2006

Formulation of an image processing technique for improving Sign2 performance

Kanthi Sarella

Formulation of an Image Processing Technique for Improving Sign2 Performance

Final Report

Submitted by

Kanthi Sarella

Abstract

This paper states and discusses an image processing technique for improving the performance of Phase I of the Sign2 Project. The Sign2 Project is a complete technological approach to the translation of ASL to digital audio and/or text. Considerable research had been carried out in the Sign2 Project and significant progress had been achieved. Tangible results were derived from the research but there had been unintended inconsistencies in the results where accuracy was concerned. This paper discusses the possible reasons for failure of the Sign2 Project to achieve the intended accuracy and proposes potential methods to ameliorate the approach to achieve better results.

Table of Contents

1. Introduction

The Sign2 Project is an investigation which is focused on a complete technological approach to the translation of American Sign Language (ASL) to digital audio or text. The phase one of this project dealt with recognition of ASL finger-spelling to text using a passive gray-scale image-processing approach and has demonstrated a moderate degree of reliability. The approaches proposed and pursued in this investigation are an extension of the methodology implemented in phase 1 of the Sign2 project.

Two approaches have been formulated to increase the accuracy of the results obtained following the gray scale approach discussed above. The minor challenges faced to accurately recognize letters in the above approach can be attributed to varying factors like different letter formulations speeds, window sizes considered to recognize the existence of a frame, inadequacy of the data available, etc.

The two approaches being proposed so as to obtain better results are the dual-camera approach and the feature extraction method. These methods are being pursued in an endeavor to sharpen the accuracy of the results while recognizing the signed letters.

2. Possible reasons for the procurement of moderate accuracy in the results

There are different theories on why the approach which was pursued initially in the Phase 1 of the Sign2 project failed to yield the intended accuracy in the results. Explained below are the assumed causes and minor challenges faced in Phase 1.

The approach pursued in the Phase 1 involved converting each of the images into a black and white image. While this provided an outline of each of the signs, the approach also obliterated part of the information which would help in recognizing each of the signs. This approach could be more effectively used in the dual camera approach described below. Since the methodology followed in Phase 1 involved a single-camera approach, it could not contribute much information to for the image processing algorithm to work on.

For example, the letters ‘M’ and ‘N’ are indistinguishable when the images of the hand signing for these two letters are converted into logical images.

Fig. 1

Fig. 2

The proposed theory to effectively use the silhouette approach is to use a dual camera approach wherein two sets of data could be gleaned. The dual camera approach is dwelt upon with more detail later on in this report.

Another proposition to improve the accuracy of the results is to take into consideration all the data that the image of a hand sign could offer. This process would include converting the original image into a gray scale image as opposed to a logical image. Features in the image thus processed are retained and the feature-extraction approach is adopted to extract the features in the image which for the input data for the conversion algorithm.

3. Dual Camera Approach

The following section gives a brief description of the dual-camera approach. The setup includes two cameras placed at a fixed distance of ‘D’ from each other. The distance between the subject and the line of which the cameras stand should be a fixed ‘L’. The distances ‘L’ and ‘D’ should be maintained constant. The dual camera approach follows the same methodology as that of the image processing technique illustrated in the previous sections. In the dual-camera approach we employ two cameras which result in

obtaining two set of databases and hence creating twice the amount of data availability. The setup for the dual-camera approach is shown in the Fig 3.

