“Image Text Extraction using OCR and Translation to American Sign Language”

LIUVENN ACE D. MAQUILING, Ateneo de Davao University

ERICK SEBASTIEN T. ESTABILLO, Ateneo de Davao University

Chapter 1

1.1 Introduction

OCR or Optical Character Recognition is the extraction of text from digital images either printed or hand written. OCR software attempts to replicate the combined functions of the human eye and brain, which is why it is referred to as artificial intelligence software [1].

There are many OCR techniques, an example of these techniques is the feature extraction method where the features found in the image are used to recognize the text in the image. There are already various existing OCR libraries and software in the present day. Some can already detect and extract different texts from different languages with different fonts but almost all the most accurate ones are for sale and there are only few that are open-source.

ASL (American Sign Language) is the most widely used sign language on earth. There are approximately 250,000 – 500,000 ASL users in USA and Canada [2]. The translation of words to ASL is not an easy process for non-experts and beginners.

1.2. Background of the Study

Optical Character Recognition (OCR) software was first used by libraries for historic newspaper digitization projects in the early 1990’s. An experiment at the British Library with the Burney collection and a co-operative project in Australia (the ACDP) [3]. OCR now has come a long way from where it started. It has now evolved into a great area of study in technology. It is now implemented on different fields of study such as data mining.

Sign languages are used all over the world as a primary means of communication by deaf people. American Sign Language (ASL) is one example, with broad uses in the United States. According to current estimates, in the United States alone, it is used regularly by more than 500, 000 people, and up to 2 million use it from time to time. There is thus a great need for systems that can interpret ASL (e.g., computer interfaces) or can serve as interpreters between ASL and English (e.g., in hospitals) [4].

1.3. Problem Statement of the Study

There are only a few people who have access to signers to translate for them when they are trying to interact with people who don't understand sign language. This is the problem that we are trying to address. There are only few signers, people who know how to use sign language, on earth. There are 400 forms of sign language around the globe and sign language is not a universal language unlike the American English that is always used by professionals. In this study, we are trying to focus on ASL.

OCR is already a widely research and we are trying to utilize this in order to get texts from images and translate them into sign language. Even though it is already a widely researched area, it is still not perfect. In accordance to this problem, we will try to fix this with an autocorrect feature of the project.

1.4. Objectives of the Study

The objective of the study is to use a development platform that has an integrated OCR feature that will extract text from images through the optical character recognition feature and to translate the extracted text through word per word translation then creating a correct sign language translation video output.

Specific Objectives

1. Use a development platform that has an integrated OCR feature.
   - To Program using C/C++/VB
   - To scan image for text with an OCR feature/library
2. To correct the misclassified words that were detected by the OCR.
3. Translate words recognized by the software.
   - Word per word translation
4. To create a series of videos that corresponds to the words detected.

1.5. Significance of the Study

This study will use OCR to extract the words properly then the text output goes through segmentation and classification that will create a grammatically correct output. This will increase the accuracy of the translation. With the segmentation and classification of the text, the translation of the text to sign language will not be complicated. This study will also use an accurate library for translation of the text. This will help deaf/mute users understand the written words that they do not know and this can also be a learning tool for both deaf/mute and non-deaf/mute users that wish to learn sign language.

1.6. Scope and Limitations

1.6.1 OCR Library Scope and Limitations

The scope of the study is the recognition of the universal English alphabet characters from an image that has at least 400 x 400 resolution with 18 as font size of the characters and specific font types, this will be discussed in the limitations. These measurements may vary with the relation of the resolution and font size because there are times when not both requirements are met, for example there are images with the correct font size but has a small image resolution.

The OCR that we used have certain limitations. One of its main limitations is the recognition of characters that have backgrounds that are very complex in the sense that the designs are not plain or one colored. This makes it difficult for the OCR to find the lines of the text for extraction because it disrupts the baseline of the text, therefore it will not recognize it as a line of characters. This can sometimes lead to the OCR library ignoring all of the characters thus there is no output. There are also times when there is a partial uniformity of the background for certain characters in the image, this makes the texts recognizable but since the background is not the same as some of the characters on the line, the OCR library only extracts the characters in the same background and excluding the characters that have a different background. This creates partial results that will always result in errors in the total extraction of the text since the text that was extracted has some missing characters.

Another limitation of the library is the typeface. There are only a few typeface that can be correctly recognized by the Tesseract OCR Library. An example would be the basic typeface Times New Roman. The types of typefaces that the library can properly recognize are the serif type styles, an example is the Rockwell font, and the sans serif type styles like the Helvetica. These font are recognizable because they are plain and have proper vertical alignment to the baseline. These fonts also do not have that much artistic characteristics, this is good since the designs sometimes create room for error, and this is because the OCR library tries to recognize the

designs of the text as part of the character creating more possibilities for different results that leads to error.

Another limitation is the alignment of the text. This is a problem where the text in the image is tilted. This is not the sideways alignment, it is when the text are tilted forwards or backwards. This makes the baseline of the texts undetectable. Therefore if this is the case there will be no output to be shown. These are the limitations of the TESSERACT OCR library.

1.6.2 Autocorrect Scope and Limitations

The autocorrect feature covers more than ten thousand words in the dictionary. But the autocorrect is also not perfect since, first limitation, it does not know what part of the word was incorrectly recognized by the OCR library.

Its accuracy significantly drops if there are two characters that the OCR incorrectly recognizes. This is because the number of possible words also increase significantly and since it cannot identify the characters that were incorrectly recognized, therefore it cannot exactly pinpoint the correct word to replace for the misclassified word.

It also autocorrects proper nouns and words that it cannot find in the dictionary. And if it cannot find a word that can replace it, it removes it from the text that the OCR primarily recognized from the image.

1.6.3 Translation Scope and Limitations

There are only 521 videos in our project corresponding to 521 words in the English dictionary. As long as the word is in the dictionary and has a corresponding sign language translation video, even if it was used as a name in the text it will be translated as a word in the dictionary and not a name of a person. There are also no available sign language translation videos for numeric figures. Even though they are not removed from the text, it will not be translated. The last problem is that there are words that have two or more meaning in accordance to what context they are being used, since the project does not know the context being used we cannot differentiate the meaning. In order to compensate for this, we picked the word that is easily understood or possibly applied in every context.

