of the x-y coordinates) for the inner points and contour points independently. The estimated boxes are fed to the rest of the system respectively.



Source: Zhou et al [26]

Figure 4.9: Description of High Level Landmark Localization System

Inner points

For the inner 51 points, another three level convolutional neural networks are trained. After obtaining the bounding box of inner points, an initial estimation of coordinates of the 51 inner landmarks is produced by the second level. Based on the initial estimation, the local regions for 6 facial components (i.e., eyebrows, eyes, mouth and nose) are computed. Then the third level is trained to refine the landmarks of each facial component independently. Finally the rotation angle

of each component is estimated and then corrected to upright, and the rotated patches are fed to the fourth level network to predict the final results.

Contour points

A simpler network cascade is applied for the contour points localization. Given the bounding box covering the cheek, the second level takes the cropped image as input and computes the coordinates of the contour points from the raw pixels. Third and fourth level networks are not utilized due to the limited time.

Implementation of Facial Landmark Detection

A new local face detector has been released (iOS and Android) by Face++. Developers can download it on the My App page on the Face++ website. Although the research paper [26] uses only 67 landmark points, it is able to detect 83 key facial points, including eyebrows, eyes, nose, mouth, and face contour.

Below Figure 4.10 is an example image used for training the CNN

Figure 4.10: User Image used for Training the System

The Face Detection returns the following results when the Figure 4.10 was fed into the system. These data will be used for recognition and training of the image.

{"face":[{"attribute":{"age":{"range":5,"value":18},
"gender":{"confidence":99.106,"value":"Female"},
"race":{"confidence":99.7558,"value":"White"},
"smiling"{"value":93.4754}},
"face_id":"e4aee1e6797f7a9e61b4e54da31104b6",
"position":{"center":{"x":57.222222,"y":42.25},
"eye_left":{"x":55.373778,"y":38.360167},
"eye_right":{"x":62.947778,"y":41.4095},
"height":12.833333,"mouth_left":{"x":51.495333,"y":43.755167},
"mouth_right":{"x":59.814222,"y":46.970667},
"nose":{"x":57.675333,"y":43.269833},"width":17.111111},
"tag":""}],
"img_height":4000,"img_id":"17bb9e2075658fa7a2266439e1bf047d",
"img_width":3000,"session_id":"75dff4ff60cf4f6e870a4ddc4788c719",
"url":null,"response_code":200}

Figure 4.11: Configuring Android Layout to be Vertical

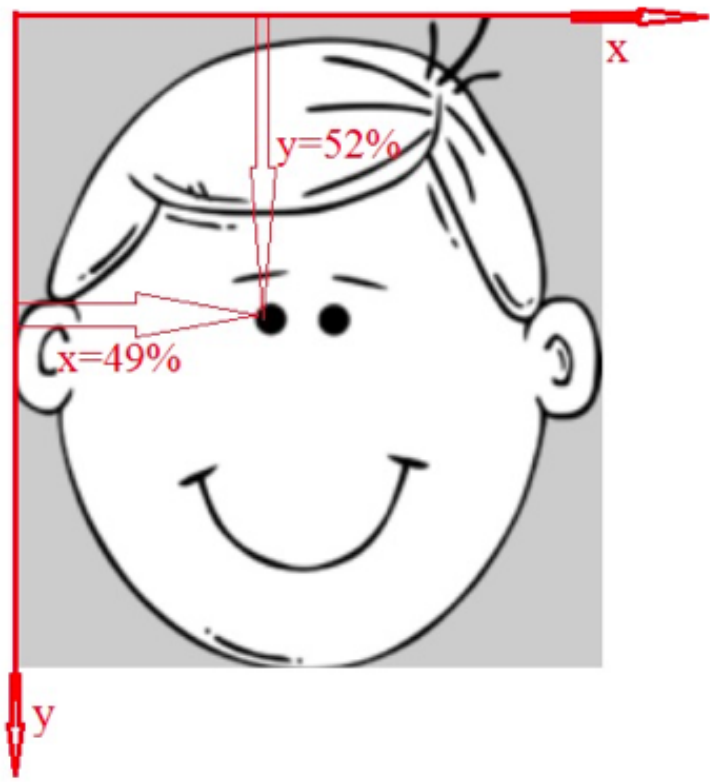


Figure 4.12: Measuring the coordinates of landmark positions

The position of the landmark location is calculated as a percentage from x and y axis as described in Figure 4.12

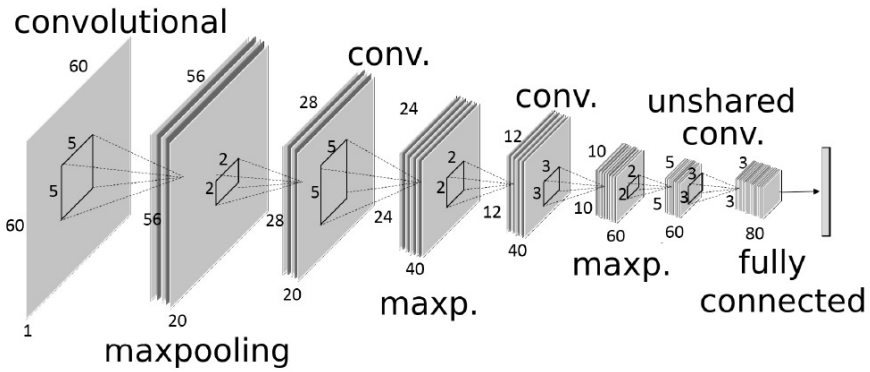
4.3.5 The Learning Architecture

Face++ uses Deep Convolutionary Neural Network (DCNN) as the basic building block of the system. The network takes the raw pixels as input and performs regression on the coordinates of the desired points. Figure 4.13 is an illustration of the deep architecture. Three convolutional layers are stacked after the input nodes. Each convolutional layer applies several filters to the multi channel input image and output the responses.

The framework uses the idea of coarse-to-fine localization. Each network level refines a subset of the landmarks inside a region computed by previous levels. In the first level, the face is divided into two parts : inner and contour. After the

second level, the facial components in the inner part are further separated.

DCNN is generally considered to be powerful enough to handle great variation in the input image, but the capacity of a single network is still limited by its size. Given insufficient prior knowledge, the network will devote a considerable part of its power to finding where the face is. To tackle the problem, the divide-and-conquer strategy is adopted, which divides the task into two steps: first to find the overall position, then to compute the relative position inside the region. For the whole face, the first step is performed by the first level networks whose supervision signal does not include the detailed structure of the points inside the bounding box, and the rest of the task is left to succeeding levels. In this way, the burden is shared across networks in different levels, and good performance is achieved by networks of only moderate size.



Source: Zhou et al [26]

Figure 4.13: Convolutionary Network employed by Face++

4.3.6 Face Recognition

Face++ Recognition method is based on CNN. The neural networks are applied to image patches, and their last layer activation is taken as the representation. To train the CNN, supervised signal is used.

Training

The neural networks are trained by stochastic gradient descent with hand-tuned hyper-parameters. To avoid severe over-fitting, the image is randomly altered by slight similarity transformation (rotating, translating and scaling) before feeding

into the network. This step creates virtually infinite number of training samples and keeps the training error close to the error on the validation set. Also, the image is flipped to reuse the left eye's model for the right eye, and left eye-brow for right eye-brow.

Image Processing

Image patch is normalized to zero mean and unit-variance, then a hyper-tangent function is applied so that the pixel values fall in the range of [1; 1]. When cropping the image inside a bounding box, the box is enlarged by 10% to 20%. More context information is retained by the enlargement, and it allows the system to tolerate small failures in the bounding box estimation step. In the fourth level, the rotation angle is computed from the position of two corner points in the facial component.

Training the current system for Face Recognition

In this project, the scenario was defined as the the user is meeting someone for the second time. It is assumed the user has met this person once and entered his/her details. Hence the system would contain only one image of the user for training the system. The current implementation of the Face++ recognition had images of 20 users for training the system.

4.4. Presentation Module

There were two main factors which were identified with relation to Presentation.

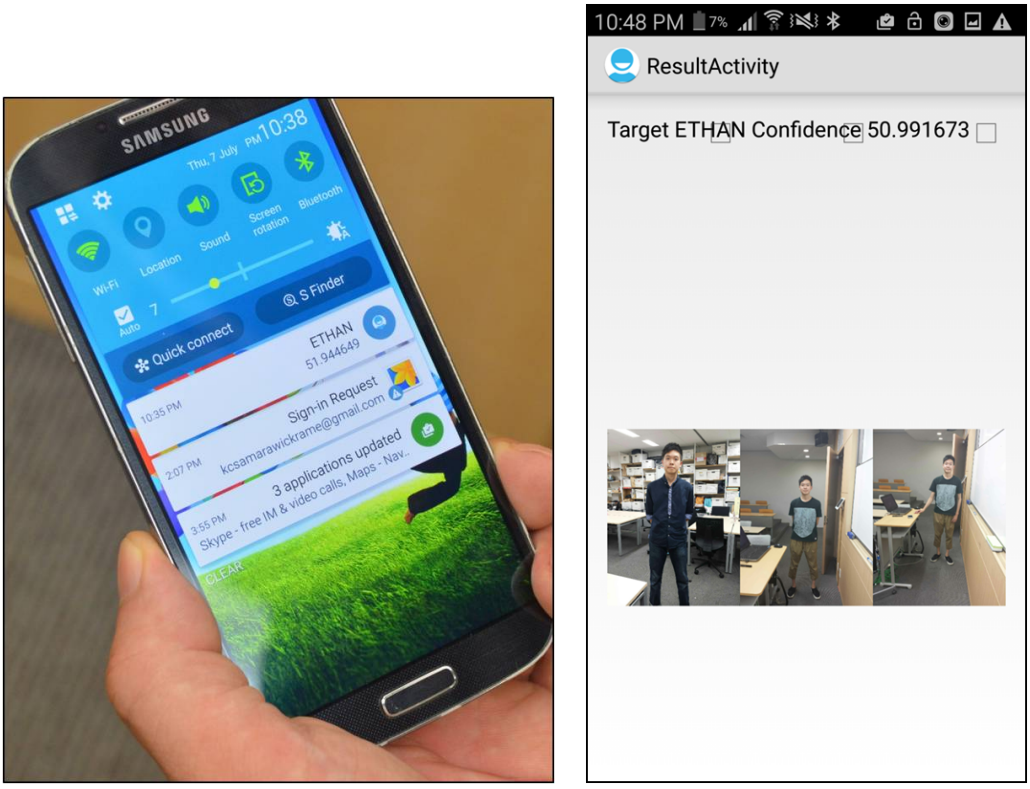
1) How to identify the accuracy of output

2) How to display data in an unobtrusive manner

One of the concerns raised by the faculty during the Interim Presentation was how can the user be sure of the result provided by the system or up to what extent can they rely on the output of the system. In order to cater this request, it was decided to show the CONFIDENCE level of the results. Depending on the Confidence level the user can take his own decision to use the information. None withstanding this, the system was developed to show past images of meeting the person with the NAME and CONFIDENCE level as well. The Confidence level and the past pictures combined together can help the user verify the accuracy of

the system's output.

