EECE 401 Senior Design 2014

Project Report

**Sign Language to English Text**

**(SLatE)**

An embedded prototype for translating American Sign Gesture to English Text

Sarad Dhungel

Roshil Paudyal

Prajjwal Dangal

Yonathan Yilma

Claude Ndzami Kolloh

Advisor : Dr. Mohamed F. Chouikha

Department of Electrical & Computer Engineering

Howard University

April 20 2015

**SUBMISSION AND APPROVAL**

This project report is submitted for partial fulfillment of the Senior Design course describing the design and implementation of our project.

Sarad Dhungel 04/20/2015

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Name Signature Date**

Roshil Paudyal 04/20/2015

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Name Signature Date**

Prajjwal Dangal 04/20/2015

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Name Signature Date**

Yonathan Yilma 04/20/2015

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Name Signature Date**

Claude Ndzami Kolloh 04/20/2015

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Name Signature Date**

**Abstract**

About 28 million population of United States comprise of hearing impaired, deaf and mute community. In order to cope with the speaking community they are compelled to adopt written English as their primary language. Regardless of their effort of trying to adapt to the hearing community they are treated as handicapped and they feel discomfort to adopt to the society. With the motive to bridge this communication gap, our project SLatE proposes a solution by developing a prototype which understands, translates and help a person not knowledgeable with ASL by interpreting ASL to English. SLatE8’s effort of real-time image processing includes a hardware component (De2i-150 atom board) with Ubuntu Linux Operating system, implementing Computer Vision Library (OpenCV) and trained cascade of gesture database that utilizes the USB camera. The displayed sign were captured via camera, which were fed into the database where each images are queried with the database of the stored silhouette images. Based on their maximum match, each equivalent image are displayed on the Display screen attached with the board. Although training result shows adequate amount of result were achieved. However more variance on the database is yet to be added for robustness of the system. In addition, some adoptions of algorithms are required for ambiguous sign gestures.