Chapter 2

2. Review of Related Works

Andrew Kae and Erik G. Learned-Miller stated that there are still some problems with OCR, many categories of documents continue to break modern OCR software such as documents with moderate degradation or unusual fonts. To address these problems the authors presented a form of iterative contextual modelling that learns character models directly from the document it is trying to recognize. Then use this method to segment and recognize the characters in an iterative process. The research is about recognizing characters by relying on the repetition of similar symbols, coupled with statistics of language, like English, to interpret a document. It is therefore a font-independent method of OCR, instead of relying on the font which is sometimes degraded or too noisy it will rely on symbols. Its approach to OCR is a form of iterative contextual modelling, building a document-specific model by first recognizing the least ambiguous characters and then iteratively refining the model to recognize more difficult characters [5].

Huenerfauth in this article talked about improving Current American Sign language animation. It is stated that current ASL animation technology cannot automatically generate expressions in which the signer associates locations in space with entities under discussion, nor can it generate many ASL signs whose movements are modified based on these locations. Its objective is to discover techniques for ASL generation that predict when to associate entities with 3D locations, where to place them, and how these locations affect sign movements. This technology will make ASL animations more understandable, enabling their use in accessibility applications for deaf individuals with low levels of English literacy. Current ASL generation technology cannot predict when a signer should associate an entity with a location in space or how to dynamically modify the movement of signs based on these spatial associations. The question addressed in this study is whether this limitation has an impact on the degree to which users understand and remember information from ASL animations [6].

Louloudis, Gatos, Pratikakis, and Halatsis discussed is about the segmentation of text to text lines and words. Text line segmentation is achieved by applying Hough transform on a subset of the document image connected components. The pre-processing stage of this paper consists of three steps. In the first step, the CCs of the binary image are extracted. The second step is the average character height AH for the whole document image is calculated based on the average height of all CCs. The final step is the partitioning of the CCs domain into three sub-domains. The post processing stage of this paper has two steps. The first is a merging technique

over the result of the Hough transform and the second deals with large components lying in the second subdomain [7].

Singh, Yadav, Verma, and Yadav created an OCR for different five fonts and sizes of printed Devnagari (an alphabet system of indian) script using Artificial Neural Network for the benefit of its application potential in banks, post offices, defense organizations and library sautomation etc. [8]. This article also discusses about using different approaches used in OCR systems:

- Matrix Matching
- Fuzzy Logic
- Feature Extraction
- Structural Analysis
- Neural Networks

Dreuw, Stein, and Ney studied on automatic sign language translation, one of the main problems is the usage of spatial information and its proper representation and translation, e.g. the handling of spatial reference points in the signing space. This article gives a way to solve the problems stated above and this is by obtaining manual features from tracking. This is called the tracking system - Relevant body parts such as the head and the hands have to be found for feature extraction, but most systems can only produce candidate regions. To extract features which describe manual components of a sign, the dominant hand is tracked in each image sequence. The translation system used in this article is the statistical machine translation so that it can automatically transfer the meaning of the source language sentence to the target language sentence [9].

Chapter 3

3. Research Methodology

3.1 Research/Project Design

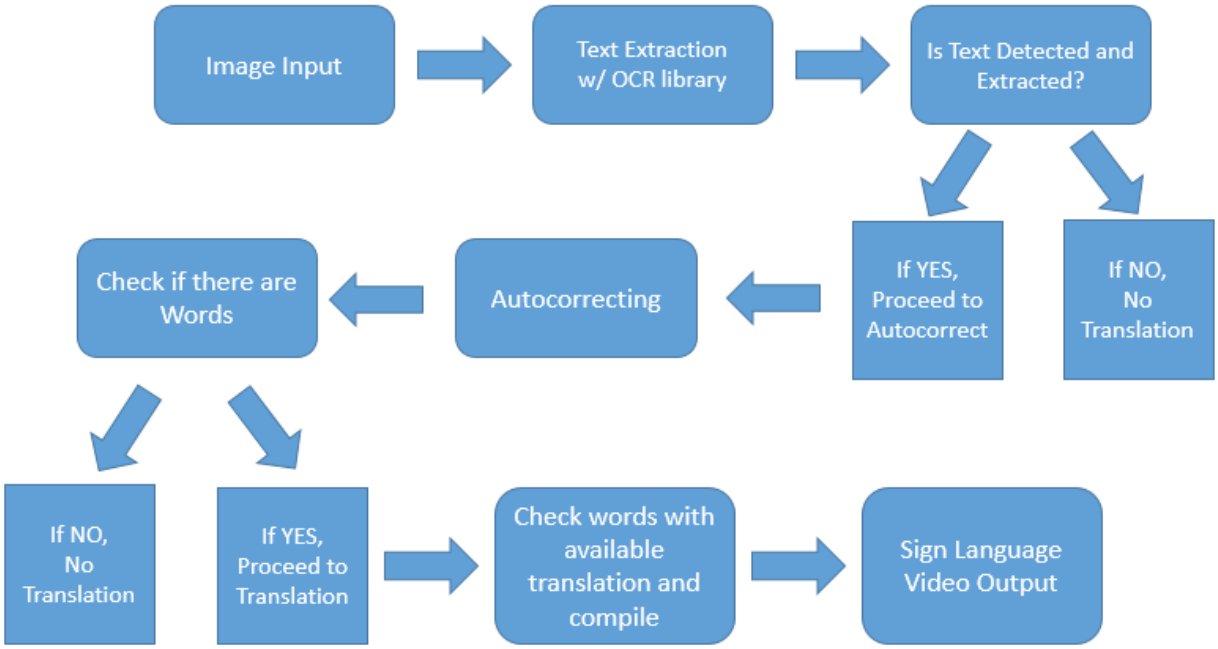


Figure 1. Conceptual Framework

3.2 Methodology

3.2.1 Data Gathering

We will be gathering images from the internet. These images will contain texts of not less than 3 or greater than 100 English words. These images will have different typefaces and backgrounds with different colors. They may vary in resolution but the ones with readable font size will be selected.