The Face++ API result gives an array of results with confidence levels and the result with the highest confidence level is chosen to be produced to the user.



(a) Mobile Phone Notification (b) Detailed Display with Past Images

Figure 4.14: Visual Output on Mobile

Figure 4.14(a) shows the initial presentation of notification to the user with name and confidence as a percentage. Receiving the notification would make a vibration and user could have a quick glance at the screen and see the information. If the user requires to see more information he could click on the notification which would redirect him to the screen in Figure 4.14(b). This screen would contain past images which would help the user verify it is the same person he's talking to. These images will act as a trigger to activate organic memory as looking at the images will trigger his own memory of the event where he met this person. The background of the images will provide cues for him to remember better. Hence he

could verify the accuracy of the system and use the results with confidence. The Evaluation results presented in Section 5.3 and Section 5.5 further confirms the decision to show Name, Confidence and Past Images as memory cues.

Chapter 5

Evaluation and Discussion

5.1. Evaluation Criteria

In this Chapter the system will be evaluated to test the implementation of the two research questions discussed in Chapter 1. A quantitative analysis will be done to find out the effectiveness of using Voice Activity Detection as a trigger and a qualitative analysis will be done to evaluate the usability of the system.

5.2. Evaluation of the effectiveness of using Voice Activity Detection as a Trigger

5.2.1 Experiment 1: Trigger Based Capture

It was observed general lifeloggiong devices capture vast amount of data but the number of usable data or data with a potential for memory enhancement is less. Hence there is a need to implement selective capture to eliminate vast amount of unusable data. The experiment was designed comparing Autographer and the current system. The two systems were used over a period of **one hour** to see how many images will be captured by both systems. Several tests were carried out to compare Autographer vs Voice Activity Detection based system and a sample test was picked randomly to evaluate the performance in terms of storage redundancy.

Table 5.1: Autographer vs Voice Activity Detection Trigger based system's Image Capturing

	Autographer	Voice Trigger based system
No. of Images	120	21
No. of Usable Images	2	16

As depicted in the Table 5.1 both Autographer and Voice Activity Detection Trigger Based system was used inside the Keio University Graduate School of Media Design over a period of one hour. Autographer captutred 120 images and Voice Activity Detection Trigger Based system captured 21 images. Table 5.2 shows further evaluation of images captured by Autographer. It can be clearly seen as soon as Autographer started it captured 8 images per minute and continued to capture an average of 2 images per minute for the next 59 minutes. However with the Voice Activity Detection Trigger based system, it reduced the data capture by 82.5%.

Table 5.2: Image Capture of Autographer vs Voice Activity Detection Trigger Based System.

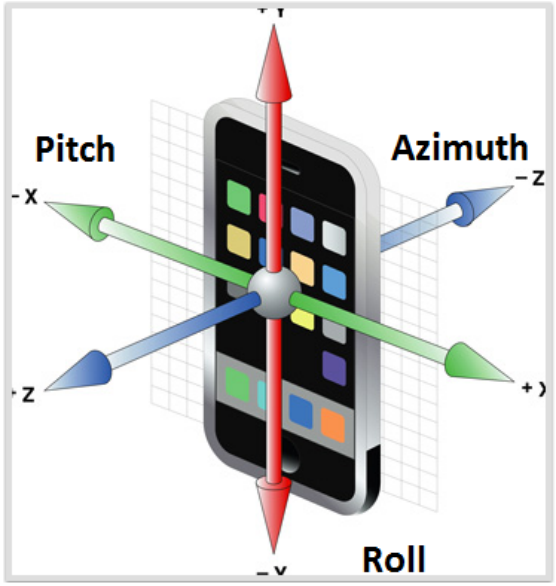
Time	Autographer	Time	Autographer
4.15 PM	8	4.46 PM	2
4.1 6PM	3	4.47 PM	2
4.17 PM	2	4.48 PM	1
4.18 PM	2	4.49 PM	2
4.19 PM	2	4.50 PM	2
4.20 PM	2	4.51 PM	2
4.21 PM	2	4.52 PM	2
4.22 PM	2	4.53 PM	2
4.23 PM	3	4.54 PM	1
4.24 PM	2	4.55 PM	2
4.25 PM	2	4.56 PM	2
4.26 PM	2	4.57 PM	2
4.27 PM	2	4.58 PM	2
4.28 PM	2	4.59 PM	2
4.29 PM	2	5.00 PM	2
4.30 PM	2	5.01 PM	1
4.31 PM	2	5.02 PM	2
4.32 PM	1	5.03 PM	3
4.33 PM	2	5.04 PM	1
4.34 PM	2	5.05 PM	2
4.35 PM	1	5.06 PM	2
4.36 PM	2	5.07 PM	2
4.37 PM	1	5.08 PM	2
4.38 PM	2	5.09 PM	1
4.39 PM	2	5.10 PM	2
4.40 PM	1	5.11 PM	2
4.41 PM	2	5.12 PM	2
4.42 PM	1	5.13 PM	3
4.43 PM	2	4.44 PM	2
4.45 PM	1	5.15 PM	2

5.2.2 Experiment 2: Context for Face Recognition

The second experiment was carried out in order to find if there's a correlation between the context values Acceleration, Azimut, Pitch, Roll and Ambient Light of the images when it comes to accurate recognition.

A conversation was carried out for 20 minutes 9 seconds and during this period 18 images were captured. The scenario was defined as meeting someone you have met only once before. The system was trained with only one image of the particular user. Out of the 18 images 11 images were accurately identified. And 7 images gave the result face cannot be detected. The images were taken in the same location under different lighting conditions, different rooms and different seating and standing positions.

An experiment was conducted to identify if orientation of the capture device has an impact on capturing a good quality image which can be recognized by the training engine. Orientation has three parameters Azimut, Pitch and Roll and these three parameters will be examined below.



Source: StackOverflow Website [29]

Figure 5.1: Rotation of the mobile phone

As the above diagram Figure 5.1 describes

- Azimuth - Azimut calculates the angle around the z-axis.
- Pitch - Pitch calculates the orientation angle around x-axis.
- Roll - Roll parameter refers to the orientation angle around the y-axis.

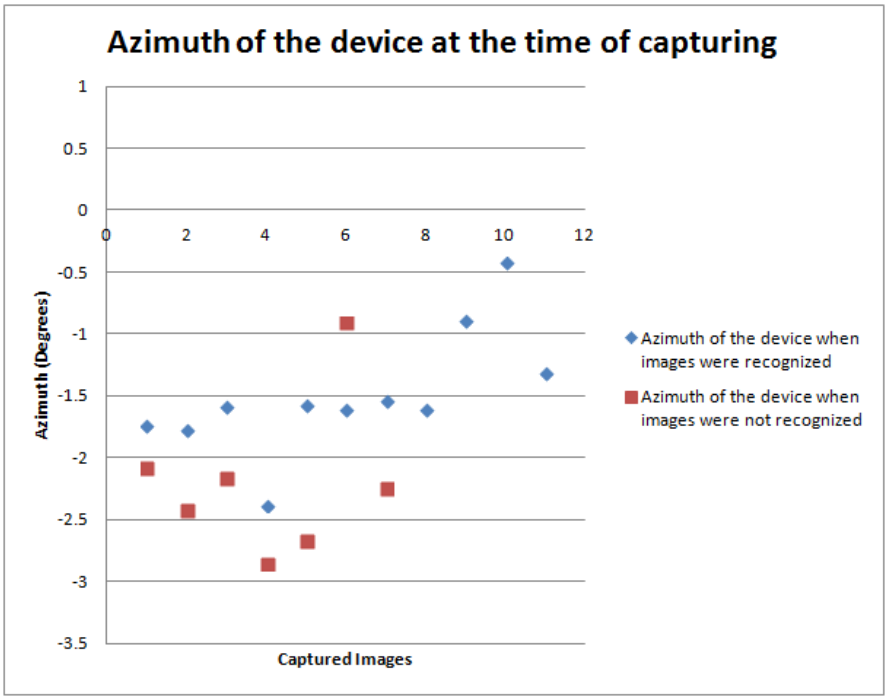


Figure 5.2: Azimuth of the image sequence

As shown in Figure 5.2 Azimuth calculates the angle around the z-axis. Observing the Azimuth values of the device at the time of capturing images which were recognized and not recognized it could been seen values of both categories range between -2.85 degrees and -0.42 degrees. This shows Azimuth parameter has remained same irrespective of weather the image was able to be recognized or not. Hence it can be concluded azimuth parameter was not the determinant factor on capturing a good image for Face Recognition.