**Contents**

1.Introduction……………………………………………………………………………………...**5**

2.Problem Definition…………...…………………………………………………………………**7**

3.Current Status of Art…………………………………………………………………………….**8**

4.Design Requirements…………………………………………………………………………..**10**

5.Solution Generation...………………………………………………………………………….**11**

6.Implementation………………………………………………………………………………...**13**

7.Performance Analysis and Evaluation…………………………………………………………**18**

8.Conclusion……………………………………………………………………………………..**23**

9.Recommendation………………………………………………………………………………**25**

10.References…………………………………………………………………………………….**25**

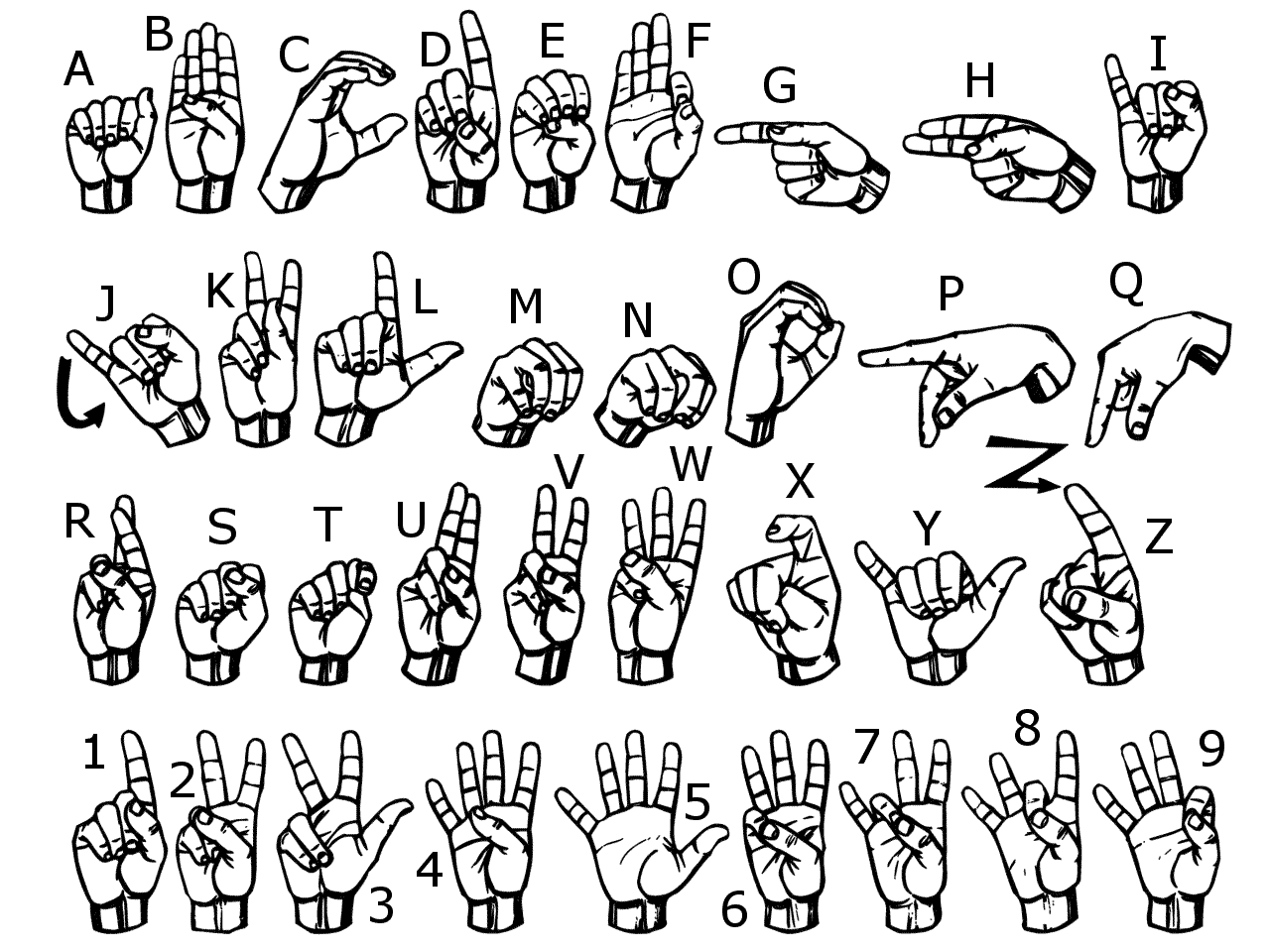
**Introduction**

Ever wonder what it is like to communicate without speaking? Or imagine a world without sound and the frustration of trying to express what you feel. You can try using hand gestures or home signs, but the problem is that they are not universal. You will need to learn a sign language that everyone understands, but unfortunately one doesn’t exist. Even a widely accepted sign language like the American Sign Language (ASL) is barely understood by anyone outside the hearing impaired community in the regions its used.

Today, ASL is the first language and main means of communication for hard of hearing and deaf individuals in the [United States](http://en.wikipedia.org/wiki/United_States) , most of Anglophone countries such as [Canada](http://en.wikipedia.org/wiki/Canada) as well as other regions of the world, including [West Africa](http://en.wikipedia.org/wiki/West_Africa)n countries and parts of [Asia](http://en.wikipedia.org/wiki/Southeast_Asia). We chose American Sign Language over other sign languages as it is widely used and comprises of easier sign spellings. American sign language (ASL) is a visual gestural language, which uses hand shapes, orientation and movement of the hands, arms or body to convey the meaning.

There is an issue that all in the hard of hearing or deaf community face in their everyday life, which is the communication barrier. For instance, social skills are a necessary component of everyday life, yet when deaf children are mainstreamed, this is one thing they cannot be directly taught. The communication barrier between deaf children and other children can cause deaf children to develop anxiety or low self-esteem.

To address this problem, over project Slate proposed an embedded prototype approach to help hearing-impaired people communicate conveniently with people unfamiliar with American Sign Language (ASL), and hence bridge the communication gap. This will also raise their self-confidence as they can better adapt into the broader community. Slate will employ Intel atom board (De2i-150) with an USB camera to stream video as input, which is then converted into text using image processing algorithms.

Img 1 - Alphabets and Numbers of American Sign Language

**Problem Definition**

The purpose of our design is to build a prototype device that will translate American Sign Language into written English. This device will bridge the communication gap between the hard of hearing community and the rest society that is not knowledgeable with sign language. In order to develop our system, some studies were made in analyzing ASL alphabet, in which we identified 26 letters, static signs (pose signs), and motion signs (Letter J and Z). In our recognition system, we want classify some signs upon characteristics of the hand shape, location, and size and be able to distinguish and translate the sign.

**Current State of Art**

Innovation requires one to consider the possibility that the product or service one wants to create might have similarities with something that is already available or already attempted. With this being said, one must do the proper research so that an idea is not duplicated. This also helps with efficiency. If a previous step in a series of steps has a known issue, one can capitalize on this by taking a different route to the solution or using one that is already available as to not make the same mistake and waste time trying to come up with a solution already available. Our team set out to be both innovative and efficient when creating a solution to the problem of communication between the deaf and/or mute and those who are not deaf and/or mute. In order to do this, we had to consider the current state of art.

First we looked at the sign language rings. [1]The sign language ring are a set of rings and a bracelet that detects sign language motions and translates them to voice. It can also translates voice to text. It uses three detachable rings that detect hand motion and a bracelet to display the text of the word gestured by the user. It can also be heard via the bracelets speakers. This was the winner of the 2013 reddot design award. This design was great. It was mobile which met one of our criteria. The disadvantage of this design was that a user that did not know sign language could not speak to someone who did know sign language. [1] The bracelets didn't show a user how to sign. Also, with four separate components, it left the possibility of one of the components getting lost. This was also not very aesthetically appealing. A small child would be irritated with a ring on his or her finger.

Next, we looked at the MyVoice. This was developed by the students at the University of Houston. [2] MyVoice is a concept that includes a handheld tool with a built-in microphone, speaker, soundboard, video camera and monitor. When placed on a hard surface, it reads a user’s sign language movements. Once MyVoice processes the motions, it then translates sign language into an electronic voice that can be heard. It can also capture a person’s voice and can translate words into sign language, which is projected on its monitor. [2] This can currently translate only one phrase. This concept was one that we set to capitalize on. It bad two way translation and was mobile. The disadvantage is that it can only translate one phrase due to the database being a very large file.

Finally we looked at Microsofts Kinect Sign Language Translator. This was developed by Microsoft China. It uses a device called the Kinect which was originally for Microsoft Xbox 360. [3]This device can detect motions of anyone in front of its camera. The engineering team in China used this to develop their product. This uses an avatar on the screen which can show a user how to sign any phrase or word written on the screen and can also translate sign language gestures to text and voice which would be displayed on the screen and or heard from the speakers of TV. [3] The disadvantage of this product was that it was not mobile. Our team set out to use the inspiration from this product to design something that was similar but mobile.