3.2.2 Digital Image Processing

We will be using a digital image processing library for the noise removal and filtering of the image for improving the accuracy of the OCR. This is where unneeded features of the image are removed in order not to disrupt the recognition of the text in the image which will increase the accuracy of the text output. This process will use different filters to remove noise. This will also greatly affect the output for the translation.

3.2.3 Text Extraction

The extraction of text from the image will be done by an OCR library. The following processes will be done by the library:

A. Line detection – the detection of the lines of the text in the image input.
B. Text Segmentation – The process where the words in one line will be segmented into words.
C. Text Extraction – The part where the words that were segmented will be compiled as an output text for the translation.

3.2.4 String Search and Auto correction

The string search is a process where the words that were extracted will be checked if they are in the English dictionary or not to verify the validity of the translation. The words from the extraction that are not the dictionary will be corrected using the autocorrect feature. We will try to search for the proper word/s to replace the word/s from the extraction that were not found in the library.

3.2.5 Sign Language Video Acquisition

The sign language video will be acquired from the internet. These videos may have variance in resolution and memory size, as long as the sign language in the video is understandable and clear but we will only get these from one source to have uniformity.

3.2.6 Sign Language Video Translation

The output of the string search and auto correction will be translated to sign language word per word. The output will be a compilation of videos from each word from the text output which means that the text and sign language will have the same sequence of words unless there is a word that does not have a corresponding sign language video, it will skip the word and move on to the next one.

Chapter 4

4. Theoretical Background

We will try to incorporate OCR, which is already a very well researched field in technology, with American sign language, which is only one of the many forms of sign language around the world, this is because sign language are sometimes native so that it will be more easier to understand. The most accurate OCR right now in this modern day is Omnipage standard, which is also costly. OCR software is used so widely that it is impossible to list all of its uses. In general, OCR is used to make documents text-searchable. The types

of documents for which OCR can be used are unlimited, ranging from contracts and invoices to articles and entire books. OCR is used in finance, education, healthcare, and the legal industry, among many others, and its uses grow with every year [10].

Chapter 5

5. Results and Discussions

5.1 Implementation Results

5.1.1 Data Gathering

Images of not less than 200 x 200 resolution were gathered from the following:

- www.pinterest.com
- www.twitter.com
- www.facebook.com
- www.tumblr.com

Google searches with keywords:

- “Quotes”
- “Sayings”

These Images have been classified into different categories so that the testing will also be categorized and these categories are:

- Images with black font and white background – 50 photos
- Images with black font and colored background – 50 photos
- Images with white font and black background – 50 photos
- Images with white font and colored background – 50 photos
- Images with colored font and colored background – 30 photos
- Images with colored font and white background – 10 photos
- Images with colored font and black background – 10 photos
-

5.1.2 Digital Image Processing

The digital image processing library that we utilized is the Leptopnica image processing library. Its main process in the image is Thresholding, this is the separation of the pixels in an image to a foreground and background which is based on a threshold that when it is higher than a certain pixel

intensity value it will be placed as a foreground and others will be a background.

Figure 2. An example of thresholding. Retrieved on September 29, 2014. Retrieved from http://mathematica.stackexchange.com/questions/16343/push-pin-art-convert-image-to-r-y-b-w-blk-color-space.

The image processing library is controlled by the OCR library that we are using. We will be unable to specifically show what the library does in every text extraction from an image since it is controlled by the OCR library.

5.1.3 Text Extraction

The OCR that we used is the TESSERACT OCR, it is called through the command prompt and then the output that is created is in the form of a .txt file. The library uses many classes and these are how the major functions work. First is the line finding, this part is where the library finds the lines of the potential baselines for the text, but there are always possibilities that the baselines are not perfectly horizontal lines. Some of these are sometimes ascending and descending, meaning it is a little bit diagonal. The main factor here in detecting these line are the height of the text that is detected, as long as it is similar and its height lines up with the other characters it can be detected as a line. After this, the lines that were found will be put in blobs. These blobs are the boundaries for the detected lines.

The next part is baseline fitting, this is where the blobs are combined with other blobs to see if they are on one baseline because there are cases that the baseline of other blobs are not perfectly aligned with its other components. Here it is analyzed to compare the location of each blob if they can be in one baseline.

After the baseline fitting is the fixed pitch detection and chopping, this is where the blobs are scanned if they have the same pitch, meaning they have the same format as the other characters in the same blob, like the spacing between the characters. Then after that, they will be divided into words which then will be chopped into characters. After that is the proportional word finding, this is the part where words do not have the same format, an example is that there is a line that has a different font size to the next line. This involves more on the vertical distance of the detected lines.

The word recognition is a major part. This part has 2 processes, the first one is the chopping of joined characters. This is where when they detect a word that has a joined part that does not follow the pitch of the characters, the OCR will try to chop the connected parts so that the characters will have the same pitch. And the second one, associating broken characters, is opposite to the first process, this is where there are unrecognized characters that do not have the same spacing as the other characters. In this part the broken characters are combined to see if they match any of the existing characters.

Linguistic analysis is where the recognized word is compared to a series of categories in order to create a more accurate result but unfortunately TESSERACT OCR only have few linguistic analysis.

The adaptive classifier is very important since TESSERACT OCR has a static classifier which can easily recognize words with the same font, its ability to distinguish between different characters and different fonts in one word or one line is weakened. This is where the adaptive classifier comes in, the adaptive classifier looks at the different characters from different fonts for every character. But the main function of the adaptive classifier is to distinguish lower case and upper case letters.

5.1.4 String Search and Auto Correction

The .txt file output will be read and checked if it is empty or not, if it is empty that means there was no text that was extracted from the image. If the .txt file is not empty, it will read the .txt file and display it in the textbox. This is the initial text for the translation.

