December 16th, 2010

Dr. Andrew Rawicz
School of Engineering Science, SFU
8888 University Drive
Burnaby, British Columbia
V5A 1S6

Re: ENSC440 Post Mortem for a Facial and Speech Identification System (FASIS)

Dear Dr. Rawicz:

The attached document, *Post Mortem for a Facial and Speech Identification System (FASIS)*, outlines the process the Ztitch Solutions team has undergone during the development of the project for ENSC440 (Capstone Engineering Science Project). Our goal is to design and implement an automatic mobile log-in system based on the user's face and speech-spoken keyword, to eliminate the need for tedious typing on small devices.

The purpose of this document is to describe how the system works, problems we have encountered, comparison of estimated and actual budget, group dynamics, and what we would do differently if we were to undertake another similar project. Finally, an individually written description of contributions and things learned will be provided.

Ztitch Solutions consists of three motivated fifth-year engineering students: Andrew Au, George Liao, and Ching-Hsin Chen. By harvesting our individual skills and knowledge, the team has successfully developed a proof-of-concept system for the course. If you have any question or concern about our design specification document, feel free to contact me by phone (778-322-7928) or by e-mail (aau1@sfu.ca).

Sincerely,
Andrew Au
Fifth-year Computer Engineering Student
Enclosure: *Post Mortem for a Facial and Speech Identification System (FASIS)*

Post Mortem for a

Facial and Speech Identification System (FASIS) for Nokia Devices

Project Team: Andrew Au
George Liao
Ching-Hsin Chen

Contact Person: Andrew Au
aau1@sfu.ca

Submitted to: Dr. Andrew Rawicz – ENSC 440
Michael Sjoerdsma – ENSC 305
School of Engineering Science
Simon Fraser University

Issued date: December 16th, 2010

Revision: 4

Abstract

While both the demand of smart devices and the number of mobile internet usage continue to rise, there has been no innovation in the way users log into their online accounts. The traditional username and password scheme has been around for nearly two decades. Ztitch Solutions challenges this conventional username and password system by introducing the Facial and Speech Identification System (FASIS), which provides a quick mobile-login solution that eliminates the need for tedious typing on the small keypads or touch screens of modern smart phones, while maintaining a good level of security. The key success of FASIS lies in the fact that it leverages existing hardware of many modern smart phones, so no hardware modification is needed.

Table of Contents

List of Figures and Tables

Glossary

API Application Programming Interface
BOE Board of Education
FASIS Facial and Speech Identification System
IC Integrated Chip
I/O Input and Output
MCU Microcontroller Unit
OEM Original Equipment Manufacturer
PC Personal Computer
RS232 Recommended Serial 232
SAPI Speech Application Programming Interface
SDK Software Development Kit
UI User Interface
USB Universal Serial Bus

1. Introduction

This document, a post mortem of our Capstone Engineering Science project, will briefly describe the current state of the proof-of-concept system, derivations from the original design, comparison of estimated versus actual budget, group dynamics, and what we would do differently if we were to undertake another similar project. In the end, individually written statements, regarding the contributions and learning experiences, will be provided from each member of Ztitch Solutions.

For explicit details regarding the functions and design of FASIS, please refer to the functional specification and design specification documents.

2. Current State of the Project

Over the course of 6 months, starting from last semester (July 2010), the three members of Ztitch Solutions have successfully developed a working prototype of the FASIS. However, many more work lies ahead if this project were to be commercialized, especially in the area of collaboration with OEMs to adopt the idea, and in improving some of the software algorithms used. The following sections will give an outline of the entire project at its current state.

2.1. System Overview

As discussed in the functional specification document, the system leverages existing hardware of many modern smart phones; in particular, the front-facing camera and the microphone. Before using FASIS, the user must be registered with Ztitch Solutions and have a facial image database, as well as the default log-in information stored in the server.

The workflow of the system is as follows: the user wants to log into an online account and requests assistance from Ztitch Solutions so that he or she does not need to type in the log-in information manually. The user then provides a face image to the Ztitch Solutions server, and the server will process and the image and identify the user. Once the user is identified, it then requests for a voice-spoken keyword. If the two data validate, then the Ztitch Solutions server will help the user log into their online account. This process is visually summarized by figure 1.