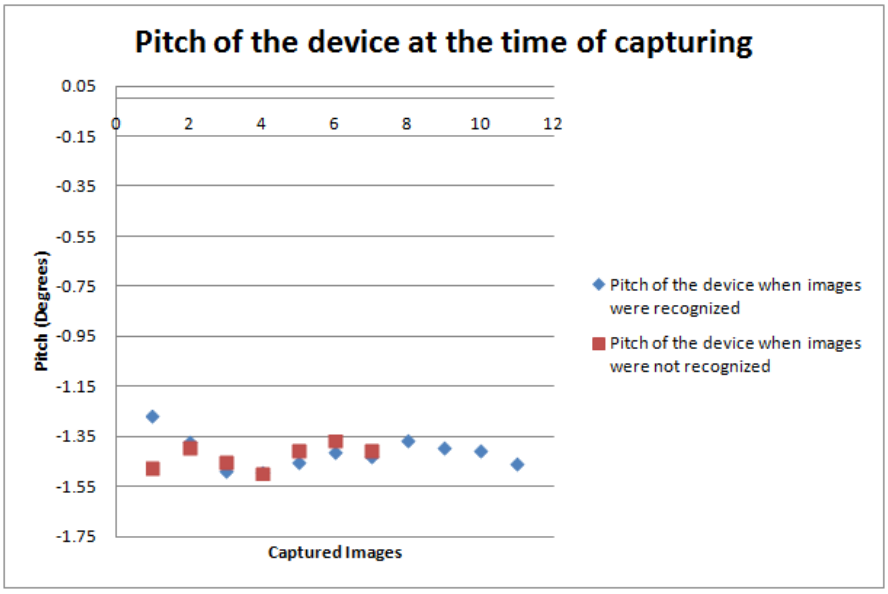


Figure 5.3: Pitch of the image sequences

As depicted in Figure 5.3 Pitch calculates the orientation angle around x-axis. And the values range from -1.50 degrees to -1.26 degrees at the time of capturing the images. Both categories does not show high variance and fall between the same range. Since pitch remains constant for both recognized and not recognized images, we can come to the conclusion that Pitch of the Capture Device was not the determinant factor on capturing an image of an already trained target to be easily recognized.

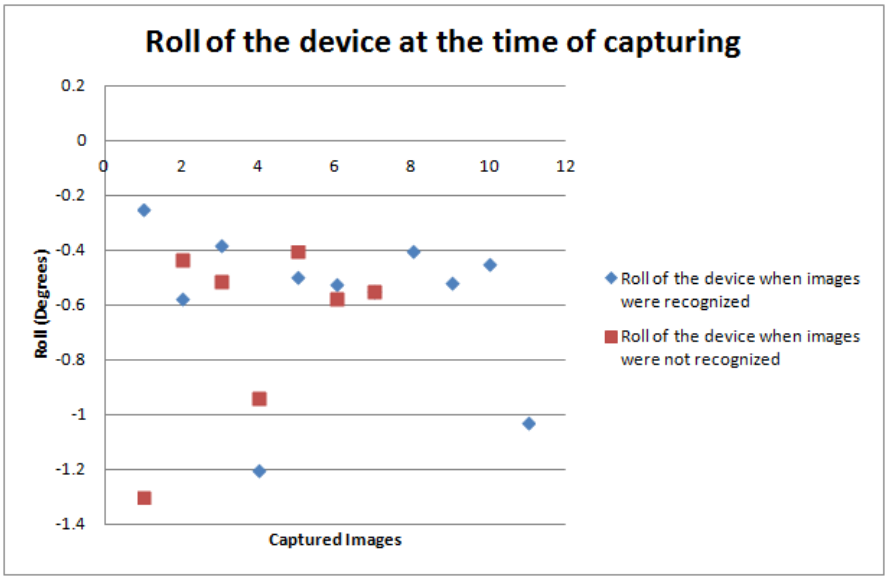


Figure 5.4: Roll of the image sequence

As shown in Figure 5.4 Roll parameter refers to the orientation angle around the y-axis. The values range between -1.57 and -0.07 degrees. It can be observed that roll remains constant throughout the time of capturing images by the device. Yet, some images being recognized and some failing to be recognized shows, orientation parameter Roll is not the determinant factor if the device is being held in a standard position.

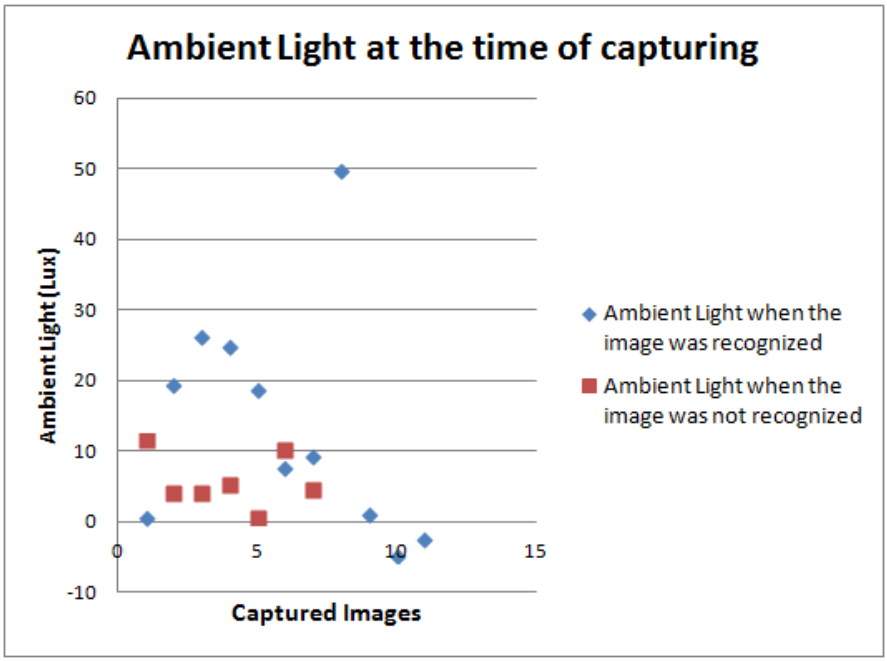


Figure 5.5: Illumination of the image sequence

According to Figure 5.5 Illumination values range between -4.61 LUX and 49.77 LUX the lowest value -4.61 for images taken in darker area and the highest value 49.77 for the image taken in average bright light inside the room. Although according to the Standard illumination scale the minimum lux value is 0, probably due to an error in the Android RGB sensor mentioned in Table 4.2 we have achieved a minus value for illumination. But as the figure depicts even for Illumination both Recognized and Not Recognized image categories fall in the same range. Hence we can deduce, in standard lighting conditions, slight variations of illumination parameter does not affect taking a good quality image which can be recognized by the recognition engine.

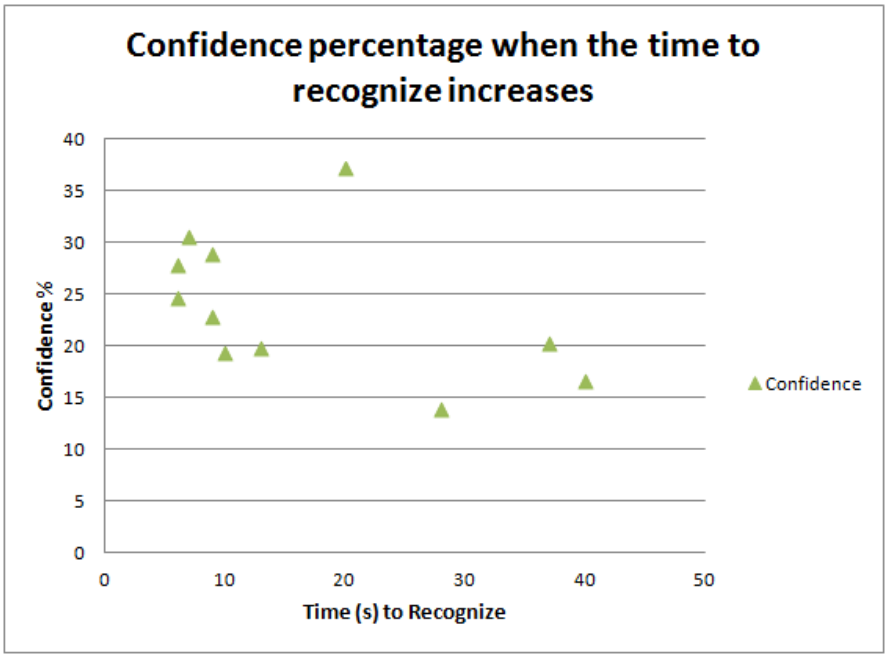
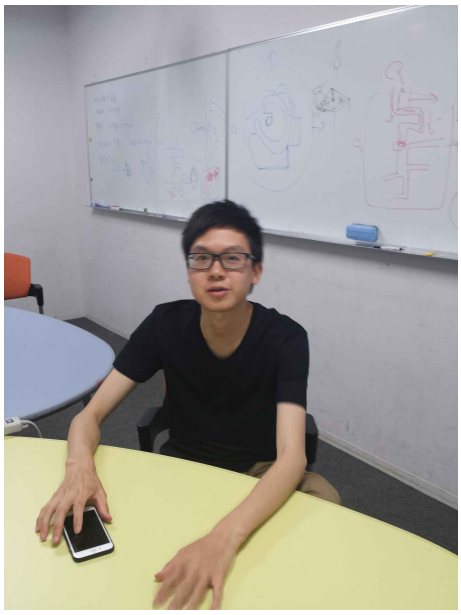


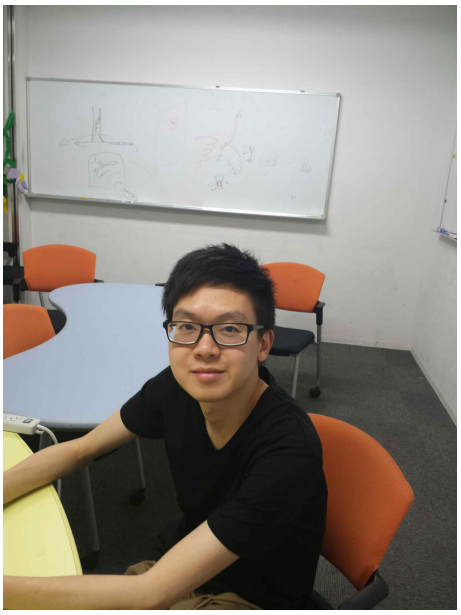
Figure 5.6: Confidence as a percentage when the time to recognize increases

As shown in Figure 5.6 some images have taken a high value as 40seconds to give the recognition result and with the confidence 16.78% that the target is the identified user. The shortest time for recognition during this session was reported as 6 seconds and the system was confident the target is 24.81% assured to be the one given as result. Observing the overall Time to recognize vs Confidence, it can be observed the images which took the shortest time to recognize had a higher confidence level of identifying the target.