**Design Requirements**

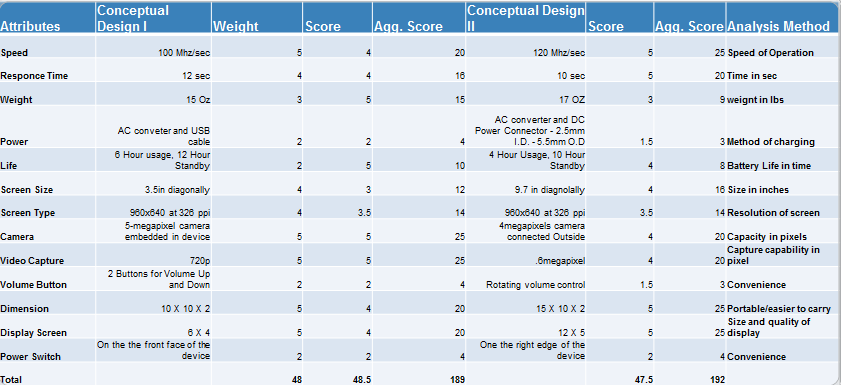
The design requirements included the Intel DE2i-150 Atom board at the core of our design. This was the requirement set by the Intel Cornell Cup committee. The Intel DE2i-150 Atom board came with the Yocto, host system running a supported Linux distribution. Our board used Ubuntu Linux as the operating system. Performance requirements include translation accuracy of at least 90%. This is after the realization that we would only do still image signs such as alphabets (excluding ‘j’ and ‘z’) and numbers. The time needed for accurate translation of a single sign should be under 3 seconds. We used Computer Vision Library (OpenCV) to perform all the image analysis. OpenCV is real time image processing software and we chose it because it can help us keep with our requirement of fast response time. The cost requirement for the project was heavily dependent on Intel DE2i-150 Atom board. We had the board offered to us through Howard University and Intel Cornell Cup. The requirements for the camera we wanted to purchase include the functionality of the webcam camera (720p and at least 640x480 pixel image capture capability) and cost (under $80). Other components of the overall device include standard mouse, keyboard, and a monitor with a HDMI or VGA connection capabilities.

FCC regulations stated the Specific Absorption Rate (SAR) of electronic devices (Electromagnetic radiation from DE2i-150 board). We met FCC compliance standard so that there was no electrostatic discharge, insured electromagnetic compatible, and followed all federal regulations. Finished product should meet the environmental requirements as stated in the most recent version of the following SAE standard: SAE Standard J1455, "Joint SAE/ Technology and Maintenance Council (TMC) Recommended Environmental Practices for Electronic Equipment Design". The size requirements are also heavily dependent on the size of the Intel DE2i-150 board and the monitor. This affects the our goal of compact, portable device but we can still deliver a system with a case that is put together as one device. The deliverable of this project is a working prototype that evaluates the desired functions and performances.