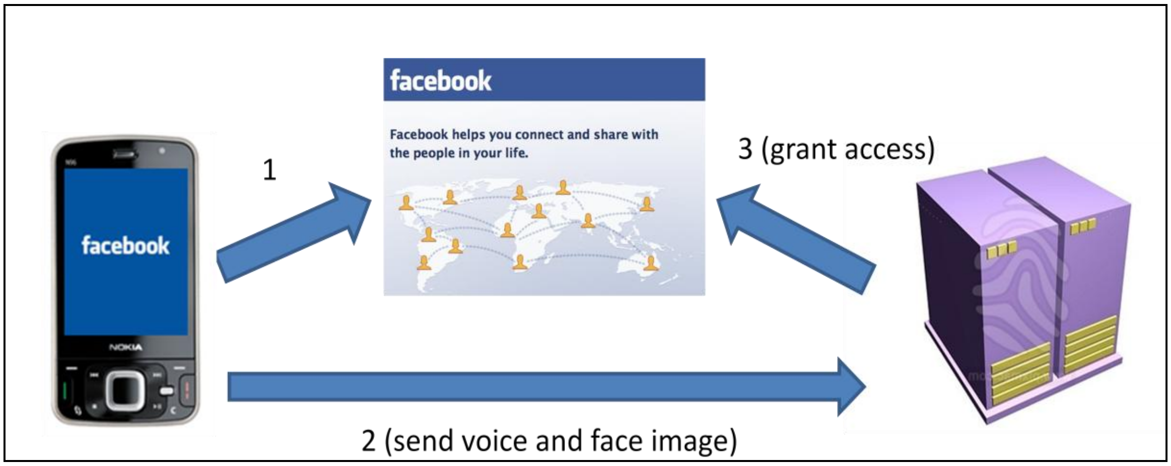

Figure 1. *A simple FASIS execution model.*

2.2. Face Localization

Ztitch Solutions proposes using skin color as a feature for tracking the location of the face in an image. The argument for this method is that color processing is much quicker than processing any other facial feature since color information is immediately exposed in an image (bitmap file). No additional calculation is needed to extract color information. Furthermore, treating the facial region as a color blob makes it invariant to rotation, which makes it very robust.

Ztitch Solutions' steps for skin color based face detection are simplified as follows:

1) Periodically capture an image from the camera
2) Filter the skin color to produce a binary image
3) Noise reduction
4) Dilation
5) Locate the coordinate of the blob

Figure 2 illustrates the five steps.

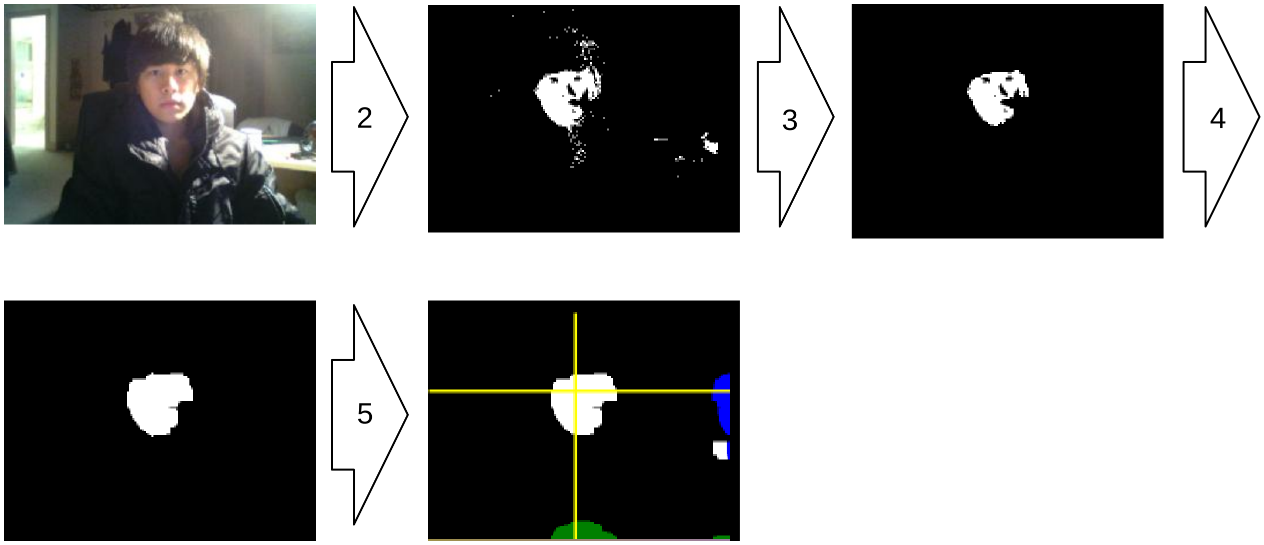

Figure 2. *Graphical flowchart of the skin-color based facial localization method*

2.3. Face Recognition

Ztitch Solutions adopts the Eigenface method of face recognition. We first prepare the images (T_i) by converting them to PGM greyscale images.