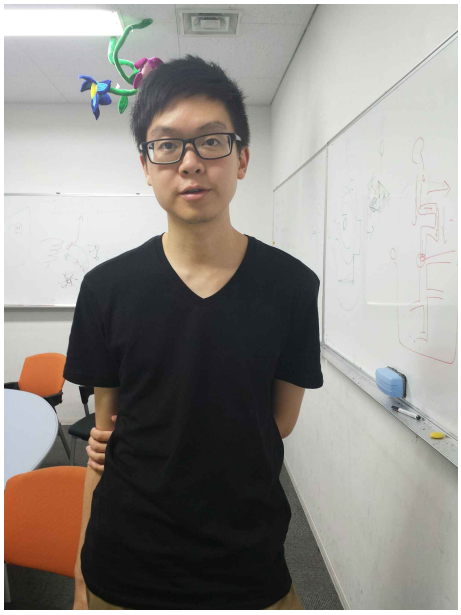
However, 7 images failed to recognize the person. Analysing the above context values of Orientation and Illumination we can conclude that in a standard environment where the user is holding the mobile phone in front of him/her, in normal lighting conditions, irrespective of being seated or standing the main factor determining if the target can be recognized is the fact that face is visible to the system without any pose variations. The images captured during this session is described in Figure 5.7. It should also be emphasized that in the training image of the user, he did not have any glasses and yet the system was able to recognize accurately.



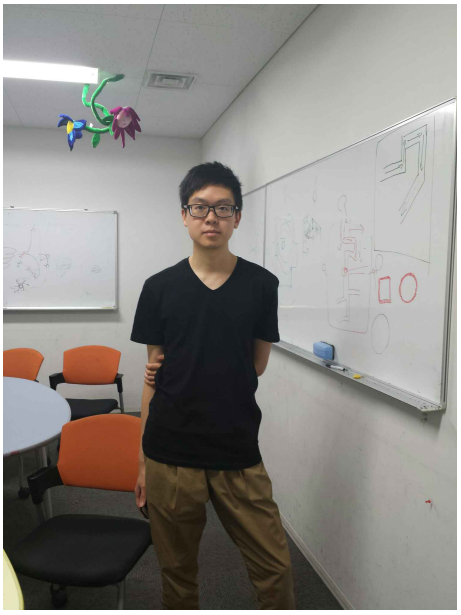
(a) 0.5meter distance Wearing Spectacles Recognized 13.95% Confidence



(b) 0.5meter distance Wearing Spectacles Recognized 22.94% Confidence

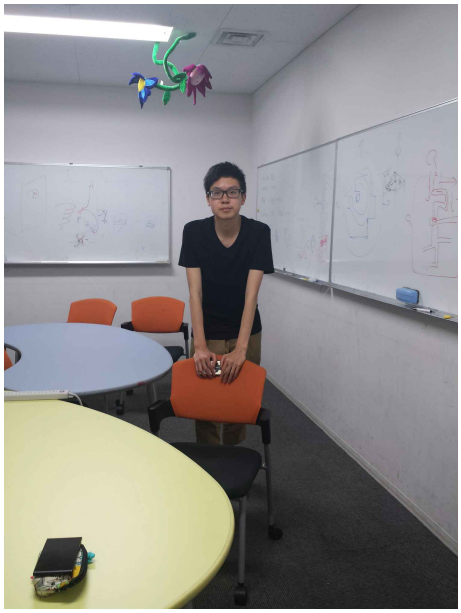


(c) 0.5meter Distance Wearing Spectacles Recognized 27.95% Confidence

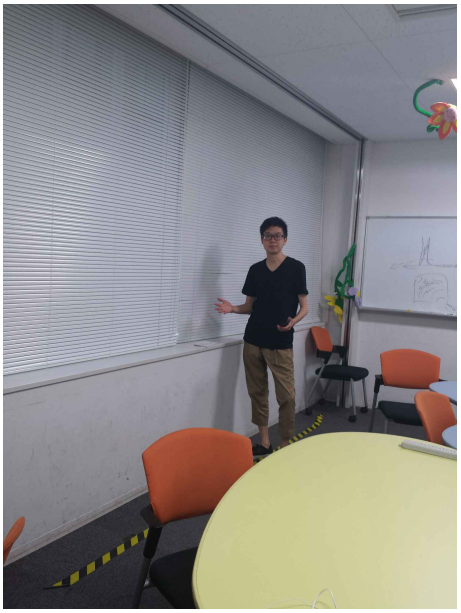


(d) 0.5meter Distance Wearing Spectacles Recognized 19.97% Confidence

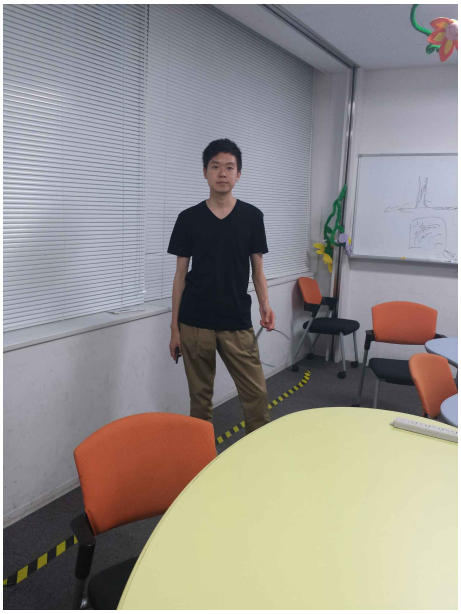
us



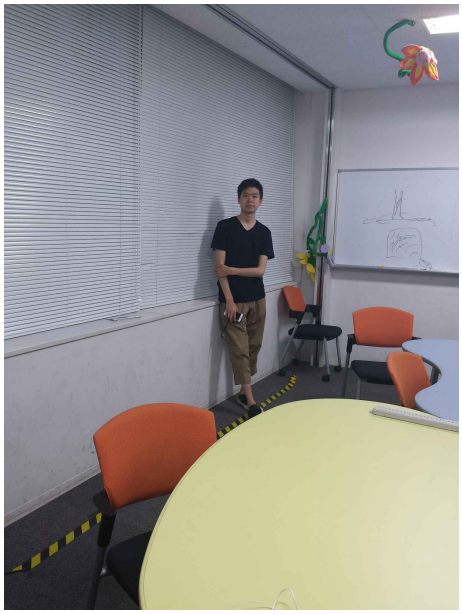
(e) 1 meter Distance Wearing Spectacles Recognized 19.53%



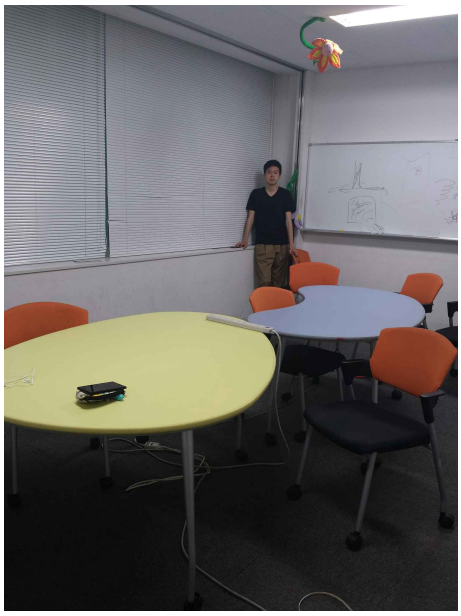
(f) 1.5 meter Distance Wearing Spectacles No Face Detected



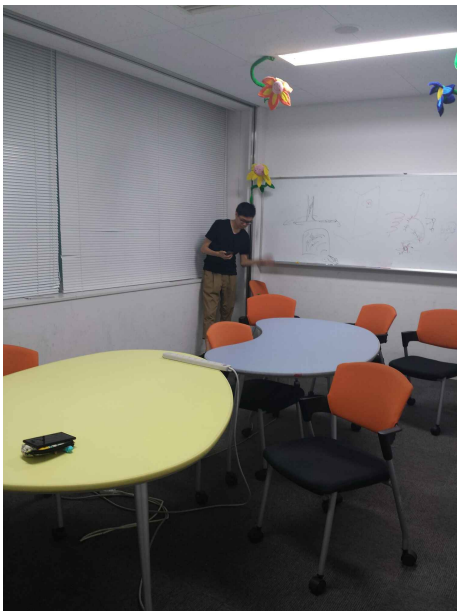
(g) 1.5 meter Distance No Spectacles Recognized 37.23%



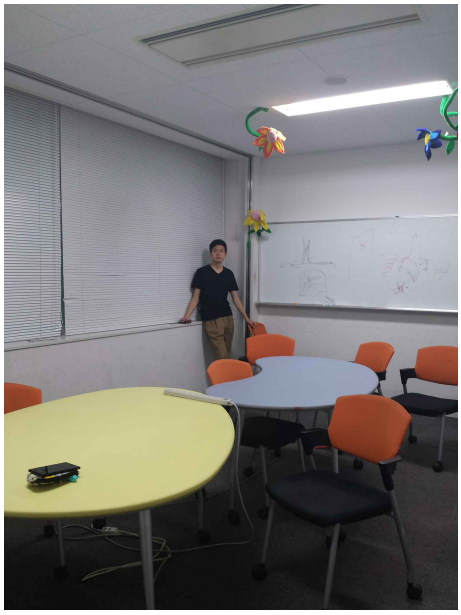
(h) 1.7 meter Distance No Spectacles Recognized 16.78%



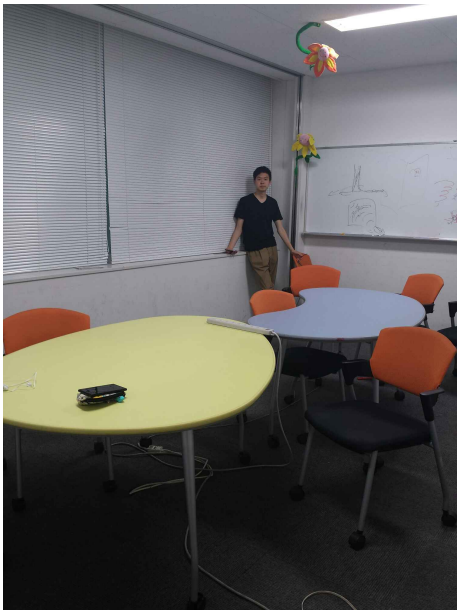
(i) 2 meter Distance No Spectacles No Face Detected