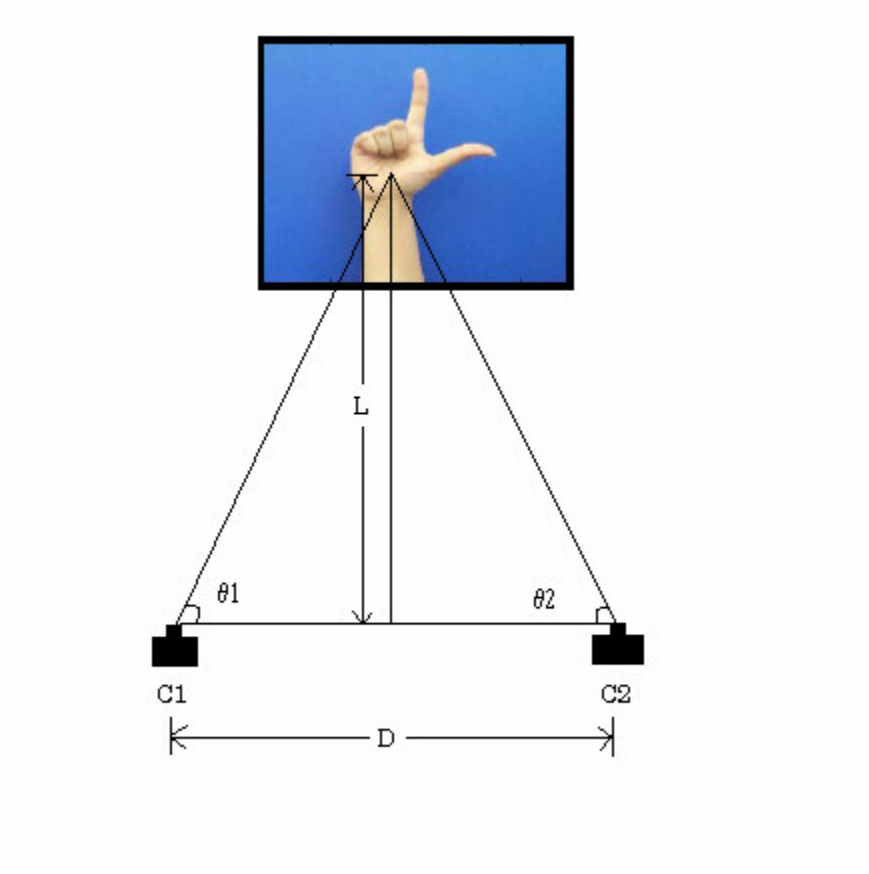


Fig.3: Dual camera approach setup

Shots taken from Camera 1 and Camera 2 are shown in the following figures.

Image taken from Camera C1

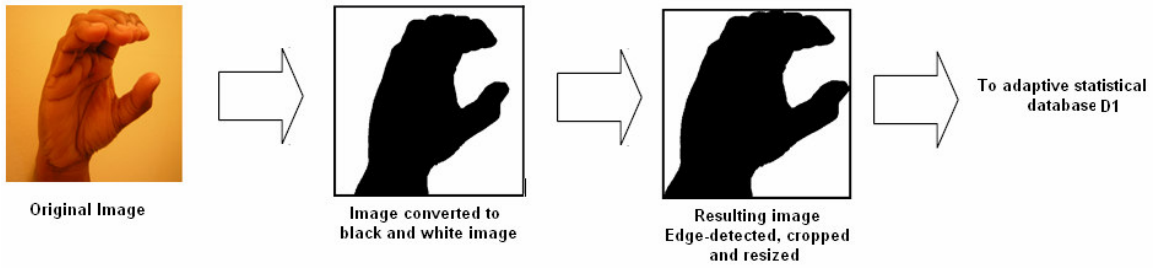
Fig.4

Image taken from Camera C2

Fig.5

A database D1 is formed which consists of the resultant processed images of each of the images shot from Camera 1 (C1) and another database D2 is formed consisting of all the resultant processed images of each of the images shot from Camera 2 (C2). When a fresh image is captured, it is processed and compared to those of the images in D1 and D2. This would result in giving a more exact result. 'A1' corresponds to the signed letter 'A' as shot from the camera C1 and 'A2' correspond to the signed letter 'A' as shot from the camera C2. 9 iterations are considered for each letter. 'A11' would be the first iteration if letter 'A' shot from camera C1. The databases shown below are considered for a single subject. When multiple subjects are considered, the database takes a third dimension with the Z-axis representing the number of subjects.

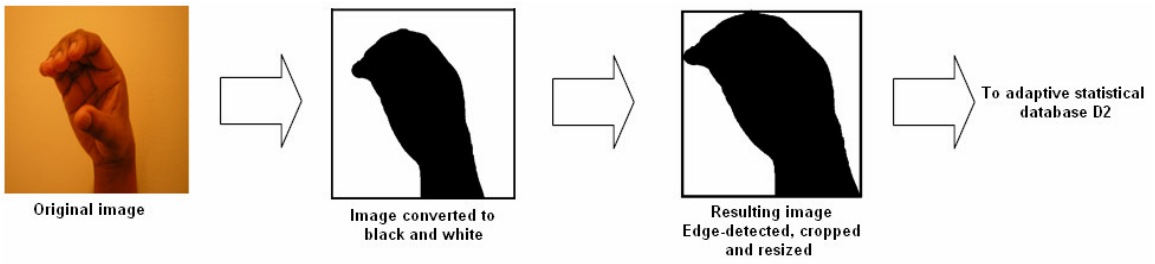
Process:



Database 1

D1	1	2	3	…	…	…	9
A1	A11	A12	A13	…	…	…	A19
B1	B11						
…	…						
…	…						
Z1	Z11						

Fig.6: Image extraction process for images shot from C1 and the corresponding statistical database.



Database 2

D2	1	2	3	…	…	…	9
A2	A21	A22	A23	…	…	…	A29
B2	B21						
…	…						
…	…						
Z2	Z21						

Fig.7: Image extraction process for images shot from C2 and the corresponding statistical database.