1) By concatenating the rows of pixels, each image is seen as one vector. A grayscale image with *r* rows and *c* columns is represented as a vector with *r* x *c* elements. All images are stored in a single matrix, **T**, where each row of **T** is an image.

2) From matrix **T**, calculate an average image **a**, and subtract it from each original image in **T**.

3) Calculate the eigenvectors and eigenvalues of the covariance matrix **S**. The eigenvectors of **S** are called Eigenfaces. They are the directions in which the images in the set differ from the mean image **a.** The challenge is to efficiently find **S** without actually computing **S** explicitly.

4) Choose the **principal components.** The *D* x *D* covariance matrix will result in *D* eigenvectors, each representing a direction in the image space. Keep the eigenvectors with the largest associated eigenvalue.

Matching a new, unknown face T_{new} to one of the known faces T_i involves two steps. First, T_{new} is transformed to the Eigenface, and then secondly, the weight is calculated. The Euclidean difference between two weight vectors is a measure of similarity between the two corresponding images x and y. If the Euclidean difference between T_{new} and other faces exceeds a certain threshold value, which we determine through trial and error, then the image can be one of the following: not the original face, or probably not a face at all.

2.4. Voice Recognition

Microsoft has provided a set of Speech API (SAPI) for 3rd party developers to incorporate speech recognition in their applications. The tool does not require the developers to have extensive knowledge in voice pattern science, and provides a high level API for Ztitch Solutions to use speech recognition in FASIS.

Ztitch Solutions implemented the SAPI runtime in the C# Visual Studio 2010 environment and integrated it in MATLAB. All the low-level details needed to control and manage the real-time operations of various speech engines were then handled by the DDI. Figure 3 provides a visual diagram of this implementation.

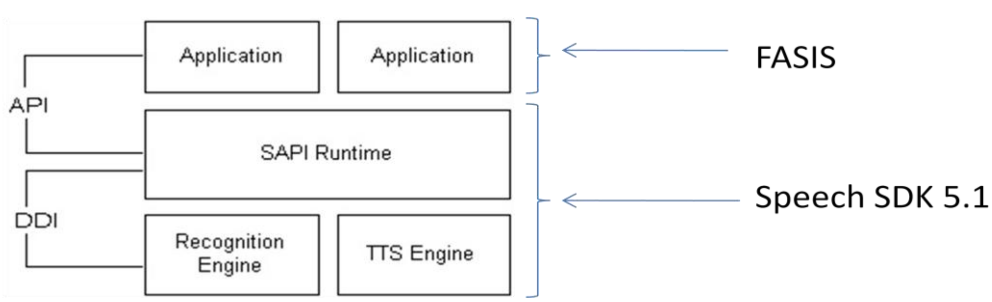

Figure 3. *An overview of the Speech API.*

2.5. Hardware

It is stressed that FASIS requires no hardware modification, but for demonstration purposes, a microcontroller was added to control the phone. In the phone's software, we wire each button of the numeric keypad to do a specific job, such as locking the phone, or unlocking the phone (each event would be visually displayed on screen).

We can easily simulate the key press of a button by sending a small current to the push button's circuitry. The general-purpose I/O pins P0-P15 of our microcontroller can each sink 25 mA and source 20mA. The HIGH command sets the specified pin to 1 (a +5 volt level) and then sets its mode to output.

For example, HIGH 14 sets PIN14 of the MCU to 5V. Combined with a 270 ohm resistor, we create 18.5mA. It also happens that 18.5mA is the same current required by the N96's internal MCU in order for the particular button to be considered push, since internally, there is a 270 ohm resistor connecting to the I/O ports of the phone's MCU.

There is also a transceiver module for relaying image and voice data from the N96 phone to the laptop. At the heart of this module is the Maxim MAX3222E, which is a 'true' RS232 transceiver. More of this will be discussed in section 3.3.