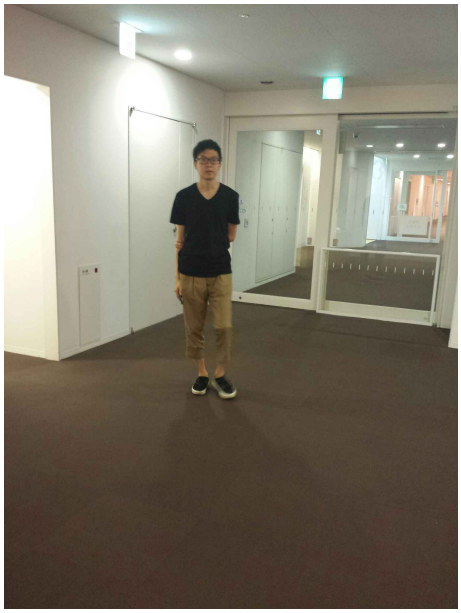
(j) 2 meter Distance Wearing Spectacles No Face Detected



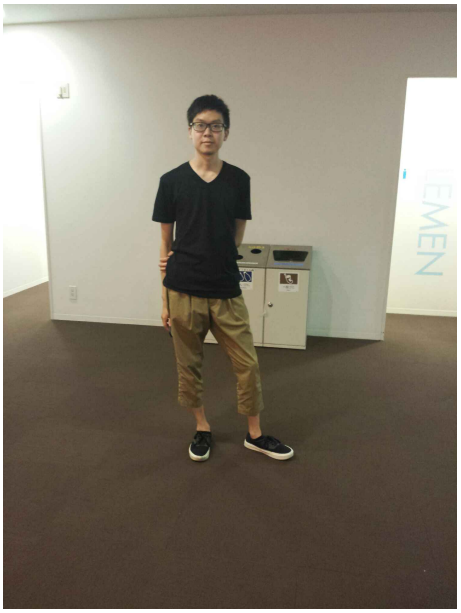
(k) 1.5 meter Distance Wearing Spectacles No Face Detected



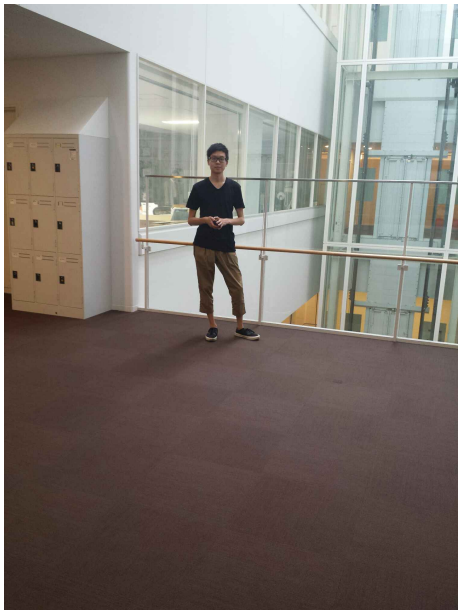
(l) 1.5 meter Distance Wearing Spectacles No Face Detected



(m) 1.5 meter Distance Wearing Spectacles No Face Detected



(n) 1 meter Distance Wearing Spectacles Recognized 20.41%



(o) 3 meter Distance Wearing Spectacles No Face Detected

(p) 2 meter Distance Wearing Spectacles Recognized 30.68%

(q) 1 meter Distance Wearing Spectacles Recognized 24.81%

(r) 1.5 meter Distance Wearing Spectacles Recognized 29.02%

Figure 5.7: Recognized and Non Recognized Images

5.3. Evaluation of the Usability of the System

The initial set up of the system was implemented with capturing the images using Samsung S4 mobile phone based on Voice Activity Detection as shown in Chapter 4. The first Expert User Evaluation was conducted with Prof. Kai Kunze. The system was able to recognize the user in 31 seconds and display the name of the target and recognition confidence on the mobile phone screen. It was advised to consider different presentation methods and obtain user feedback. Hence it was decided to conduct a workshop style session where the user will first be given the opportunity to experience the current system and answer the questionnaire based on his/her experience on the Usability of the System, mainly targeting unobtrusive feedback by the system. Before the second experiment was started the system was improved not only to display the results on the mobile phone screen but also provide audio output as a result.

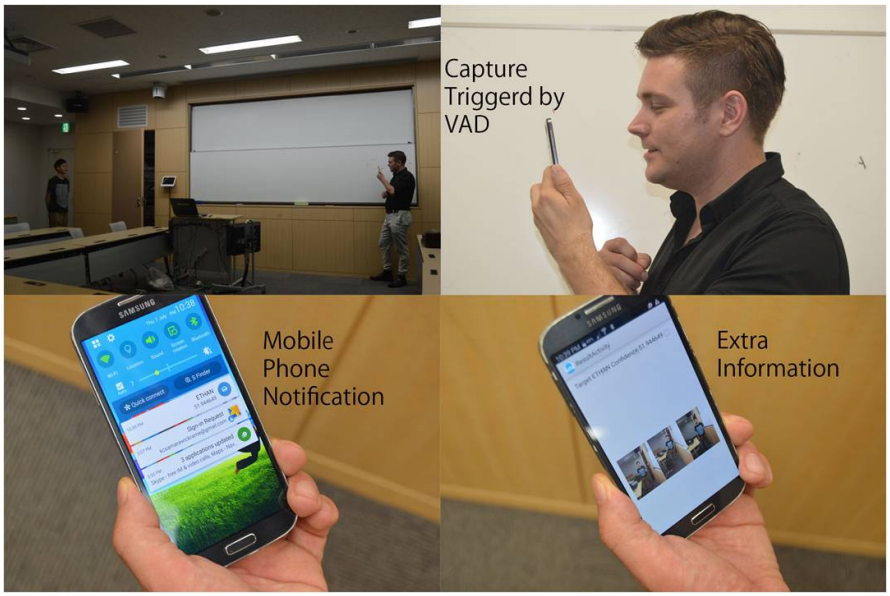


Figure 5.8: Prototype Design

Figure 5.8 shows the final prototype test carried out by users. Depending on the first iteration of evaluation, it was decided to include a Mobile Phone Notification as the primary method to display results which would then lead to a screen with past images depending on the user's preference.