**Solution Generation**

**Selection of Top Design**

Table 1 - Comparison between two initial Conceptual Designs



**Top-Design**

Img 2 - Top conceptual design selected

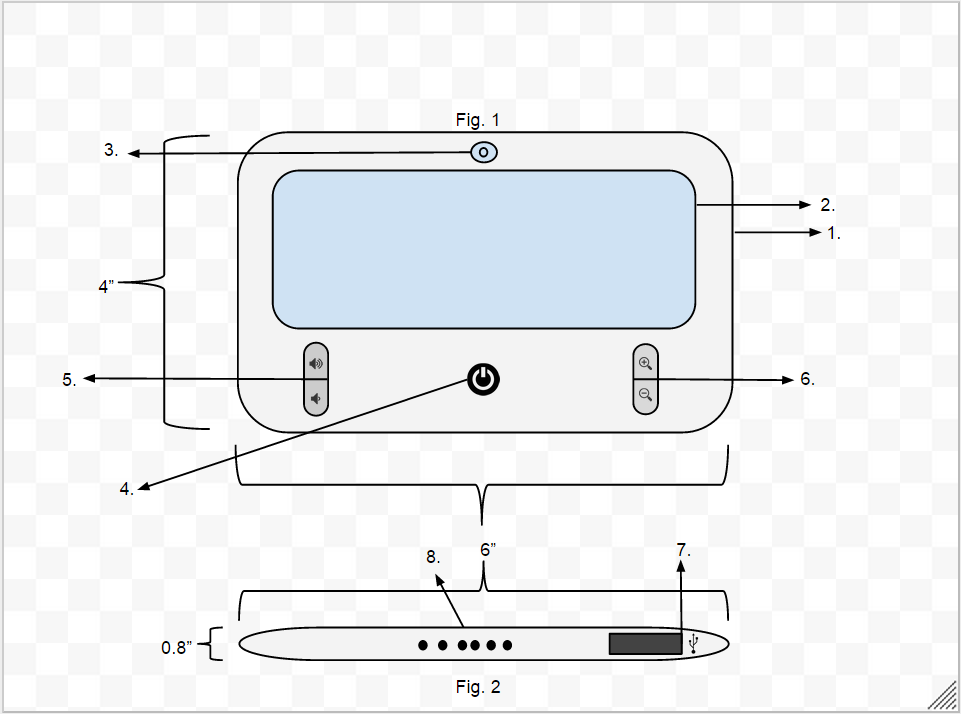


Table 2 - Parts of Initial Top Conceptual Design

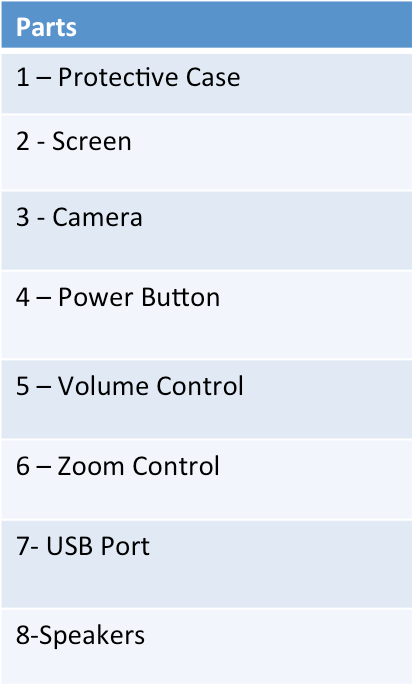
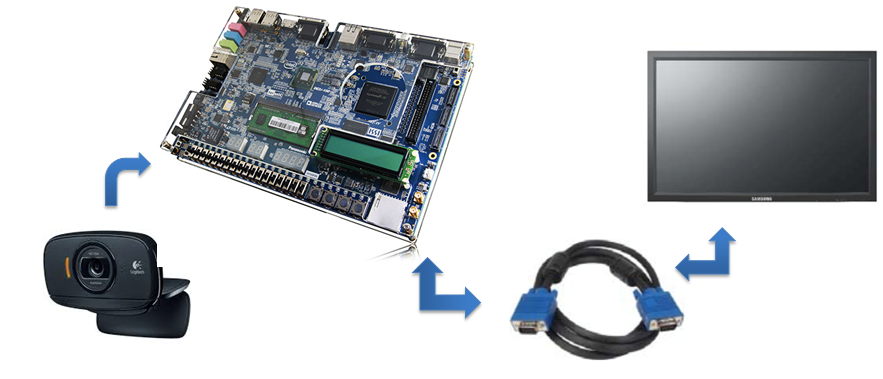


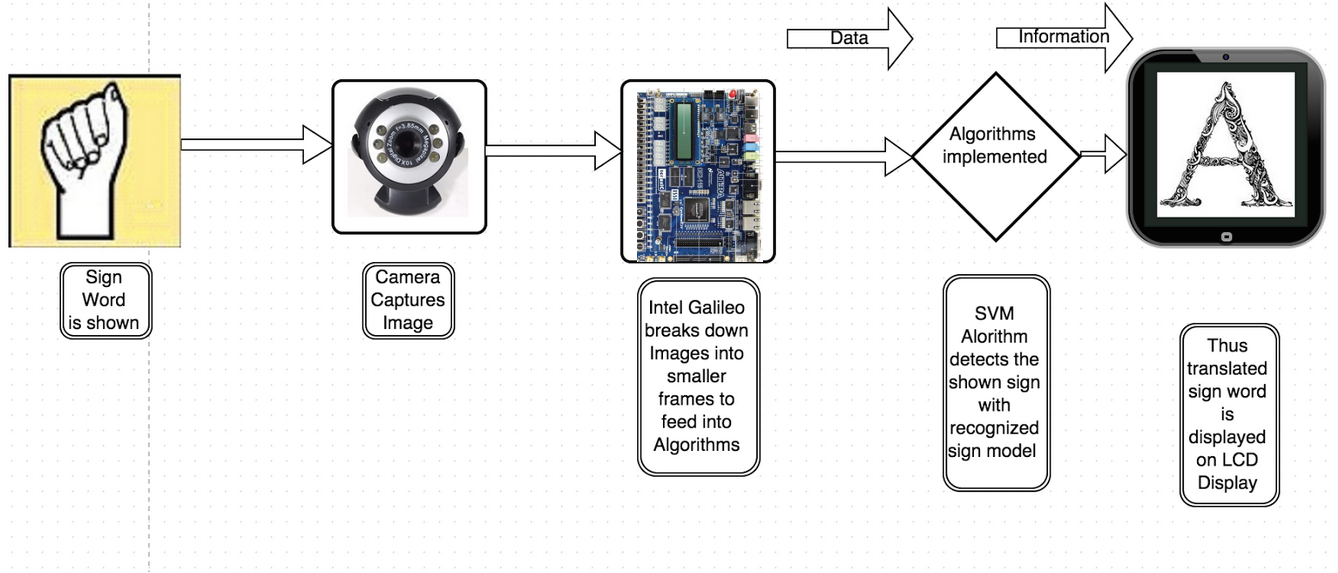
Image 3 - Hardware Components



**Implementation**

The final design of our prototype was modified over the time as per the requirement of the Intel Cornell Cup Competition. This approach is a short-term goal of our long-term vision i.e to develop a portable device capable of translating sign gestures in real time.