The Mean Square Error is calculated for each of the images and stored in the error matrix. The set of images in the database that correspond to a given letter and have the lowest cumulative error reveals the highest probability of the correct letter being returned.

This approach presents a sturdier database which provides more information and hence increases the probability of recognizing a correct letter for a signed hand signal.

4. Feature Extraction Method

The approach that is intended to be pursued is the feature extraction method. In this approach, the images are converted into gray scale images as opposed to the black and white images in the previous methods. The aim of this approach is to retain the features of the image and to work on them as they provide more data than the black and white images. The features that are extracted in this approach are the gray scale intensities of several pixels picked at the intended coordinates. The coordinates are chosen in a way to

cover all the crucial areas of the image where there is the highest likeliness of change in features for each letter.

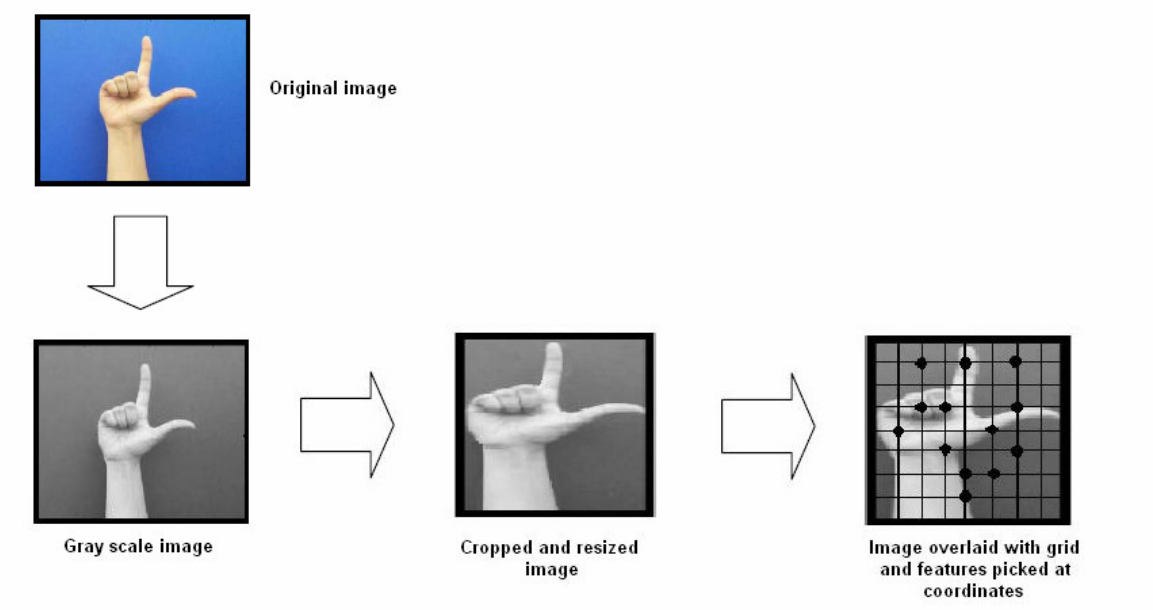


Fig.8: Feature extraction process of an image

The gray scale intensities thus extracted are converted into a statistical representation of the intensities per pixel and fed into a database for later reference.

The algorithm for recognition of newly fed images follows the same algorithm used for the feature extraction of the original images. The gray scale intensity information extracted from the fresh input image is compared to the information in the database to obtain an appropriate recognition of the letter signed. Neural networks are trained to recognize the pattern formed by each of the signed letters. Employment of the neural network helps in normalizing the slight differences obtained while recognizing the letter and smoothens them to the appropriate weighted pattern of each of the letters.

5. Conclusion

The primary focus of this study was to examine image processing as a tool for the conversion of American Sign Language in to digital text. Further this study promises to be used in the real time application to fully recognize American Sign Language. This can be further developed into a system which can be integrated in to the upcoming telecommunication devices with cameras to bridge the communication gap between the hearing and deaf/hard of hearing communities. System can be enhanced in terms of increase in the data processing speed and data storage by using the compression techniques and feature extraction techniques. This analysis presents possible reasons for the inaccuracy of the results of the phase one of the Sign2 process and proposes future enhancements to the system using some of existing and upcoming technologies.

6. References

[1] http://www.deaflibrary.org/asl.html

[2] Kanthi Sarella, Divya Mandloi and Chance M. Glenn. Implementation of Image Processing Approach to Translation of ASL Finger-spelling to Digital Text

[3] Chance M. Glenn, Divya Mandloi, Kanthi Sarella and Muhammed Lonon. An Image Processing Technique for the Translation of ASL Finger-Spelling to Digital Audio or Text.