The Questionnaire for Usability Evaluation

1. Participants Age Distribution

Table 5.3: User age distribution

Age	Participants
20-25	4
26-30	4
31-40	2
41-60	0

2. How often do you meet a person you should have met before but cannot remember?

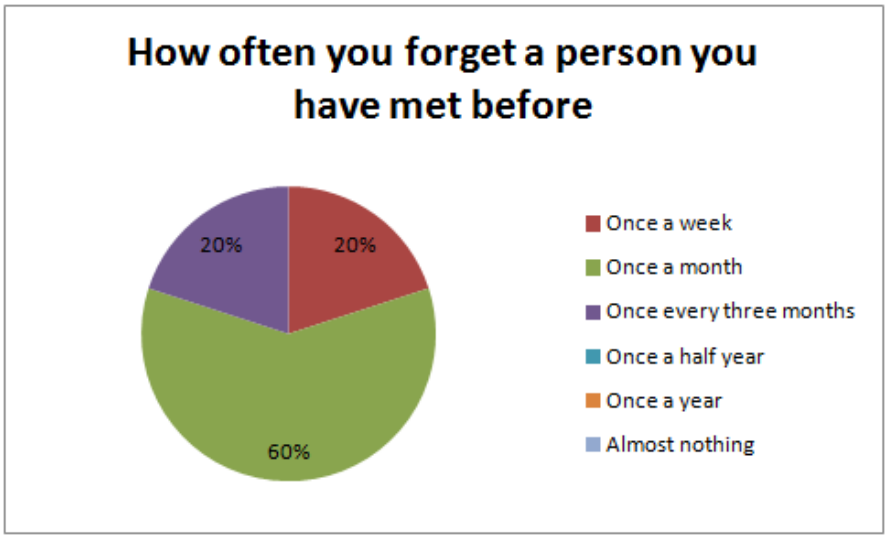


Figure 5.9: The tendency of forgetting a person

3. Do you feel very embarrassed with the situation?

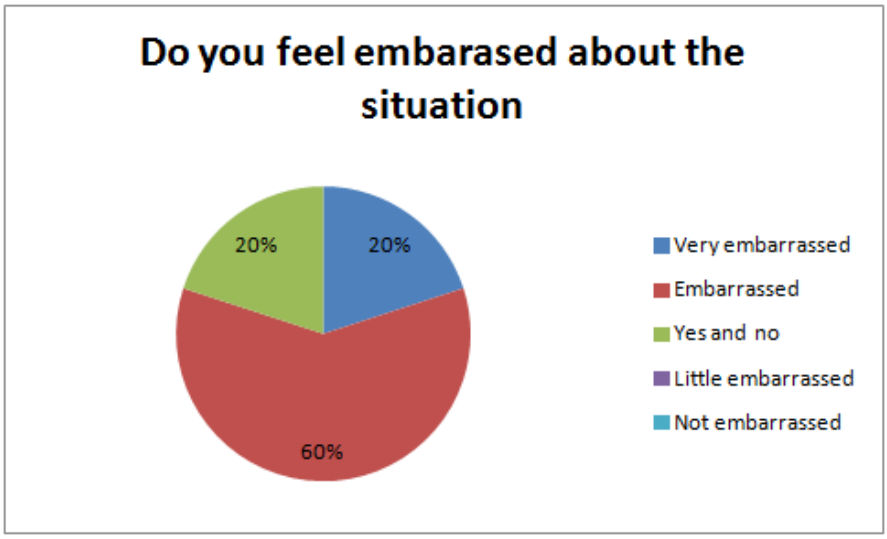


Figure 5.10: How embarrassing it is to forget

4. How big is the problem of asking them about previous meetings?

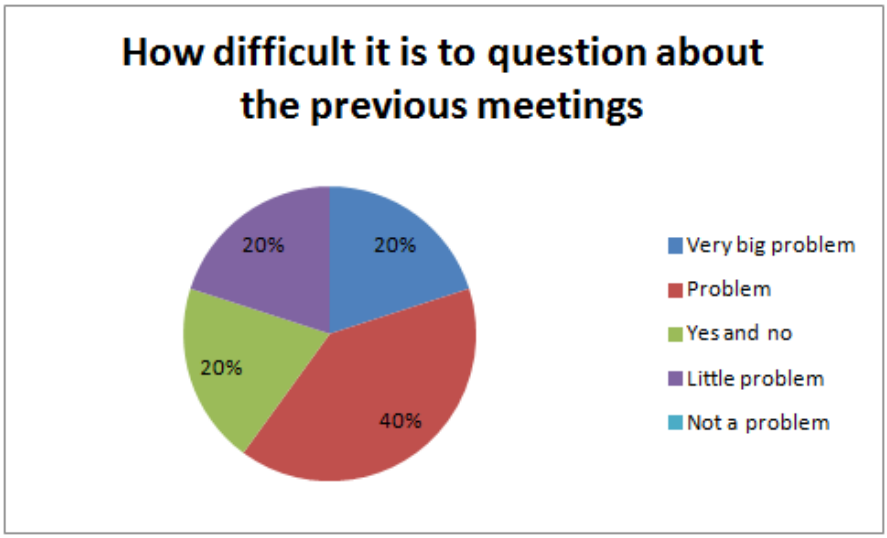


Figure 5.11: User's feeling about asking about previous meetings

5. What cues remind you of the person?

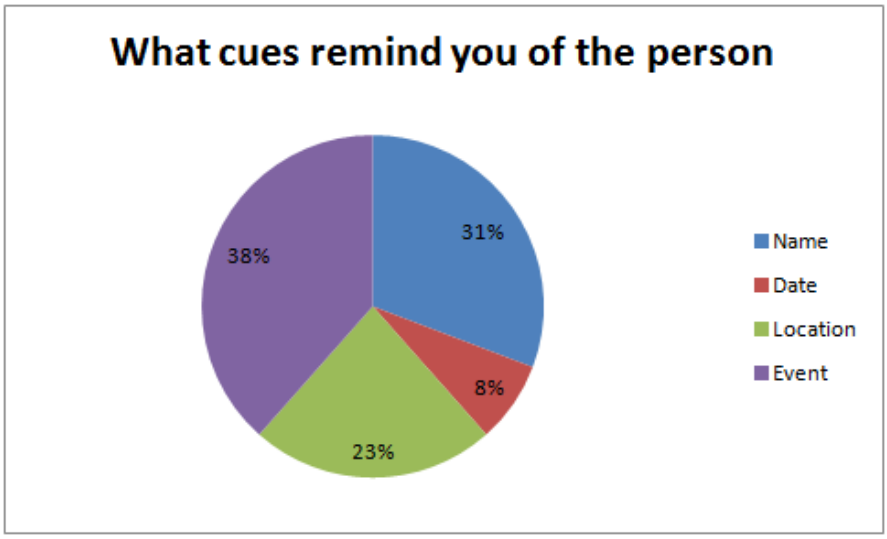


Figure 5.12: Memory cues

6. Would you carry a system like that with you which would help you remember people when you see them next time?

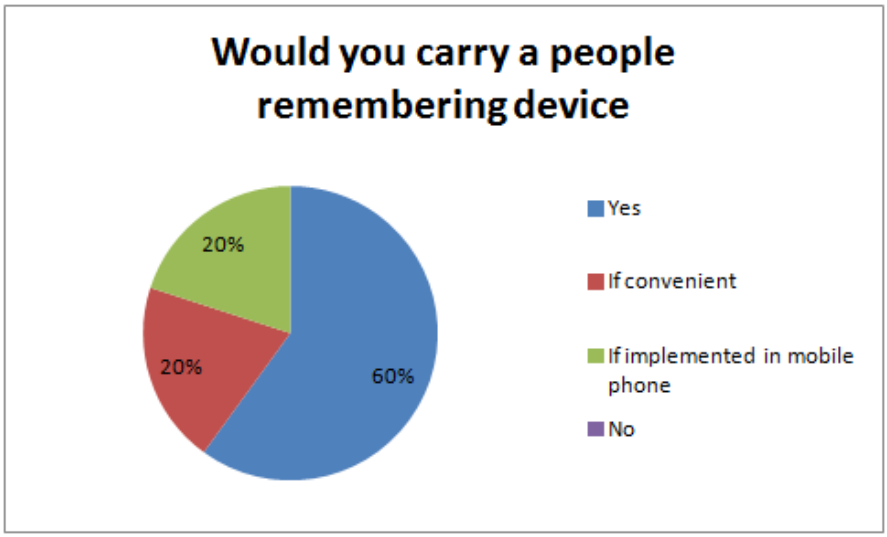


Figure 5.13: Opinion about using a device for that helps them remember people

7. Would you use the current implementation?

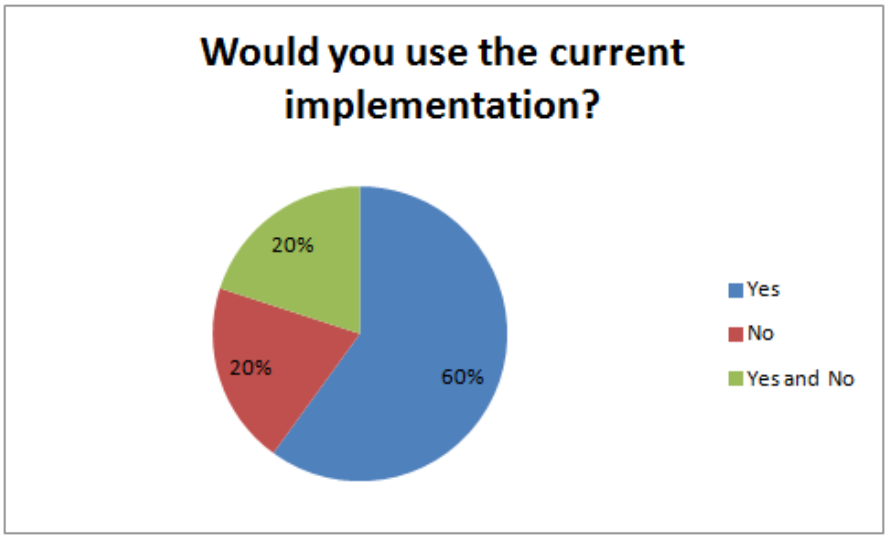


Figure 5.14: How comfortable to use the current system

If the answer is no what is the reason?

8. Which device would you like to see the result as output?

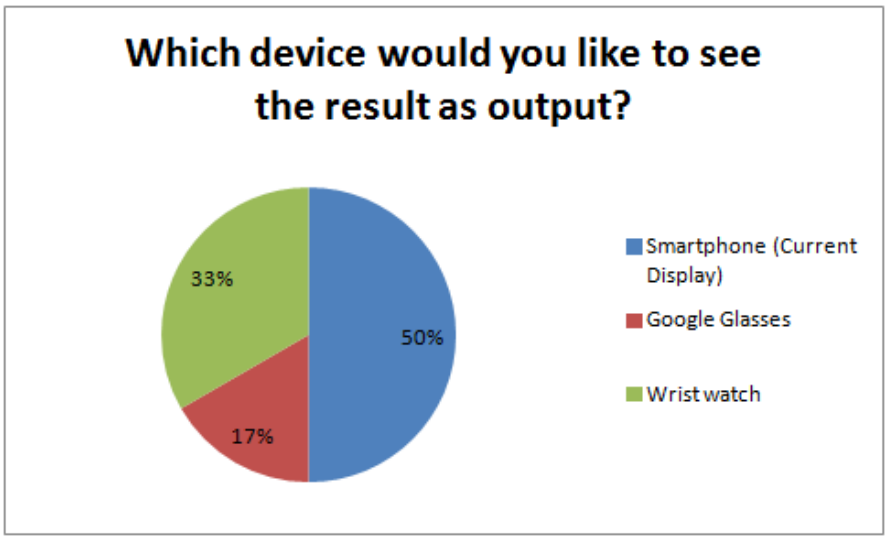


Figure 5.15: Displaying results back to the user

9. Would you wear it everyday or to specific events?

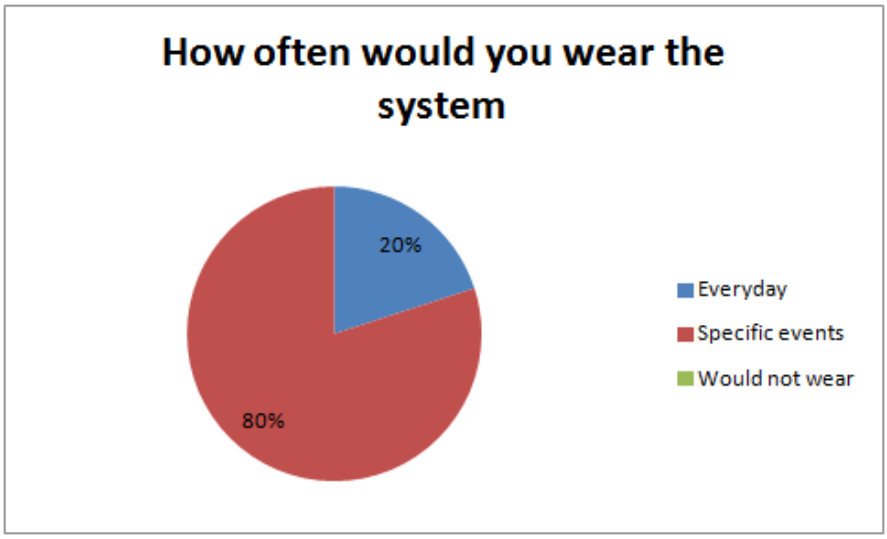


Figure 5.16: What occasions to wear the system/How often

5.4. Discussion of the Usability Test Results

The user test was carried out among students and lecturer's at Keio University Graduate School of Media Design. The participants were from different nationalities comprising of Sri Lankan, German, Taiwanese, Omani and Japanese.

Examining the user study results it can be seen majority of the participants were among 20-30 years of age. This shows the majority were a young group of people who do not suffer from natural short term memory loss situations that come with aging.