Image 4 - System Architecture

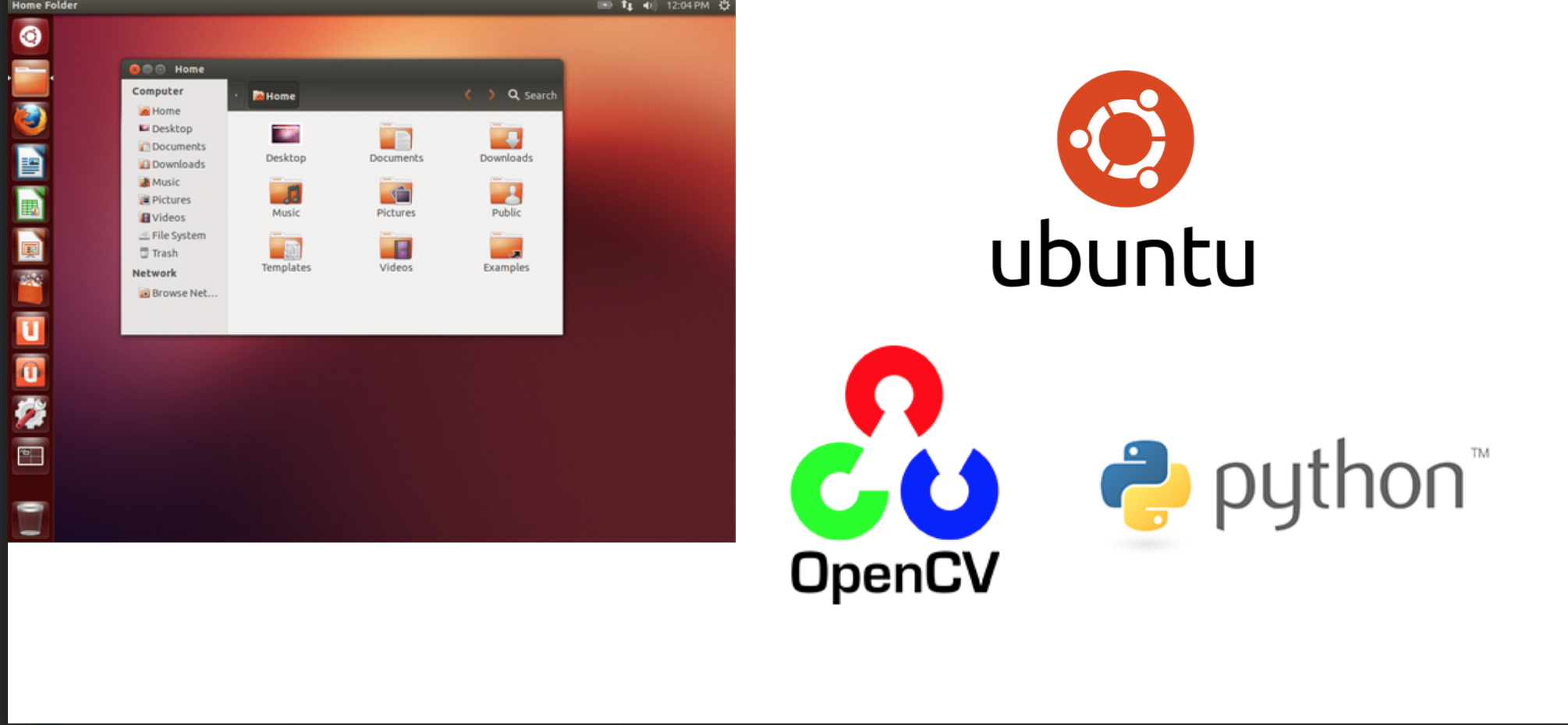


The above flow chart represents the working hierarchy of our system. The Hardware component constitutes of USB camera and De2i-150 Intel Atom board. Once the camera captures the images of each frames, then it send to the Atom board for further processing. Thus processed images data sets are then compared with stored attributes of the data set of the images. Then the corresponding sign gestured is then displayed on the screen. The majority of work such as hand detection, background segmentation and database comparison is done by various algorithms implemented on the Linux environment of the board.

**Hardware Architecture**

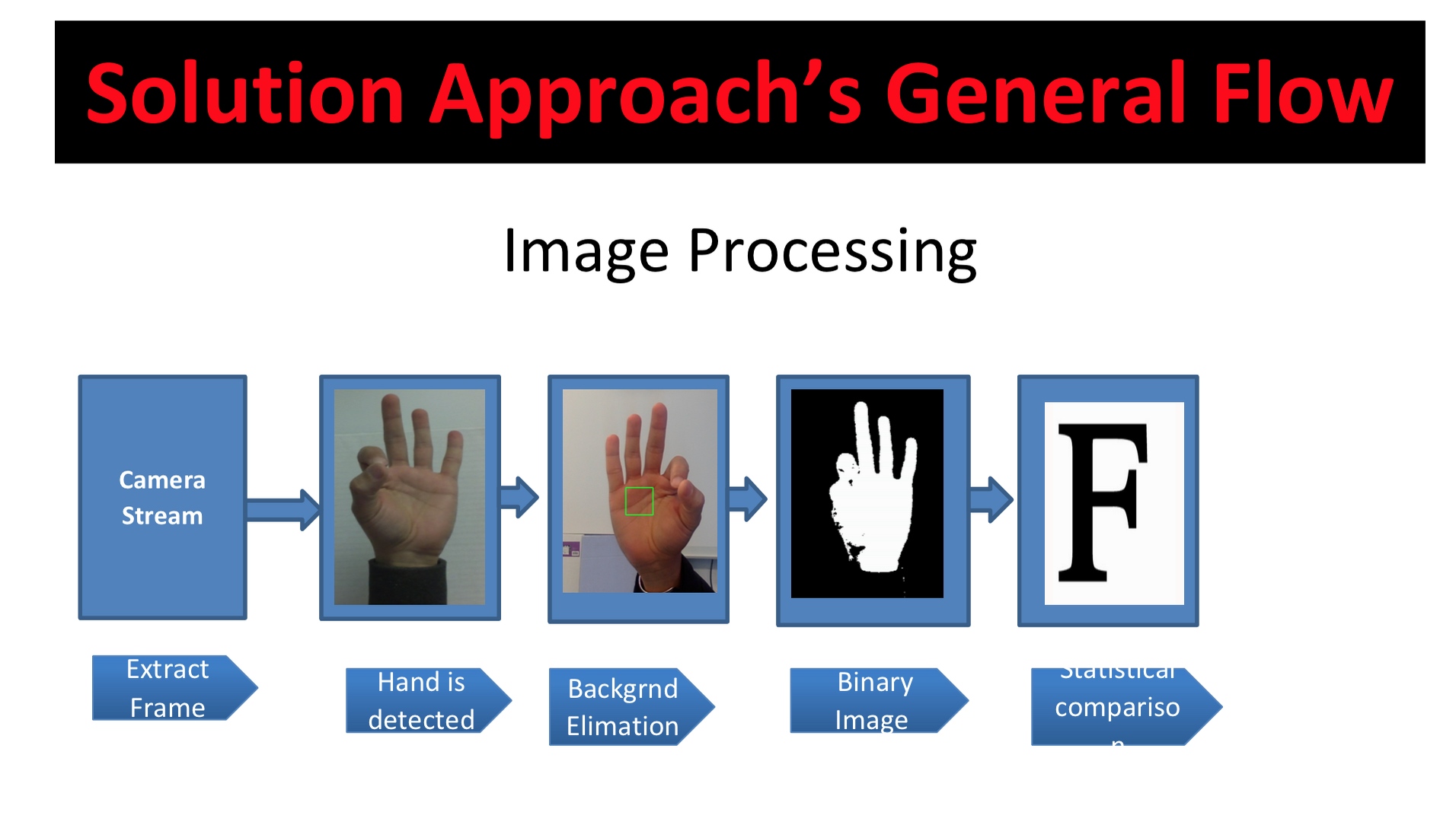
The De2i-150 Atom board came with inbuilt Yocto Project Linux which we lacked familiarity. Therefore, we imaged Ubuntu 10.04 Linux environment onto the Atom Board which eased the installation of computer vision library called OpenCV. The benefit of installing OpenCV is that the complex image processing functions could be easily imported into programming environment such as python, C/C++. We chose python over other languages because of familiarity and proficiency.