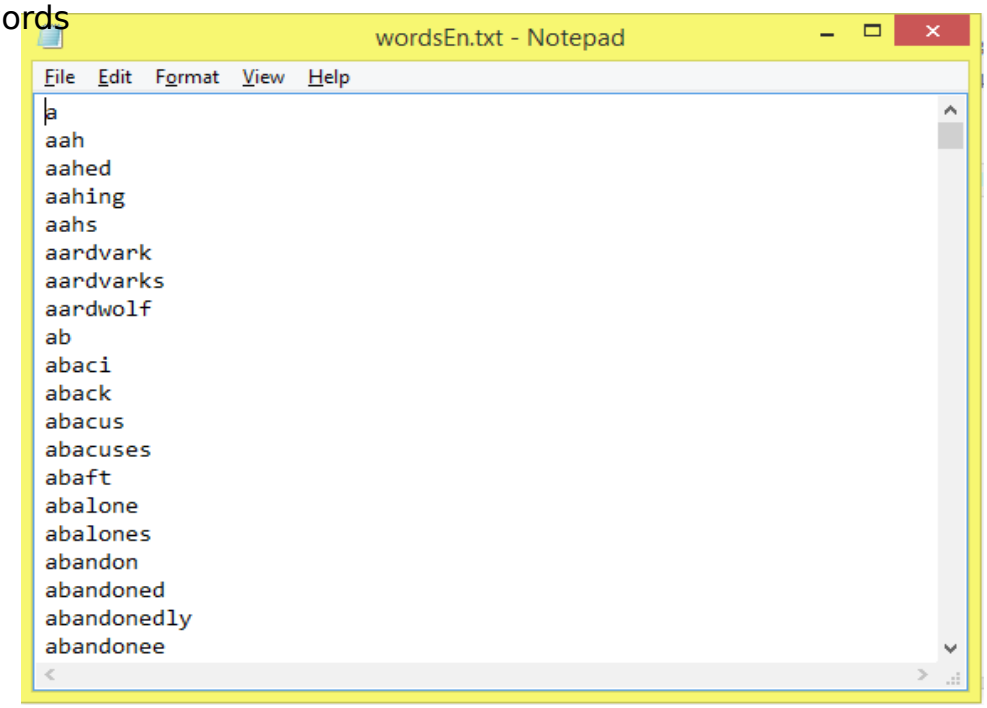
5.1.4.1 Text Cleaning

It will separate the initial text word by word and check if there are special characters in every word. It will iterate every character in every word then eliminate or remove all the special characters that were detected.

5.1.4.2 String Searching

The words in the cleaned text will be separated again word per word and each word will be checked to see if it is a valid word in the English dictionary. This is done by iterating through a .txt file, one word in every line of the .txt file that was retrieved from: http://www-01.sil.org/linguistics/wordlists/english/ which has all the English words in the English dictionary. It will only check if they have the same length and then compare if they are the same. The .txt file was also modified to add some words that are not in the dictionary but has a sign language translation. This is to prevent the word from being auto corrected.

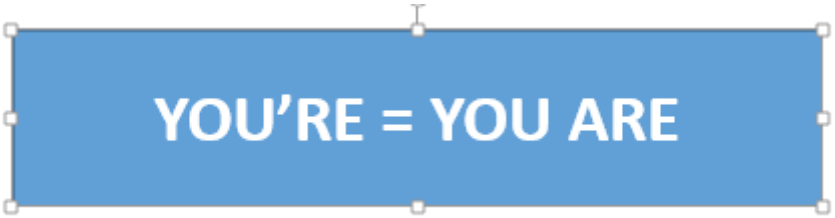
Figure 3. A snip of the .txt file that contains all the English words



5.1.4.3 Auto Correction

The auto correction part has 3 stages. If the word passes a stage it will not move on to the next. The following stages are:

5.1.4.3.1 Checking for Contracted words



The word will iterate through a .txt file that was created by us that has some of the contracted words that we usually use. As I mentioned earlier, we modified the .txt file which contains all the English words because there are some contracted words that already have a sign language translation. This .txt file that we created is for those contracted words that do not have a corresponding sign language video translation.

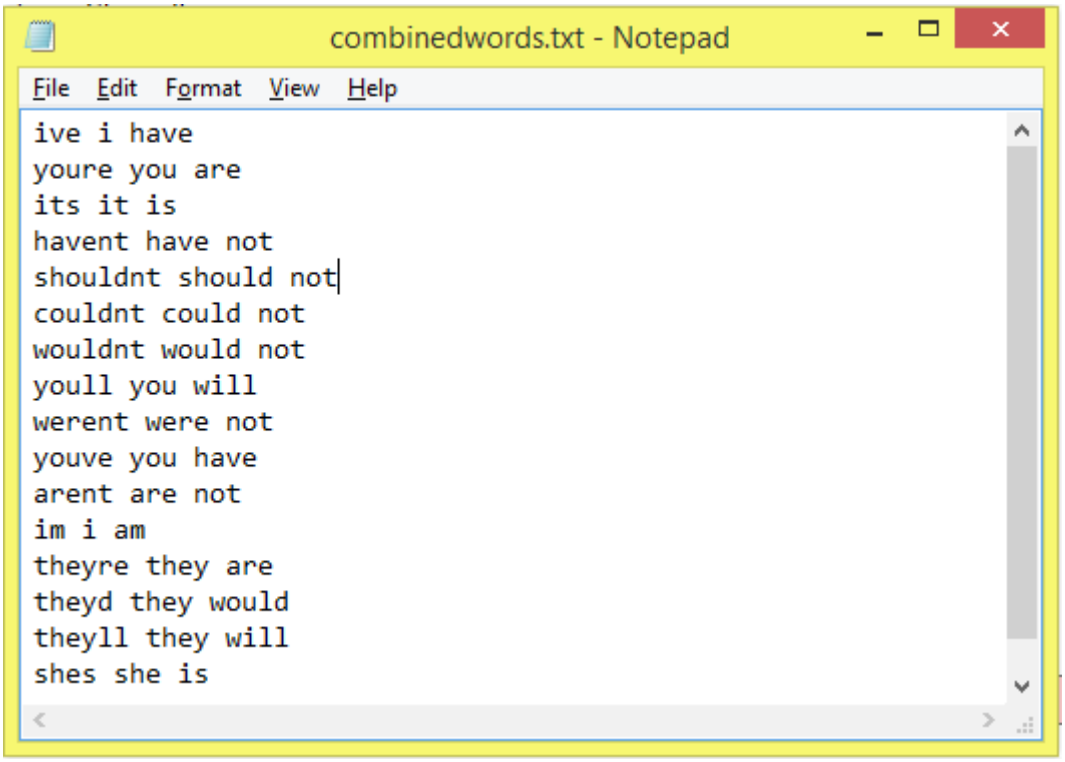


Figure 4. A snip of the .txt file of the contracted words.