Yet, the interesting fact is most of them experience meeting someone they have met before but not able to recall the name. This falls in line with the initial argument of the need for a memory assistance system irrespective of the age group. Majority of the users comprising of 60% voted they have a tendency to forget someone's name at least once a month. While 20% said they face this situation once a week and another 20% once every three months. The less possibility of this happening once a week could be due to the reason that user group being 20 to 40. If it had more participants in the range of 50 - 60 or participants who are traveling more often and meeting new people more often could also have affected the results.

The users were next inquired about how they feel when they forget someone's name and fail to address them by name when the other party greets them by name. 20% of the users responded they will be very embarrassed while 80% said they will be embarrassed. This response also validates the initial argument that a memory assistance system is important to overcome feeling embarrassed at these situations.

Next the users were asked if they face this situation, how difficult it is for them to ask the users for their name or where they met. 40% of the users classified these kind of situations as a problem while the rest of the 60% divided by 20% each among Very Big Problem, Yes and No and Little Problem. Since the opinion of the majority was that it is a problem it could be deduced that implementing such a system is useful in maintaining human relationships or building a good rapport with clients.

The following question was based on identifying important Memory Cues that would help them remember someone. 38% of the users responded they would like to know the event they met the person as the most important memory cue while name was the second most important choice. The third most important cue was the location while it was surprising to find out that date was the least important. The selection of Event as the most important cue as opposed to Name, agrees with the research survey done by Iwamura et al [14] where the users said event is the most important memory cue rather than name.

Afterwards, the participants were asked if they were willing to carry a memory assistance system. 60% of the respondents were willing to use a memory assistance system while the rest equally voted for If Convenient and If Implemented in a Mobile Device.

Then the users were asked if they would use the current implementation. While majority were happy to use the current implementation during the discussion it was identified still there is room for improvement. Results of this discussion will be given in the latter section named Post Questionnaire Interview with the Participants.

Unobtrusive presentation is the most important factor about the research and what distinguishes the current research among other researches available in terms of contribution. Hence the users were asked how they would like to see the results.

50% of the users wanted to see the results in the smartphone and later when further questioned they replied they would like to see the results in notification bar instead of result page for unobtrusiveness and they didn't find the current audio output to be important. The second choice was wrist watch (smart watch) whereas only one user wanted to see the results in Google Glass.

Finally, while the system was accepted by majority of the users and they were happy with the current implementation 80% of them noted this would be used in specific events such as business meetings.

The results proved the decision to develop situation specific lifelogging is a valid requirement in the era of lifelogging. This study also confirmed that while the current system overcomes the problems with using Head Mounted Display(HMD) or google glass to display results it needs to be made more unobtrusive to be widely accepted.

Post Questionnaire Interview with the participants

- Majority of the users preferred Smartphone display while Google glass was also preferred by one user

  Considering the background of the user who's in the HCI technical field it can be decided as the reason for preferring Google Glass implementation. But majority of the users who attend business meetings and travel frequently do not posses a Google Glass. If this system is to be accepted by a wider group of users it is important to consider what the majority of business users would use. Similarly, in [14] Iwamura et al says when conducting the user test for their system which involved a head mounted display; a half of respondents thought that the most appropriate device for the system was Smartphone. He also mentions the reason could be that the wearable devices were still not common even in 2014 when their research was published. Still, glasses would be a good option when the users are ready to accept it as an unobtrusive mechanism.

- Smart watch display option was rejected by users because

  - People do not wear watches
  - Feel its disturbing the communication to look at the watch

- During interview users mentioned Event and Name were the most important cues they would like to see
- Users would like to use the system in specific events such as Tech Events, Business Meetings
- It was discussed with users how they would like to add data in a real world system.
  The current system had the target images pre-trained. But when it comes to production use it's highly important to decide how the users will train the system with face images. So users were given three options to choose from.
  - Manually annotate images with user data
  - Business Card Reading
  - Checking newly added Facebook contacts and linking them with the app to obtain data

As expected, users were reluctant to accept the idea of manually adding related data such as name, event you met etc; for face images. But both Business Card Reading and Checking newly added Facebook contacts and linking them with the app to obtain data were accepted by users

5.5. User Test with Participants from Queens University, Canada

Two candidates volunteered to use the system for recognizing target faces and identifying the target's names. Both users were Masters Students at Queens University, Canada. They are both from technical background where one user has lived in Sri Lanka, worked in Japan and Canada while the other user is Canadian. Both were in the age group 26-30.

One user mentioned he face a situation where he meets somebody once a month while the other user said she faces it only once every three months.

Both users said they feel “Embarrassed” when they face a similar situation. And also it’s a “Problem” asking about the name and previous meeting. One user selected “Event”, “Name” and “Place” as important memory cues while the second user choose only “Event”. When asked about if they are willing to carry a system to help them remember people when they see them next time they responded “If Convenient” and “If Implemented in a daily-use device” respectively.

One user mentioned he would use the current system but “I would like to see this on a wearable device so that interaction is subtle.” The other user said she’s not sure.

It was surprising to see both users selected “Google Glass” as the method for presentation. This can be due to both of them being students with technical background researching on Human Computer Interaction. But during discussion it was revealed they do not have experience of wearing Google Glass for over a period of one hour. Hence, it’s clear their choice is biased towards their research interests rather than experience. In 5.4 it was revealed one user opted for “Google Glass” who has a similar background. Hence, it can be deduced “Google Glass” implementation will be a more appropriate presentation mechanism for those who are researching in “Human Computer Interaction” field.

Both users were happy to use the system in “Specific Events”. In order to obtain user data for annotating the images; one user preferred Business Card Reading while the second user preferred Searching newly added Facebook contacts and matching with new images captured by app and extracting name and other info from Facebook.

The following questions were newly included in the Questionnaire which didn’t exist in the previous user test.

1. What are the most important memory cues? Arrange according to your preference

    (a) Name

    (b) Event or Location

    (c) Time

    (d) Past Images

(e) Location displayed on Map

The previous Questionnaire didn't contain Past Images and Location Displayed on Map and users choose Event and Name. But when given two more options, Past Images and Location Displayed on Map; it was an interesting revelation that users thought Past Images were the most important memory cue to remember someone.

2. What are your thoughts on the User Interaction aspect of the system?

   - The interactions should be subtle as I dont want to get distracted by the system when I am in a conversation, yet want to use it to find the cues about person that I am talking to.

   - If I check my phone or see my watch when I talk to people, it is not polite. The interface should be invisible to other people.

3. What is your opinion about the recognition speed? Do you feel the current speed is enough for a real world application?

   - Current speed is pretty good for a real-time application.

   - Yes

4. Do you think notification mechanism and audio output contribute for unobtrusiveness compared with Google Glass?

   - Yes, but Google Glass could be used better to show visuals. Having said that, Google Glass can be socially awkward to use in a conversation.

   - Yes

5. Any suggestions for improvement?

   - I would like to see this implemented on a wearable device such as Google Glass or a wearable camera which will enable subtle interactions.

6. What do you think about Voice Activity Detection as a Trigger?

- I think it works well when you want to use the system without activating any physical buttons and such. Voice commands/activity detection is suitable for this project. However, it is important to consider that the system would be able to perform well in a noisy environment.

7. Do you feel using trigger based capture is more useful for memory augmentation related projects rather than current lifelogging systems like Autographer and Sesnsecam which captures in a timely manner?

    - Yes. The triggers are associated with interesting moments of an interaction or a conversation between users so the same triggers will have a higher chance to help users recall a situation/meeting compared to other systems.

    - If the trigger based capture is reliable, it is better than the other method.

The interesting revelations of the user test conducted at Queen's University, Canada among students of Human Computer Interaction is even though they feel some may feel socially awkward using "Google Glass" during a conversation, still they would prefer to see the results in "Google Glass". Hence, it is understood the user's opinions are biased towards their research interests and their personal preferences.

It was also interesting to find that users accepted "Voice Activity Detection" trigger based capture which is a novel contribution by the current research in the Lifelogging Devices.

Scrutinizing the findings of the two user tests conducted at Keio University Graduate School of Media Design and Queens University it can be concluded "Past Images", "Event" and "Name" were the most important memory cues. And different display mechanisms can be introduced depending on the user's background.

Chapter 6

Conclusion

Lifelogging is becoming popular among technical community and there's on going research on how to use the philosophy of Lifelogging for Human Memory Enhancement. However it was identified there's high amount of significance being given to increasing the technical features of the Lifeloggibg devices such as including many sensors, decreasing the sizes if the device and increasing storage capacity etc. However there's less attention being given to researching on how these lifelogs can be analysed and give productive output to the user rather than overwhelming the user with vast amount of data.

6.1. Selective Capture vs Total Capture

An experiment was carried out in order to find out the effectiveness of total capture vs designing lifelogging systems targeting specific scenarios. Autographer was used for total capture and images were analyzed on how useful these images are to automatically analyse and give an output. As explained in Chapter 5 autographer generated 120 images per hour but out of those 120 images only 2 images had images which contained faces. On the other hand Voice Activity Detection Trigger based system proved Selective Capture is more useful in which it contained 16 images which were useful for Face Recognition.

6.2. Voice Activity Detection as a Trigger

Voice Activity Detection is considered as the starting point of a conversation and data captured while talking to someone captures important information such as the face of the other person which can be used for creating a system that can help people remember better. The experiments also proved images captured based on Voice Activity Detection based trigger solved the issue of redundant space.

6.3. Impact of context for capturing a good picture triggered by Voice Activity Detection

An experiment was carried out by capturing context information such as Acceleration, Orientation and Ambient Light. The experiment was carried out in standard conditions as explained in Chapter 5. The results proved that context data did not have a significant difference in Recognized Images and Non Recognized Images. This proved the determinant factor in capturing a good image for face recognition is being able to detect the face.

6.4. Conclusion

Based on the experiments carried so far it was identified Voice Activity Detection Based Trigger solves the problem of redundant storage while supporting Selective Capture for Lifelogging. As future enhancement it is suggested to first detect the face before capturing the image to further enhance the quality of Capturing process.

High importance should be given to the Presentation as well and further experiments needs to be carried out analysing how this data can be produced to the user depending on situations.

The current system was able to recognize the person in 2seconds in the best case scenario and recorded 31seconds as the worst case scenario. With the current available technologies this performance can be accepted but in future with the advancement of technology the current face recognition module can be replaced by more robust and faster mechanism.