Image 5 - System Software Components



**Software Architecture**

Image 6 - Software Architecture



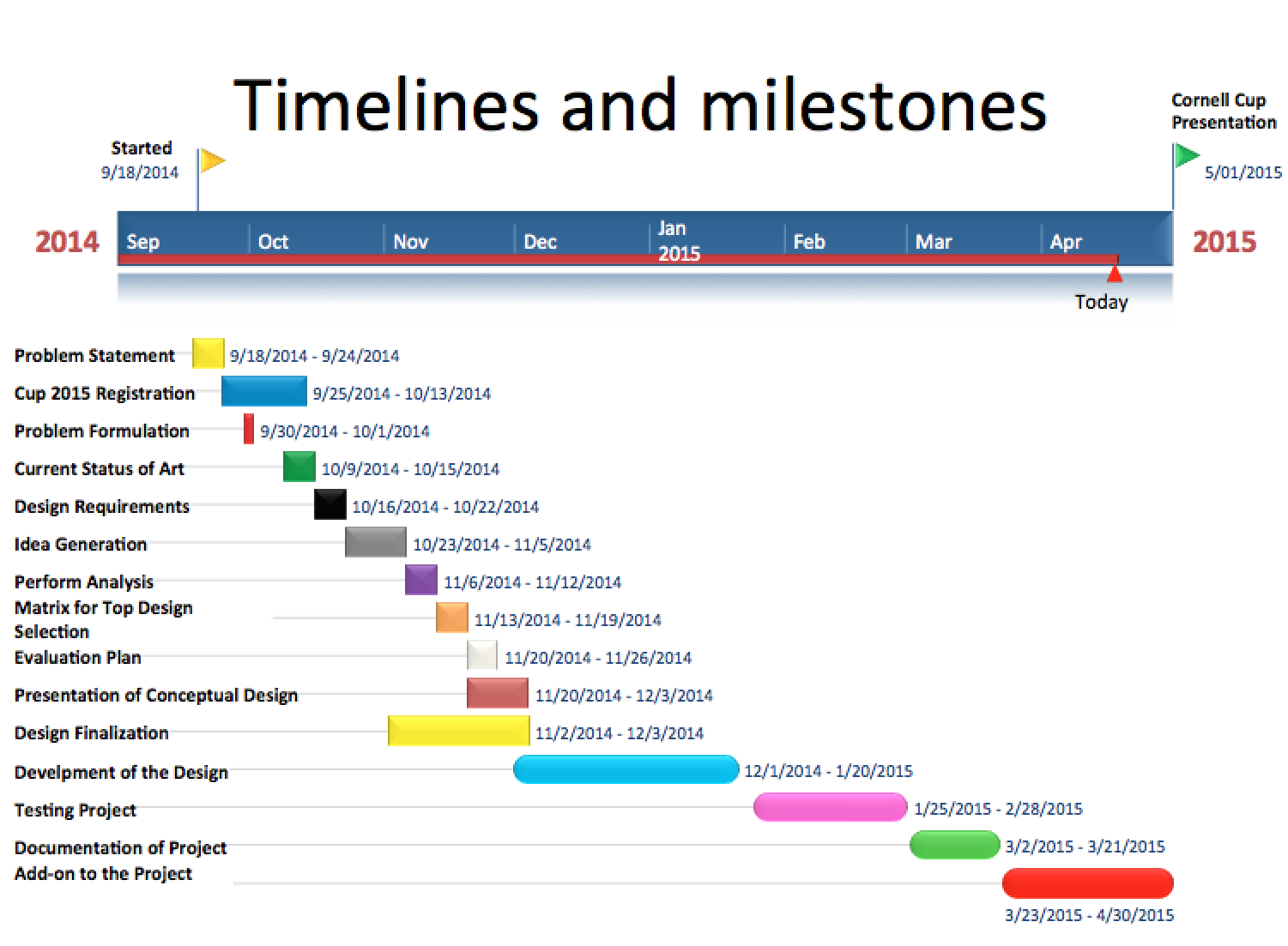
The frames of each sign letters are extracted from the video images. After each frame is extracted, it is crucial to distinguish hand from background so that our system could process. There are various color spacing that are applied for processing colored images. Initially, the image acquired from video are in RGB color model.[4] The degree of correlations between RGB components is lofty which produce a perceived sensitivity to illumination changes and would hinder the performance of the segmentation process especially for methods that rely on skin color in detection aim. [4] To overcome this obstacle, color spacing was chosen from available color spacing such as YCrCb, HSI/ HSL/HSV which have different characteristics. We chose Hue Saturation Color (HSV) color spacing over others because of ease of extracting features and because of robustness to lighting chances with the aid of white background. Thus, the hand was extracted from background using the Hue Saturation Value (HSV) method of skin color detection by rearranging the HSV ranges