3. Problems, Deviations, and Recommendations

This section discusses the problems of the originally proposed system, and their solutions leading to the design deviations. At the end, recommendations will also be given as to what Ztitch Solutions should do differently if the team were to embark on a similar project.

There are three deviations of the final proof-of-concept system compared to the original proposed system. The first deviation is the addition of voice recognition, and the second being the way user authentication is handled. The third deviation deals with the hardware transceiver module.

3.1. Addition of Voice Recognition

The most noticeable deviation of the current system compared to the proposed design solution was the implementation of voice recognition as a secondary security measure (section 2.4.). In the original design, only the face image was required to gain authentication. However, critics of the original system pointed out that the original system would be vulnerable to intruders using a printed image of the real owner. This was indeed a security flaw of the design.

Therefore, a new design solution was adopted. The face image acts like the username in the conventional username/password scheme, and the spoken-keyword acts as the password. Although this is just an analogy, the idea is the same in that it maintains a good level of security. The name of the project was renamed from "Mobile Facial Identification System" to "Face and Speech Identification System."

3.2. Change in the Authentication Framework

The second change was made in the way the user's authentication is handled. Initially, as suggested in the proposal, the system would lock user out of the phone entirely and the user would not be to use the phone at all without authentication. However, the system is now an online-based solution for webpage login, such as accessing a Facebook account, an eBay account, etc.

The reason for this change stemmed from the fact that all processing would be done on the server side (on the Ztitch Solutions server). That means the user would only be able to use the phone if there were an internet connection. Of course, it would be inconvenient if the user cannot access his or her phone because of a lack of internet connection, so FASIS was modified to become a website log-in tool only. The phone's basic functions would not be compromised if there were no internet connection.

3.3. Modifying the Transceiver Module for Low Voltage Devices

RS232-compliant parts need a minimum of two supply voltages, one that is greater than +5V and one that is less (more negative) than -5V. These two supplies are needed to guarantee the minimum +/-5V output swings required for the transmitters. See figure 4. In systems that already have +/-12V, this isn't a problem. However, in the phone, +/-12V (or other voltages) isn't available. To solve this problem, Maxim has designed a wide array of parts that can be powered from a single supply by including power-supply converters built directly onto the IC.

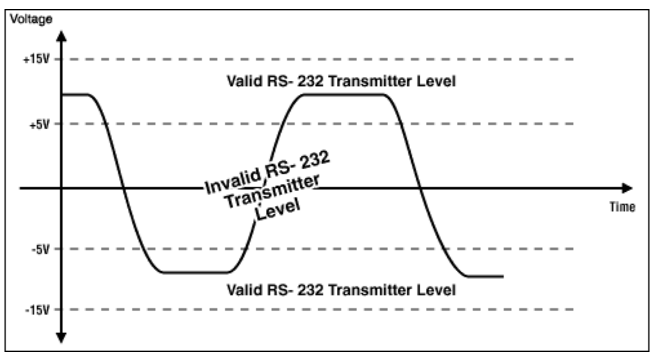

Figure 4. *RS232 transmitters must swing at least ±5V. This means that they must be powered with supplies ±5V*

Although single 5V-only parts are great in a lot of cases, an increasing number of applications require parts that will run from a single 3.3V supply. For instance, the battery voltage of a Nokia phone is 3.6V. Operation with 3.3V is not just important in 3.3V-only systems, but also where the RS232 parts must interface to 3V logic.

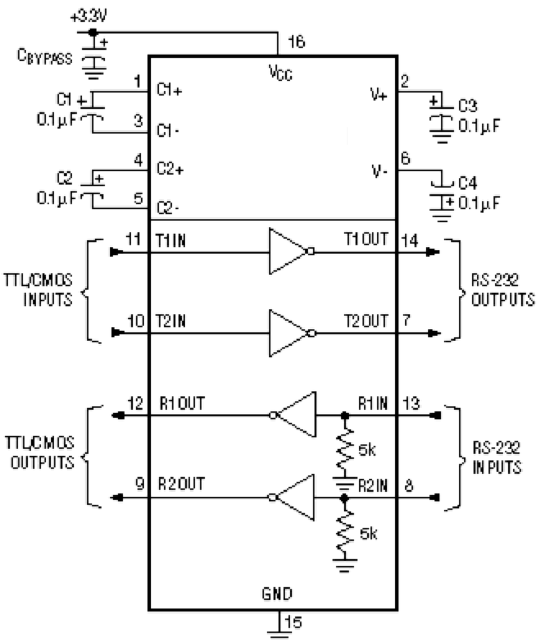

Figure 5. *RS232 application circuit for 3.6V using the MAX3222E chip.*

As shown in figure 5, the circuit has two charge-pump power supplies built on-board. These RS232 parts are special because of their low-dropout transmitters. These transmitters can meet the minimum-required +/-5V swings, while running from charge-pump power supplies as low as +/-5.5V. This makes it possible for these parts to run from a single 3.0V supply and still be fully compliant with the RS232 specification. Although these parts will run on supplies as low as 3.0V, they have also been designed to run as high as 5.5V.