One of the concerns raised about the Face Recognition engine of the system was it's security, since the images are uploaded to the cloud. Performing the processing on the cloud eliminates the need for high end processor's on mobile devices and system can run on basic smartphones. Yet, there is room to investigate how the analysis can be performed on mobile phone itself, without uploading images to the cloud while maintaining the processing speed.

According to Gurrin et al [30] "As with all new technologies there are early adopters, the extreme lifeloggers, who attempt to record as much of life into their "black box" as they can. While many may not want to have such a fine-grained and detailed black box of their lives, these early adopters and the technologies that they develop, will have more universal appeal in some form, either as a scaled down version for certain applications or as a full lifelogging activity in the years to come." The presented reserach is a scaled down version of Lifelogging and while this will be able to appeal the early adopters and technologists, it will take longer to be accepted by the general community.

6.5. Future Enhancements

There are many other areas where the system can be tested. For example, at this moment, there is a tendency for experimenting ways for using Lifelogging for Alzheimer's disease patients. In around 2005 Microsoft [31] started a trial with a 63-year-old patient from the clinic with amnesia resulting from a brain infection. The patient was given a SenseCam and asked to wear it whenever she anticipated a "significant event" the sort of event that she would like to remember (i.e. not just something routine or mundane). After wearing SenseCam for the duration of such an event, she would spend around one hour reviewing the images every two days, for a two-week period. Without any aids to recall, she typically completely forgot everything about an event after five days or less. However, during the course of this period of assisted recall using SenseCam, her memory for the event steadily increased, and after two weeks she could recall around 80 percent of the event in question. This system required the user to manually start capturing any significant moments but the current voice activity detection trigger based capture would be more useful since the system decides the "interesting" moment

to capture the image.

Another viable area the system can be evaluated is, treating Clinical Stress. The first symptoms of patients with clinical stress is reserving oneself and moving away from others and not communicating. Since the Voice Activity Detection Trigger based system monitors user's conversations, this system can be used for treating clinical stress. This system can be used to monitor how often such patients communicate with other people and see their progression in meeting new people.

Based on user tests, it can be identified that future research can also focus on evaluating different presentation mechanisms depending on the user's background such as Technical Community and Business Community.

Although not explored in the current research, it is also interesting to see if activity recognition can be performed on the captured images. It will be a strong memory cue to remind what the other person was doing at the time of meeting again, to remember someone better.

With the novel introduction of Voice Activity Detection, in future with technological advancements it might be possible to extract background information such as name, place etc. from the conversation itself. This technology does not exist at the moment and out of the scope of the current project but would be of extreme use if can be implemented in future with technical advancements.

References

[1] Vannevar Bush. As we may think. *interactions*, 3(2):35–46, March 1996.

[2] M. Lamming and M. Flynn. 'Forget-me-not': intimate computing in support of human memory. In *Proceedings of FRIEND21: Symposium on Next Generation Human Interfaces*, Tokyo, Japan, 1994.

[3] G. "Bell and J" Gemmell. " : How e-memory revolution will change everything". 2009.

[4] Abigail J. Sellen and Steve Whittaker. Beyond capture: A constructive critique of lifelogging. *Commun. ACM*, 53(5):70–77, May 2010.

[5] Vaiva Kalnikaite and Steve Whittaker. *Human-Computer Interaction: The Agency Perspective*, chapter Synergetic Recollection: How to Design Lifelogging Tools That Help Locate the Right Information, pages 329–348. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[6] Episodic memory and semantic memory - types of memory - the human memory. `http://www.human-memory.net/types_episodic.html`. Accessed: 2016-05-13.

[7] A.M. Surprenant and Neath. chapter Systems or process, pages pp. 9–25. Hove, Psychology Press, 2009.

[8] Abigail Sellen, Andrew Fogg, Mike Aitken, Steve Hodges, Carsten Rother, and Kenneth R. Wood. Do life-logging technologies support memory for the past?: an experimental study using sensecam. In *Proceedings of the 25th ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 81–90. ACM, 2007.

[9] Matthew L. Lee and Anind K. Dey. Capture & access lifelogging assistive technology for people with episodic memory impairment.

[10] Robert S Wilson, Carlos F Mendes De Leon, Lisa L Barnes, Julie A Schneider, Julia L Bienias, Denis A Evans, and David A Bennett. Participation in cognitively stimulating activities and risk of incident alzheimer disease. *Jama*, 287(6):742–748, 2002.

[11] Vaiva Kalnikaite, Abigail Sellen, Steve Whittaker, and David Kirk. Now let me see where i was: Understanding how lifelogs mediate memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 2045–2054, New York, NY, USA, 2010. ACM.

[12] Steve Mann. Wearable computing: A first step toward personal imaging. *Computer*, 30(2):25–32, February 1997.

[13] Sreekar Krishna, Greg Little, John Black, and Sethuraman Panchanathan. A wearable face recognition system for individuals with visual impairments. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '05, pages 106–113, New York, NY, USA, 2005. ACM.

[14] Masakazu Iwamura, Kai Kunze, Yuya Kato, Yuzuko Utsumi, and Koichi Kise. Haven't we met before?: A realistic memory assistance system to remind you of the person in front of you. In *Proceedings of the 5th Augmented Human International Conference*, AH '14, pages 32:1–32:4, New York, NY, USA, 2014. ACM.

[15] S. Hodges, E. Berry, and K. Wood. Sensecam: A wearable camera that stimulates and rehabilitates autobiographical memory. *Memory*, 19(7):685696, 2011.

[16] The sensecam helps memory recall in alzheimers, howpublished = `http://seniortechdaily.com/the-sensecam-helps-memory-recall-in-alzheimers/`, note = Accessed: 2016-06-14.

[17] Omg life autographer, howpublished = `http://www.theverge.com/products/autographer/6165`, note = Accessed: 2016-06-14.

[18] Chris Davies. Narrative Clip 2 gets WiFi, Bluetooth and 8MP upgrade, year = 2015, url= http://www.slashgear.com/narrative-clip-2-gets-wifi-bluetooth-and-8mp-upgrade-04361493/, urldate = Jan 4 2015.

[19] Morgan Harvey, Marc Langheinrich, and Geoff Ward. Remembering through lifelogging: A survey of human memory augmentation. *Pervasive and Mobile Computing*.

[20] Matthew L. Lee and Anind K. Dey. Wearable experience capture for episodic memory support. *2012 16th International Symposium on Wearable Computers*, 0:107–108, 2008.

[21] Voice activity detector (vad) algorithm users guide. http://www.ti.com/lit/ug/spru635/spru635.pdf.

[22] G. Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.

[23] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, January 1991.

[24] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.

[25] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, August 1997.

[26] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin. Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 386–391, Dec 2013.

[27] Face++ privacy policy = www.faceplusplus.com, note = Accessed: 2016-05-14.

[28] Face recognition with opencv = http://docs.opencv.org/2.4/modules/contrib/doc/facerec/facerec_tutorial.html#conclusion, note = Accessed: 2016-08-01.

[29] ios sensor fusion, howpublished = https://sarofax.wordpress.com/2011/07/10/ios-sensor-fusion/, note = Accessed: 2016-06-17.

[30] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. Lifelogging: Personal big data. *Found. Trends Inf. Retr.*, 8(1):1–125, June 2014.

[31] Using sensecam to alleviate memory loss = http://research.microsoft.com/en-us/um/cambridge/projects/sensecam/memory.htm, note = Accessed: 2016-08-01.

Appendix

A. Source Code of Rotating Image After Resizing

```
//STEP 1: Get rotation degrees
 Camera.CameraInfo info = new Camera.CameraInfo();
 Camera.getCameraInfo(Camera.CameraInfo.CAMERA_FACING_BACK, info);
 int rotation = this.getWindowManager().getDefaultDisplay().getRotation();
 int degrees = 0;
 switch (rotation)
 {
  case Surface.ROTATION_0:
   degrees = 0;
   break; //Natural orientation

  case Surface.ROTATION_90:
   degrees = 90;
   break; //Landscape left

  case Surface.ROTATION_180:
   degrees = 180;
   break;//Upside down

  case Surface.ROTATION_270:
   degrees = 270;
   break;//Landscape right
  }
int rotate = (info.orientation - degrees + 360) % 360;
```

```
//STEP 2: Set the 'rotation' parameter
Camera.Parameters params = mCamera.getParameters();
params.setRotation(rotate);
mCamera.setParameters(params);
```

Figure 6.1: Source Code of rotating actual image

B. Source code of Resizing the Image by maintaining the Scale of the Image

```
bmpImage = getScaledBitmap(bmpImage, originalWidth, originalHeight,
            originalWidthToHeightRatio, originalHeightToWidthRatio,
            maxHeight, maxWidth);
}

private static Bitmap getScaledBitmap(Bitmap bm, int bmOriginalWidth,
int bmOriginalHeight, double originalWidthToHeightRatio,
double originalHeightToWidthRatio, int maxHeight, int maxWidth)
{
if(bmOriginalWidth > maxWidth || bmOriginalHeight > maxHeight) {

if(bmOriginalWidth > bmOriginalHeight) {
bm = scaleDeminsFromWidth(bm, maxWidth,
        bmOriginalHeight, originalHeightToWidthRatio);
} else if (bmOriginalHeight > bmOriginalWidth){
    bm = scaleDeminsFromHeight(bm, maxHeight,
        bmOriginalHeight,originalWidthToHeightRatio);
        }
    }
  return bm;
}

private static Bitmap scaleDeminsFromHeight(Bitmap bm, int maxHeight,
int bmOriginalHeight, double originalWidthToHeightRatio) {

int newHeight = (int) Math.max(maxHeight, bmOriginalHeight * .25);
int newWidth = (int) (newHeight * originalWidthToHeightRatio);
bm = Bitmap.createScaledBitmap(bm, newWidth, newHeight, true);
return bm;
}
```

```
private static Bitmap scaleDeminsFromWidth(Bitmap bm,
int maxWidth, int bmOriginalWidth,
double originalHeightToWidthRatio) {
//scale the width
int newWidth = (int) Math.max(maxWidth, bmOriginalWidth * .35);
int newHeight = (int) (newWidth * originalHeightToWidthRatio);
bm = Bitmap.createScaledBitmap(bm, newWidth, newHeight, true);
return bm;
}
```

Figure 6.2: Source Code of Scaling the Resized Image