* 0 > H > 360 ⇒ OpenCV range = H/2 (0 > H > 180)
* 0 > S > 1 ⇒ OpenCV range = 255\*S (0 > S > 255)
* 0 > V > 1 ⇒ OpenCV range= 255\*V (0 > V > 255)

The maximum and minimum value of the color profile is selected which is then compared with the threshold value to eliminate the background. After the background is eliminated the binary representation of the hand is created by selecting the largest contour, which is determined by comparing the HSV values of the neighboring pixels. After the largest contour i.e the region around the hand is detected then different features of the hand such as image moment, perimeter, minor axis length, extent, major axis, minor axis and solidity are extracted which are then statistically compared with values of same features of the stored hand gesture’s database. The algorithm would compare to the close proximity of the given features of each sign letters and then display the corresponding letter on the screen after comparison.

**Timeline for our Project**

Image 7 - Timeline of Project

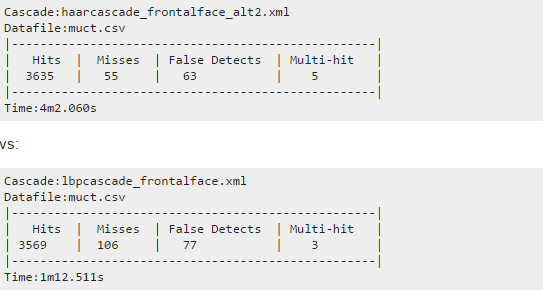


**Performance Analysis and Evaluation**

Initially while doing research of the project, we came across different resources that have done the sign language to text detection with different approach and yet research was undergoing for improving the accuracy. The motion sign gestures were still out of realm of detection. Since the project was sounding more ambitious, we wanted to test our idea priminarily to achieve some result.

We wanted to start with the approach which we found promising initially. We choose so-called HAAR classifier based on its accuracy on the facial detection. However this training method of Machine Learning was tedious as the training would require several hours to even days depending upon the number of images, ideally for robust system it required ( >5000 positive samples and >2000 negative samples). Positive samples are the images of interest which contains the images we want our system to detect whereas negative samples were the images of non-interest or ideally the one we don’t want our system to detect. We then switched to Local Binary Pattern (LBP) recognition approach upon considering the drawback of the HAAR classifier. Unlike HAAR classifier, LBP was very fast methods of training the system with approx 5-10% lower accuracy than HAAR. It would train the system in many levels of training in few hours for which HAAR classifier would take around a day or two.

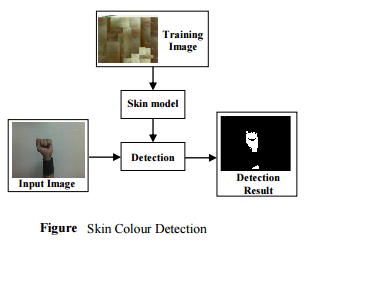
The table below is the comparison between the HAAR classifier and LBP classifier for the facial detection.



After training our system with LBP cascade system, our system was able to detect the hand with good accuracy in different background. However, while trying to distinguish between different still sign gestures, our trained system failed. Despite of training the system for hours, the system was not intelligent enough to distinguish between letters. Hence we had to come up with different approach for sign gesture recognition.

Then we decided to switch to skin color segmentation using HSV color spacing. This approach was simpler and faster means of gesture recognition. Basically this system would extract the feature of HSV values then compare with the threshold value then eliminate the background from hand by binarizing the hand from background.

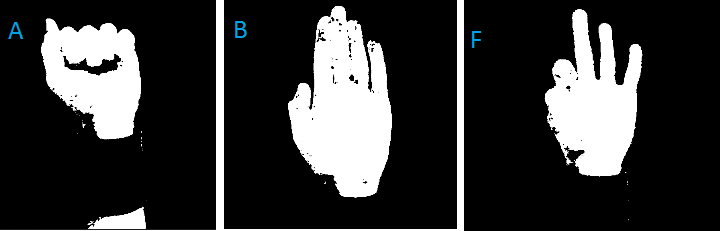
Image 8 - Colour Detection model



Once the hand is extracted from background, the different features of hand such as image moment, perimeter, minor axis length, extent, major axis length, minor axis length, and solidity are extracted. [5] The image moments are pixel intensities with greatest intensity in the region. Perimeter is computed by calculating distance between each adjoining pairs of pixel bordering the region. Extents are the ratio of pixels in the region to pixel in the total bounding box. Major axis length of the ellipse returns a scalar that specifies length (in pixels) of major axis. Minor Axis length of the ellipse returns a scalar which specifies length (in pixels) of minor axis. Solidity is calculated as Area/convex area which returns a scalar specifying the proportion of the pixels in the convex hull in the region.[5]

We applied the HSV approach to distinguish some letter like A, B, C, D, F, Q, Y which had the largest hand contour. The features of those sign letters after extracted were statistically compared with the features of features of images in databases.