It will compare it to the first word in every line. If they are the same it will replace it with the rest of the line which is the corresponding separated words of the contracted word. If it is not a contracted word then it will be passed to the second stage.

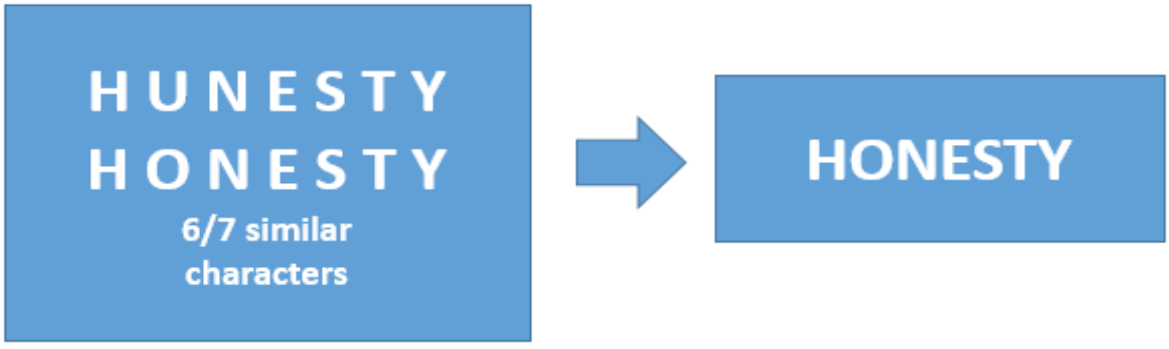
5.1.4.3.2 Separating of two combined words.

The word will be checked from the very first letter to the last. It will divide the word into two. The two words will then be checked if they are in the dictionary, it is the same process in the string search. If the two words are in the dictionary, the word that was not found in the dictionary will be

then replaced by the two words that was in the dictionary. It will continually iterate to the last letter, if it reaches the last letter without finding two words that are in the dictionary. It will be passed on to the third stage. The only problem is that the separation is first come-first serve which means that if there are more than one possible separation of the words, it will choose the first one it detects.

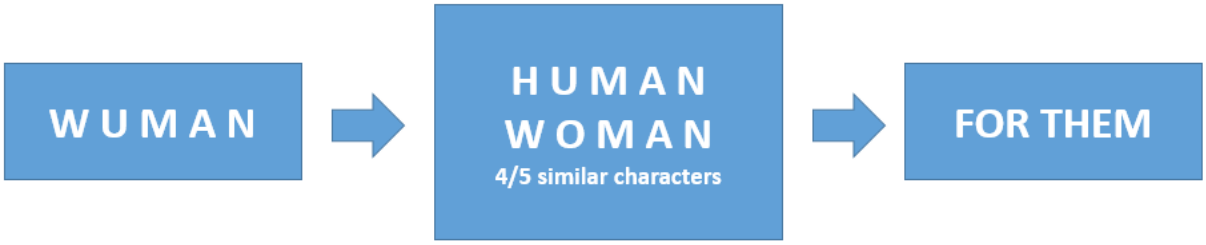
5.1.4.3.3 Auto correction

The word will be compared to all the words in the dictionary with the same length as the word. The words with the same length will be compared to each character with the same index of the word. The word that will replace the incorrect word is the word with the most similar letters to it.



.

It is also the same with the separation of combined words. The first one it detects with the same number of similar characters will be the one to replace the word.



Unless the similar character count is more than 60% than the length of the word, it will remove the word from the text that was primarily extracted by the OCR library. The purpose of this is to eliminate the words that has too much error or gibberish words.

5.1.5 Sign Language Video Acquisition

There are more than 500 sign language videos that are currently available for translation that we acquired from: http://www.signingsavvy.com/. These videos have low memory but the sign language is understandable and it is very clear. Almost all of the videos do not exceed the duration of 2 seconds. This is only appropriate because the translations that they are doing are for only one word.

5.1.6 Sign Language Video Translation

The sign language videos are named according to what word they are translated to, in this case we only have to iterate the word through the name of the sign language videos. If the word does not have a corresponding sign language video translation, it will move on to the next one. After the words are done iterating through the name of the videos, the videos will then be compiled into one single playlist. It will then be played as the sign language translation for the text in the image input.

5.2 Testing

5.2.1 OCR Testing

A. Images with black font and white background.

Picture Number	Number Of words	Number of words correctly	Number of words detected	Number Of Misclassified Words	Number of words not detected

		detected			
1	**38**	**38**	**38**	**0**	**0**
2	**9**	**9**	**9**	**0**	**0**
3	**15**	**14**	**15**	**1**	**0**
4	**14**	**14**	**14**	**0**	**0**
5	**7**	**7**	**7**	**0**	**0**
6	**18**	**18**	**18**	**0**	**0**
7	**10**	**10**	**10**	**0**	**0**
8	**46**	**45**	**46**	**1**	**0**
9	**38**	**38**	**38**	**0**	**0**
10	**16**	**15**	**16**	**1**	**0**
11	**13**	**11**	**12**	**1**	**1**
12	**69**	**67**	**69**	**2**	**0**
13	**15**	**15**	**15**	**0**	**0**
14	**20**	**20**	**20**	**0**	**0**
15	**27**	**0**	**6**	**27**	**0**
16	**19**	**0**	**0**	**0**	**19**
17	**16**	**16**	**16**	**0**	**0**
18	**13**	**5**	**17**	**8**	**0**
19	**34**	**32**	**34**	**2**	**0**
20	**15**	**15**	**15**	**0**	**0**
21	**32**	**32**	**32**	**0**	**0**
22	**41**	**37**	**39**	**4**	**0**
23	**17**	**14**	**16**	**2**	**1**
24	**20**	**19**	**22**	**1**	**0**
25	**19**	**19**	**19**	**0**	**0**
26	**25**	**25**	**25**	**0**	**0**
27	**14**	**14**	**14**	**0**	**0**
28	**19**	**19**	**19**	**0**	**0**
29	**22**	**19**	**22**	**3**	**0**
30	**5**	**5**	**5**	**0**	**0**
31	**17**	**17**	**17**	**0**	**0**
32	**75**	**75**	**75**	**0**	**0**
33	**15**	**15**	**15**	**0**	**0**
34	**38**	**38**	**38**	**0**	**0**
35	**15**	**15**	**15**	**0**	**0**
36	**12**	**12**	**12**	**0**	**0**
37	**34**	**34**	**34**	**0**	**0**
38	**92**	**92**	**92**	**0**	**0**