3.4. Recommendations for Similar Projects

If the team of Ztitch Solutions were to undertake a similar project, some recommendations would be given as follows (in order of priority - 1 being most important):

1. Work on a device with better developer support. For FASIS, the team worked on a Nokia Symbian phone only because it was provided for free. If cost was not a

concern, then iPhone's iOS, or Google's Android should have been used. These platforms are fully supported by their respective companies: updates come more often, and many online resources (SDK's, guides, references, and examples) are available. Having better developer support also means that more API's are directly available to the developer. This could save us more time, without having to come up with our own solution to work around the lack of a specific API.

2. Work on a device with more customer reach. Currently, in the global smart phone market, iPhone's iOS and Google's Android operating systems dominate the majority of the market (significantly more so than the proposal's time of writing). Naturally, it would make more sense, business-wise, to work on either of these two devices.

3. For software, don't waste too much time researching or planning, but start coding as soon as possible. The reason for this is that certain things will become clearer, once the developer starts coding, that would otherwise not be during the research/planning stage. Software development is basically repetitive cycles of trial and error. Furthermore, the overhead cost of software development is very low (almost no cost at all, except for time), so even if the developer makes a mistake, going back and recoding is not very costly.

4. Think of potential problems early ahead, and ask for third party opinions. For example, if we thought ahead of the potential security flaw of section 3.1 and internet issue of 3.2, then we could have saved some time not having to redesign certain things.

4. Timeline/Budget

Table 1 provides a comparison of the initial estimated cost and the final actual cost of the FASIS project. Thanks to Nokia who provided us with a device, and sourcing online retailers for cheapest components, we were able to significantly reduce the cost of this project from 730 CAD to 250 CAD.

Table 1. *Comparison of the estimated cost and the actual cost*

Item	Estimated cost (CAD)	Actual cost (CAD)
Nokia N96 smart phone unlocked	$600	$0 (provided by Nokia)
Active SIM card plus subscription	$0	$0
Transceiver module	$100	$60
Parallax BOE	$200	$160
Software Tools, SDKs, drivers, etc.	$0	$0 (free for students)
Cases, cables, and accessories	$30	$30
Total Cost	**$730**	**$250**

Figure 6 is a Gantt chart illustrating the project schedule of FASIS. The project began early July of 2010 (the previous semester), so the team had plenty of time to research and to prepare the proposal and functional specification documents. In turn, this gave the team a lot of headroom for development and integration. Furthermore, each member placed this Engineering Science Capstone Project course at the highest priority compared to other tasks at hand. Therefore, in the end, Ztitch Solutions did not deviate from the originally projected schedule, and the proof-of-concept system was finalized for the demo in early December 2010.