Image 9 - Examples



Thus we were able to distinguish the Sign Letters A, B, C, D, F, L, Q, Y with accuracy >60%. However, letters A, B, C and F were detected with accuracy >80%. One of the reason for it, we had larger dataset for those four letters than remaining other letters. Yet we need to add more data set into our database to increase the robustness of our system for faster processing time and accuracy. Also, improvise our system to detect remaining sign alphabets. Thus after testing for the features extracted from database of hand, we are confident that we can increase detect remaining sign letters of the sign alphabets by time of Intel Cup presentation. The table below shows the accuracy measured in percentage of the sign letters four signs Letters A, B, C and F.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Overall**  **Accuracy** | Image Moment | Perimeter | Minor axis length | Major axis length | Solidity |
| **Database of Letter A** |  | 5.51948528217 | 0.691833887043 | 1.44656491632 | 0.722574688162 | 0.907768090671 |
| Sign A |  | 5.02273160677 | 0.62265049833 | 1.01259544142 | 0.58528549741 | 0.70805911072 |
| **Accuracy in %** | **82** | 91 | 90 | 70 | 81 | 78 |
| **Database of Letter B** |  | 7.8821463054 | 0.651832566238 | 1.93272981819 | 0.855741927189 | 0.904770162847 |
| Sign B |  | 6.14807411821 | 0.52798437865 | 1.39639729364 | 0.71026579956 | 0.7328638319 |
| **Accuracy in %** | **82** | 78 | 81 | 85 | 83 | 81 |
| **Database of Letter C** |  | 9.128985726396264 | 0.8017255601170044 | 1.2511731960429398 | 0.8217310117038028 | 0.60099888814810509 |
| Sign C |  | 7.39447843838 | 0.61732868129 | 1.02596202076 | 0.71490598018 | 0.48079911051 |
| Accuracy in % | **81** | 81 | 77 | 82 | 87 | 80 |
| **Database of Letter F** |  | 10.2822215018 | 0.431242580135 | 1.53924821043 | 0.760218224193 | 0.649429012001 |
| Sign Letter F |  | 8.22577720144 | 0.3191195093 | 1.20061360414 | 0.63098112608 | 0.57149753056 |
| **Accuracy**  **In %** | **81** | 80 | 74 | 78 | 83 | 88 |

**Conclusion**

Although not a complete success as to our original goal, the project did succeed partially and showed our team that the complete solution to our problem is possible. We originally set out with vision to have a portable device that offered two-way translation between those who use American Sign Language and those who do not in an accurate and timely fashion that could be incorporated into the lives of the average person. We decided to break down our ambitious project into hierarchy of goals and as a result our original idea was changed. Also, due to the Intel Cups requirements constraint, we had to use the De2i-150 that altered plan of creating portable device. As, we continued to work; we found that incorporating motion gestures would require more research and programming time than we had originally thought. With a deadline of April 30th, 2015, we had to change our goal from being able to translate all gestures of American Sign Language to just still image signs. Also, we changed the goal of the project from two-way communication into just a one-way communication (from sign language to English text). The approach we worked with for long time, using a trained cascade classifier to recognize and distinguish each sign, did not give us the results as we anticipated. The secondary approach, determining characteristics of each sign and statistically comparing the live feed image frame to that of database sign’s characteristics, was successful, and was tested with adequate accuracy. Hence allowed us to be able to translate 10 letters of the American Sign Language thus far. Our algorithm used Hue Saturation Value (HSV) color segmentation to determine the hand from the background taking threshold of sample color. Then binary image of each color profile was created which was then summed together to determine largest contour that is a hand and hence it is distinguished from the background by binary representation. If the hand gesture were not the largest object in the image, the hand gesture would not be detected. We used a clear background for our tests because it allowed the hand to be the largest object and we used long sleeve shirts to allow a cut off of the hand gesture at the wrist. After distinguishing hand from background, we extracted the features of the hand such as solidity, minor axis length, major axis length, perimeter, and image moment and then created database for each sign letters. Thus, the frames of sign gesture detected by camera would be compared to those features of each sign database. Then the one matching with higher percentage accuracy would be displayed on the screen as corresponding sign letters. We faced a challenge in trying to distinguish between signs that are similar because the characteristics used to determine each sign were analogous. Using the advancements we achieved, we hope to move forward, overcome the challenges, and present a fully complete still image American Sign Language to English text translator as we promised.

**Recommendations**

* Introduce more dataset into the database of gesture with currently low accuracy to bring up to the marginal accuracy.
* After careful considerations, we recommend that any future attempt to continue this project should consider increasing the library of letters and words.
* Incorporating motion gestures along with the still images would be a great improvement over this first prototype.

**References**

[1]RedDot Award Concept, <http://www.red-dot.sg/en/online-exhibition/concept/?code=1033&y=2013&c=16>

[2]University of Houston News

<http://www.uh.edu/news-events/stories/2012/may/0529MyVoice.php>

[3]Microsoft Research

<http://research.microsoft.com/en-us/collaboration/stories/kinect-sign-language-translator.aspx>

[4] Zhan Gao, “Appearance-based Hand Gesture Detection”.

[5] <http://www.mathworks.com/help/images/ref/regionprops.html>

[6] <http://wonderopolis.org/wonder/how-can-you-talk-without-speaking/>