39	**7**	**7**	**7**	**0**	**0**
40	**28**	**28**	**28**	**0**	**0**
41	**8**	**7**	**8**	**1**	**0**
42	**21**	**21**	**21**	**0**	**0**
43	**28**	**9**	**21**	**12**	**7**
44	**19**	**13**	**18**	**6**	**0**
45	**19**	**18**	**19**	**1**	**0**
46	**46**	**34**	**46**	**12**	**0**
47	**20**	**19**	**20**	**1**	**0**
48	**28**	**28**	**28**	**0**	**0**
49	**13**	**13**	**13**	**0**	**0**
50	**8**	**6**	**8**	**2**	**0**
TOTAL	**1209**	**1093**	**1163**	**88**	**28**

In this category, out of the 50 images that we tested there are 19 images who has an error. But out of the 1209 words that were in the images 1093 were correctly extracted from the image while the number of words misclassified are only 88 and the number of words that were not detected were only 28. This leads to a 7% error rate in the misclassification of the text in this category and a 2% error rate for not detecting words.

B. Images with black font and colored background.

Picture Number	Number Of words	Number of words correctly detected	Number of words detected	Number Of Misclassified Words	Number of words not detected
1	**11**	**11**	**12**	**0**	**0**
2	**12**	**12**	**12**	**0**	**0**
3	**13**	**9**	**13**	**4**	**0**
4	**17**	**12**	**16**	**5**	**0**
5	**4**	**4**	**4**	**0**	**0**
6	**9**	**3**	**9**	**6**	**0**
7	**8**	**2**	**9**	**6**	**0**
8	**5**	**4**	**5**	**1**	**0**
9	**20**	**9**	**10**	**1**	**10**
10	**6**	**4**	**6**	**2**	**0**
11	**7**	**7**	**7**	**0**	**0**
12	**10**	**10**	**10**	**0**	**0**
13	**26**	**26**	**26**	**0**	**0**
14	**17**	**17**	**17**	**0**	**0**
15	**23**	**23**	**23**	**0**	**0**

16	**4**	**4**	**4**	**0**	**0**
17	**21**	**21**	**21**	**0**	**0**
18	**16**	**16**	**16**	**0**	**0**
19	**24**	**14**	**24**	**10**	**0**
20	**11**	**11**	**11**	**0**	**0**
21	**17**	**16**	**16**	**0**	**1**
22	**26**	**26**	**26**	**0**	**0**
23	**13**	**13**	**13**	**0**	**0**
24	**13**	**10**	**12**	**3**	**0**
25	**13**	**0**	**3**	**3**	**10**
26	**22**	**22**	**22**	**0**	**0**
27	**16**	**16**	**16**	**0**	**0**
28	**16**	**13**	**16**	**3**	**0**
29	**28**	**18**	**28**	**10**	**0**
30	**22**	**22**	**22**	**0**	**0**
31	**4**	**4**	**4**	**0**	**0**
32	**18**	**18**	**18**	**0**	**0**
33	**23**	**23**	**23**	**0**	**0**
34	**17**	**17**	**17**	**0**	**0**
35	**26**	**26**	**26**	**0**	**0**
36	**12**	**11**	**12**	**1**	**0**
37	**17**	**17**	**17**	**0**	**0**
38	**14**	**12**	**14**	**2**	**0**
39	**21**	**4**	**21**	**17**	**0**
40	**12**	**12**	**12**	**0**	**0**
41	**10**	**6**	**10**	**4**	**0**
42	**8**	**8**	**9**	**0**	**0**
43	**14**	**14**	**15**	**0**	**0**
44	**12**	**9**	**13**	**3**	**0**
45	**12**	**12**	**13**	**0**	**0**
46	**9**	**9**	**10**	**0**	**0**
47	**11**	**11**	**13**	**0**	**0**
48	**15**	**13**	**15**	**2**	**0**
49	**18**	**16**	**18**	**2**	**0**
50	**9**	**9**	**9**	**0**	**0**
TOTAL	**732**	**626**	**718**	**85**	**21**

Out of the 50 instances of the test set in the current category, 21 out of 50 images experienced an error. Out of 732 words in the test set 85 were misclassified and 21 words were not detected. This shows that there is a 12%

error rate for the misclassification of the text and a 3% error rate for not detecting words from the image.

C. Images with white font and black background.

Picture Number	Number Of words	Number of words correctly detected	Number of words detected	Number Of Misclassified Words	Number of words not detected
1	**12**	**12**	**12**	**0**	**0**
2	**7**	**7**	**7**	**0**	**0**
3	**8**	**8**	**8**	**0**	**0**
4	**27**	**27**	**27**	**0**	**0**
5	**14**	**12**	**14**	**2**	**0**
6	**14**	**13**	**14**	**1**	**0**
7	**11**	**8**	**11**	**3**	**0**
8	**12**	**7**	**13**	**5**	**0**
9	**14**	**13**	**14**	**1**	**0**
10	**3**	**0**	**0**	**0**	**3**
11	**24**	**24**	**24**	**0**	**0**
12	**17**	**17**	**17**	**0**	**0**
13	**14**	**9**	**14**	**5**	**0**
14	**12**	**11**	**12**	**1**	**0**
15	**13**	**13**	**13**	**0**	**0**
16	**18**	**18**	**18**	**0**	**0**
17	**20**	**17**	**22**	**3**	**0**
18	**8**	**7**	**8**	**1**	**0**
19	**14**	**14**	**14**	**0**	**0**
20	**15**	**15**	**15**	**0**	**0**
21	**10**	**10**	**10**	**0**	**0**