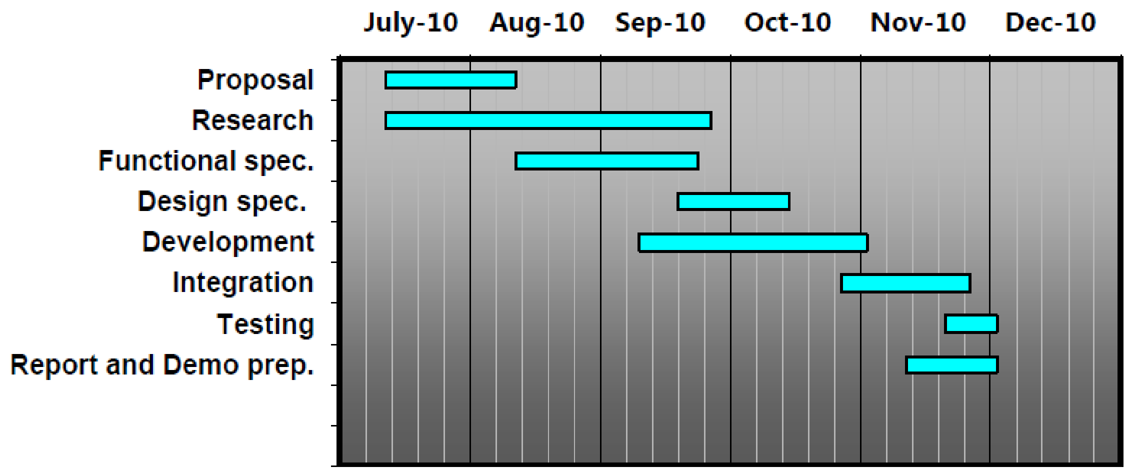

Figure 6. *Final Gantt chart of the FASIS project schedule.*

5. Future Work

A lot of work is in stored for the FASIS project. They are listed below (in no particular order):

1) Improve the facial recognition algorithm by adopting the Eigenfeatures method which is an enhanced variation of our Eigenface method, by taking into consider facial metrics, which is measuring the distance between face features.

2) Collaborate with Symbian to allow low level access of the operating systems. Special API's are needed so that hardware modification is not needed. For example, to make the project become a reality, we would need permission from Symbian to access certain native API's in order to handle specific aspects for the user, such as automatically helping the user authenticate into webpage.

3) Setup dedicated servers to handle thousands of users and to store user faces (training sets), as well as processing the face and voice recognition algorithms.

4) If successful with Nokia Symbian, we can expand our customer base by collaborating with the other mobile software giants, such as Apple, Google, RIM, and Microsoft.

5) Work on the details which were skipped in the proof-of-concept:

- Image compression
- Key encryption / decryption
- Reducing ambient noise during voice recognition
- Handling multiple failed attempts

6. Individual Reflections

Andrew Au:

While this project course is by far one of the most challenging courses of my undergraduate career, it remains as the most memorable. The initial learning curve was extremely steep, but once I got past certain milestones, the project became both rewarding and enjoyable. Getting to work with two of my best friends, whom I knew since our first semester at Simon Fraser University, was a major plus.

Although I was labeled the team leader, we were all contributing equally and there were many times I see either Ching-Hsin or George taking up the leadership role. Each one of us is dedicated to this project and commercializing it sometime in the future, and there is no turning back. Overall, my responsibilities and contributions in this project are as follows:

- Planning the FASIS system model from a high-level perspective, and partitioning the work load among the three members
- Writing the facial localization software to track human face based on skin color
- Writing the facial recognition (in collaboration with George)
- Writing the voice recognition system (in collaboration with Ching-Hsin)

Ching-Hsin Chen:

I consider myself to be very lucky to be able to work on this project. When Andrew and George first proposed this idea for our Engineering Science Capstone Project, I was a little skeptical regarding whether or not the idea was feasible and if we could finish it on time.

Through 6 months of hard work, starting from the previous semester, the dream became a reality and we have our prototype system working quite well. My expertise (from working at Broadcom) was in the hardware side, so in this project I contributed mostly in the hardware development phase. The following list is a summary of my contributions to the project:

- Implementing hardware (BOE kit and transceiver), and developing the gateway between the hardware and the phone itself
- Assisting Andrew in writing the voice recognition system
- Microcontroller scripting
- QA and debugging of the overall system

George Liao:

Looking back at the past 6 months, I found the project to be almost addictive. Despite taking four other courses at the same time, I often find myself spending the majority of the time on this project; not only because of the challenge, but also because I am quite certain that we will be able to take this project into the market.

One of the best things about our team is that while all three of us were best friends since we first came to SFU, we all remained professional in our work. We all contributed equally and there is no doubt that we can continue working together successfully, whether it is FASIS, or a new project in the future.
The list below outlines my main responsibilities:

- Develop the image processing framework in MATLAB
- Interfacing the hardware and software side
- Integrating image compression
- Researching the different possibilities of our design approach, and figuring out "the next steps"

7. Conclusion

If everything goes according to plan, the system may be able to surface in the market within a year or two. However, for this course, the major challenge of this project is developing the facial matching techniques and integrating them with the phone to successfully produce a prototype system. The end result of this Engineering Science Project is gaining valuable technical experience as well as interpersonal soft skills, such as team management and group dynamics. The purpose of this document is to serve as a post mortem to the project, explaining our problems, deviations, and recommendations. The document also examines the group's budget, timeline, and future work. Finally, individual reflections are given regarding this memorable learning experience.