22	**7**	**7**	**7**	**0**	**0**
23	**20**	**20**	**20**	**0**	**0**
24	**14**	**9**	**17**	**5**	**0**
25	**12**	**1**	**9**	**8**	**3**
26	**11**	**11**	**11**	**0**	**0**
27	**8**	**8**	**8**	**0**	**0**
28	**10**	**10**	**10**	**0**	**0**
29	**11**	**9**	**11**	**2**	**0**
30	**23**	**21**	**23**	**2**	**0**
31	**12**	**12**	**12**	**0**	**0**
32	**21**	**21**	**21**	**0**	**0**
33	**18**	**14**	**17**	**3**	**0**
34	**20**	**17**	**20**	**3**	**0**
35	**9**	**9**	**9**	**0**	**0**
36	**11**	**11**	**11**	**0**	**0**
37	**18**	**18**	**18**	**0**	**0**
38	**5**	**5**	**5**	**0**	**0**
39	**18**	**18**	**18**	**0**	**0**
40	**20**	**18**	**21**	**3**	**0**
41	**24**	**24**	**24**	**0**	**0**

42	**21**	**4**	**14**	**10**	**7**
43	**29**	**23**	**27**	**4**	**2**
44	**25**	**25**	**25**	**0**	**0**
45	**14**	**14**	**14**	**0**	**0**
46	**5**	**5**	**5**	**0**	**0**
47	**17**	**7**	**15**	**8**	**2**
48	**22**	**16**	**22**	**6**	**0**
49	**15**	**14**	**15**	**1**	**0**
50	**3**	**3**	**3**	**0**	**0**
TOTAL	**730**	**636**	**719**	**77**	**17**

Out of the 50 instances of the test set in the current category 22 out of 50 images has an error. Out of 730 words in the test set 77 were misclassified and 17 words were not detected. This shows that there is an 11% error rate for the misclassification of the text and a 2% error rate for not detecting words from the image.

D. Images with white font and colored background.

Picture Number	Number Of words	Number of words correctly detected	Number of words detected	Number Of Misclassified Words	Number of words not detected
1	**12**	**12**	**12**	**0**	**0**
2	**7**	**1**	**5**	**4**	**2**
3	**12**	**12**	**12**	**0**	**0**
4	**16**	**16**	**16**	**0**	**0**
5	**3**	**3**	**3**	**0**	**0**
6	**18**	**18**	**18**	**0**	**0**
7	**16**	**16**	**16**	**0**	**0**
8	**20**	**19**	**20**	**1**	**0**
9	**11**	**0**	**3**	**3**	**8**

10	**18**	**0**	**0**	**0**	**18**
11	**5**	**5**	**5**	**0**	**0**
12	**14**	**14**	**14**	**0**	**0**
13	**16**	**0**	**10**	**10**	**6**
14	**4**	**3**	**4**	**1**	**0**
15	**10**	**9**	**10**	**1**	**0**
16	**11**	**0**	**0**	**0**	**11**
17	**9**	**7**	**9**	**2**	**0**
18	**5**	**5**	**5**	**0**	**0**
19	**11**	**11**	**11**	**0**	**0**
20	**16**	**5**	**12**	**7**	**4**
21	**12**	**12**	**12**	**0**	**0**
22	**13**	**12**	**13**	**1**	**0**
23	**7**	**7**	**7**	**0**	**0**
24	**12**	**0**	**1**	**1**	**11**
25	**13**	**13**	**13**	**0**	**0**
26	**10**	**9**	**10**	**1**	**0**
27	**20**	**20**	**20**	**0**	**0**
28	**21**	**21**	**21**	**0**	**0**
29	**5**	**2**	**4**	**2**	**1**

30	**12**	**6**	**12**	**6**	**0**
31	**15**	**15**	**15**	**0**	**0**
32	**15**	**12**	**15**	**3**	**0**
33	**12**	**12**	**12**	**0**	**0**
34	**59**	**58**	**59**	**1**	**0**
35	**32**	**31**	**33**	**1**	**0**
36	**9**	**9**	**9**	**0**	**0**
37	**20**	**20**	**20**	**0**	**0**
38	**14**	**14**	**14**	**0**	**0**
39	**19**	**19**	**19**	**0**	**0**
40	**10**	**0**	**0**	**0**	**10**
41	**33**	**30**	**34**	**3**	**0**
42	**21**	**19**	**20**	**1**	**1**
43	**16**	**12**	**17**	**4**	**0**
44	**23**	**23**	**23**	**0**	**0**
45	**18**	**18**	**18**	**0**	**0**
46	**14**	**0**	**0**	**0**	**14**
47	**15**	**7**	**12**	**5**	**3**
48	**11**	**0**	**5**	**5**	**6**
49	**15**	**0**	**14**	**14**	**1**
5	**25**	**0**	**6**	**6**	**19**

0					
TOTAL	**755**	**557**	**643**	**83**	**115**

Out of the 50 instances of the test set in the current category 27 out of 50 images has an error. Out of 755 words in the test set 83 were misclassified and 115 words were not detected. This shows that there is an 11% error rate for the misclassification of the text and a 15% error rate for not detecting words from the image.

E. Images with colored font and colored background

Picture Number	Number Of words	Number of words correctly detected	Number of words detected	Number Misclass Word
1	**9**	**9**	**9**	**0**
2	**10**	**8**	**10**	**2**
3	**9**	**2**	**6**	**7**
4	**10**	**3**	**6**	**6**
5	**11**	**10**	**11**	**1**
6	**13**	**2**	**14**	**11**
7	**9**	**8**	**9**	**1**
8	**14**	**14**	**14**	**0**
9	**19**	**11**	**19**	**8**
10	**11**	**7**	**10**	**3**
11	**10**	**4**	**9**	**5**
12	**28**	**25**	**28**	**3**
13	**25**	**8**	**25**	**17**
14	**15**	**0**	**4**	**4**
15	**16**	**11**	**18**	**5**
16	**15**	**15**	**15**	**0**
17	**5**	**5**	**5**	**0**
18	**7**	**6**	**7**	**1**
19	**8**	**0**	**0**	**0**
20	**33**	**6**	**34**	**27**
21	**12**	**7**	**12**	**5**
22	**8**	**7**	**8**	**1**
23	**6**	**4**	**7**	**2**
24	**9**	**1**	**6**	**5**
25	**13**	**10**	**13**	**3**
26	**10**	**2**	**10**	**8**
27	**25**	**25**	**25**	**0**

28	**15**	**15**	**15**	**0**
29	**10**	**2**	**2**	**0**
30	**36**	**35**	**36**	**1**
TOTAL	**421**	**262**	**387**	**126**

Out of the 30 instances of the test set in the current category 23 out of 30 images has an error. Out of 421 words in the test set 126 were misclassified and 33 words were not detected. This shows that there is a 30% error rate for the misclassification of the text and a 9% error rate for not detecting words from the image.

F. Images with colored font and white background

Picture Number	Number Of words	Number of words correctly detected	Number of words detected	Numbe Misclass Word
1	**8**	**1**	**8**	**7**
2	**10**	**10**	**10**	**0**
3	**17**	**12**	**17**	**5**
4	**3**	**0**	**0**	**0**
5	**11**	**7**	**11**	**4**
6	**12**	**11**	**12**	**1**
7	**16**	**16**	**16**	**0**
8	**9**	**4**	**9**	**5**
9	**10**	**10**	**10**	**0**
10	**15**	**11**	**15**	**4**
TOTAL	**111**	**82**	**108**	**26**

Out of the 10 instances of the test set in the current category 7 out of 10 images has an error. Out of 111 words in the test set 26 were misclassified and 3 words were not detected. This shows that there is a 23% error rate for the misclassification of the text and a 3% error rate for not detecting words from the image.

G. Images with colored font and black background

ture mber	Number Of words	Number of words correctly	Number of words detected	Number Of Misclassified Words	Numbe words detect

		detected			
1	21	12	16	4	5
2	14	11	14	3	0
3	12	12	12	0	0
4	7	1	7	6	0
5	10	8	10	2	0
6	3	0	0	0	3
7	16	15	16	1	0
8	15	13	15	2	0
9	15	14	15	1	0
10	13	0	0	0	13
	126	86	105	19	21

Out of the 10 instances of the test set in the current category 9 images has an error. Out of 126 words in the test set 19 were misclassified and 21 words were not detected. This shows that there is a 15% error rate for the misclassification of the text and a 17% error rate for not detecting words from the image.

H. Basic words or phrases for Translation

Phrase/Words	Number of Words	Number of words correctly detected	Number of words detected	Number Of Misclassified Words	Number of words not detected
Hello	1	1	1	0	0
I Love You	3	3	3	0	0
Good Night	2	2	2	0	0
Thank You	2	2	2	0	0
I Like You	3	3	3	0	0
Total	11	11	11	0	0

Out of the 5 instances of the test set in the category of black font and white background no images has error. Out of 11 words in the test set 0 were misclassified and 0 words were not detected. This shows that there is a 0% error rate for the misclassification of the text and a 0% error rate for not detecting words from the image.

6. Conclusion and Recommendations

6.1 Conclusion

This research project implemented an image to Sign Language translator with the use of the Tesseract OCR libraries.

Word correction was also implemented through string searching and string manipulation techniques. Sign Language videos were produced using the http://www.signingsavvy.com online resource.

Category	error for misclassification	error for non-detection
Font – Black BG - White	**7%**	**2%**
Font - Black BG - Colored	**12%**	**3%**
Font - White BG - Black	**11%**	**2%**
Font - White BG - Colored	**11%**	**15%**
Font - Colored BG - Colored	**30%**	**9%**
Font - Colored BG - White	**23%**	**3%**

Font - Colored BG - Black	**15%**	**17%**
AVERAGE	**15.6%**	**7.3%**

Evaluation results revealed images that have black font and white background have the least error rate while images that have colored font and colored background have the highest error rate. Images that have one property set as colored has a higher error rate than the black and white. The category with the least non-detection error rate are the black font and white background category and white font and black background category.

The Translation of the image text to sign language has still a considerable percentage of error but with the proper specifications of the font, background and image resolution, the accuracy of the classification can be improved.

The translation for the basic words that are often translated were translated with no error using the category with the least error percentage which is black font with white background.

6.2 Recommendations

The recommended actions for the improvement of the research are to:

- Improve the accuracy of the OCR
- Improve the auto correction methods
- Incorporate a grammar function for the words with the same meaning
- Extend the scope of words
- Incorporate phrase translation

These will significantly increase the accuracy of the translation of the image text. There are still many forms of sign language around the world and each has their own set of rules and properties different from American Sign Language. It will be better if there was a universal sign language so that there will be only one set of rules and properties to follow. With this, it makes it easier to create a dynamic sign language translator for all people.

References:

[1][3] Holley, Rose. 2008. How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs. Retrieved on September 26, 2014. Retrieved from http://eprints.rclis.org/12908/1/ANDP__How_Good_Can_it_Get.pdf

[2] Mitchell, Ross. 2004. How Many People Use ASL? And Other Questions Without Good Answers. Retrieved on September 26, 2014. Retrieved from https://research.gallaudet.edu/Presentations/2004-04-07-1.pdf

[4] Ding, Liya and Martinez, Alex. 2010. Modelling and Recognition of the Linguistic Components in American Sign Language. Retrieved on September

26, 2014. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2757299/

[5]Andrew Kae and Erik G. Learned-Miller. "Learning on the fly: Font free approaches to difficult OCR problems" Proceedings of the International Conference on Document Analysis and Recognition (2009).

[6]Matt Huenerfauth. Improving Spatial Reference in American Sign Language Animation through Data Collection from Native ASL Signers. Universal Access in Human-Computer Interaction. Applications and Services Lecture Notes in Computer Science Volume 5616, 2009, pp 530-539

[7]Louloudis, G., Gatos, B., Pratikakis, I., & Halatsis, C. Text line and word segmentation of handwritten documents. Pattern Recognition,2009, 42(12), 3169-3183.

[8]Singh, R., Yadav, C. S., Verma, P., & Yadav, V. Optical Character Recognition (OCR) for Printed Devnagari Script Using Artificial Neural Network. International Journal of Computer Science & Communication,2010, 1(1), 91-95.

[9]Dreuw, Philippe, Daniel Stein, and Hermann Ney. "Enhancing a sign language translation system with vision-based features." Gesture-Based Human-Computer Interaction and Simulation. Springer Berlin Heidelberg, 2009. 108-113.

[10] Retrieved on September 26, 2014. Retrieved from http://www.cvisiontech.com/reference/ocr/the-most-accurate-ocr-software.html?lang=eng