CHAPTER 1

Project Introduction

Topic: - Text Extraction from Mobile Captured Images
Team: - Deepak Sharma
Objective: - My objective is to develop a software application for text extraction and recognition for mobile captured images. In addition, extracted text will be translated into another language so that a person travelling at foreign places can understand, instantaneously, meaning of various sign board and available text information in his/her surroundings.

CHAPTER 2

Early Relevant Work

Purposed project is not just a single unit it has a number of components --- components has been discussed in the chapter 3, for each components I have found remarkable work from my predecessors. My objective is to Use their work for developing this application and modifying their approaches according to the needs of the application.

Observations:

Before purposing or synthesising any method, making observations is imperative. I have made below two observations and leaned from my predecessor about the nature of the text images.

Cursive and Non-Cursive Languages

Any written piece of information are made of word and components of the word, characters can be connected to each other (Cursive Writing) or they can be independent (Most sign board in English language follow this style). If characters of the word are independent then characters can be recognize much easily as we don't have to separate the characters from word because developing a classifier for characters is feasible. Developing classifier for entire dictionary or in some cases extended dictionary of the language is a near impossible task.

In regard of cursive languages, again any possible algorithm for separating or segmenting associated character from the world, will differ for each language. For example one approach for Persian has been discussed [A], we can't use this approach for Devnagari*[AA]. An alternative would be by developing an algorithm based on template matching approach, -- [B] has discussed such approach for Arabic Text -- though developing application with template matching is easier and faster than approach similar to [A], because a universal algorithm, irrespective of language, can be suggested but in terms of efficiency approach with former category can beat the later approach.

Now for developing a truly global application, application must be relevant to 50% of global population. Hence the target languages from the business point of view should be

:

Language/Script (Writing Nature of sign Boards)	Percentage of Native Speakers in the world population.
Mandarin (Cursive/composite)	14.4
Spanish / English/French/Italian/German (Non Cursive/Cursive)	6.15 + 5.14 + 5.43+.90+1.39
Hindi/Marthi/gujrati/bungali/Punjabi (Cursive)	4.70+1.10+.74+1.44+3.11+1.40

Data Source Wikipedia [X]

For the first version of this application I will be developing prototype for Non cursive English language but these finding will help while giving direction to the project for future application. **If Professor Kinsman thinks/feels the objective of this project need extended then I can push myself for implementing algorithm for Cursive Writing also.** Though my personal observation says Sign Board in various countries of North America, South America, **Europe in parts of Africa and Asia are not written in cursive way of writing for** Spanish, English, French, Italian and German languages**.**

Properties of Character and words in Image: -

Though we have no control on the shape and size of character but still I have amazed myself while perusing [C], [D], [E], [F] [G] and specially [O].

In my observations I have found many sign board does have an outlining on its boarder, so while removing unwanted horizontal and vertical lines, can be eliminated by a simple approach suggested in [C] where they performed morphological opening operation on each connected component to extract possible horizontal/vertical line segments, though they have empirically selected length of structuring element as they have worked on the intuition: aspect ratio of character does not vary much for a piece or writing due to the structure of characters. This information hold true for most of the languages.

[Y], Mathworks, based on their research, has suggested there are several geometric properties that are good for discriminating between text and non-text regions, including: Aspect ratio, Eccentricity, Euler number, Extent and Solidity.

[D] Has extended the idea and purposed below rules for eliminating a connected component from labelled image:

Height > 0.8 × height of the image;
Width > 0.4 × width of the image;
Height < 15 and area < 150;
Height < 0.06 × height of the image;
Height/width > 16;
Width/height > 2;
Area > 0.08 × total area of the image.
Area > 8 × average area;
Area < 0.15 × average area;
Area < max area/20 (noise elimination).
Height > 1.8 × averageheight;
Width > 1.8 × averagewidth;
Any object which is enclosed by another one is eliminated.

[F] Has proclaimed approaches for removing non-text connected components based on the

$$\left(A/(W \times L) \leq \frac{1}{4.5} \text{ or } > .95\right) \quad \text{and}$$
$$\left(\min(W/L, L/W) \leq \frac{1}{5}\right)$$

Above conditions, a non-text element must hold these conditions.

[H] Has suggested based on morphological processing we can differentiate background and foreground keeping the structure and properties of text/characters in centre.

On the other hand [G] and **Professor Kinsman has discussed in the class how using DTC transform** we can synthesis a high pass filter which can remove the constant background and we can have only the text area. **This approach is useful when the text region is far reach from the user and capturing only the sign board is not possible for user**.

Early Approaches and their limitations:

This application is a sequential composition of various small tasks for computer vision and image processing, we can divide this application in two major parts:

1. Text Extraction
2. Character Recognition.

There are a major research has be going on in the field of text extraction and I have found a number of generals on the topic.

While learning the techniques for text extraction from image I have perceived that there are two kind of methodologies have been used by my most of predecessors, Fist the most common for such task is to apply pre-processing followed by text region extraction and then after removing unwanted part from grayscale or binary image, sending the binary image to some OCR system (Example: [D]).

Others (Example: [I] and [G]) which were based of transforming image by wavelet or DCT transform and performing request filtering in order to distinguish the text from unwanted elements of the image. When the text and images (Unwanted) are present in a juxtaposition later approach is more efficient than former.

A ubiquitous issue with algorithms/procedure I have found that no approach work on every possible text image. And most common challenges offered by environmental factors, present in images like lighting conditions, illumination, reflection.

I have found approach [D], [C], [G], [I], [J], [M] and [N] which are relevant to my objective of text extraction. Approach [D] has not discuss the issue of skew and variation in luminosity while defining pre-processing steps. This approach has no solution for situation when image is taken for distance and external background is present in image.

Approach [C] has again not discuss the variation in the luminosity and the suggested structuring element for boarder detection using morphological operation is sensitive to images as it is not robust. Approach [H] is working under assumption: "regions with text in them will have significantly higher values of average edge density, strength and variance of orientations than those of non-text regions and a morphological dilation operator can easily connect these high edge density regions together while leaving those whose position are far away to each other isolated. In their proposed method, they have used a morphological dilation operator with a 7×7 square structuring element to the obtained binary image to get joint areas referred to as text blobs."

In my suggested application a 7x7 square structure will not be able to connect the text characters as sign boards generally consist of large size character and increasing the size of structuring element would result into connecting two characters.
Similar issue with. Approach [N] has suggested a fixed size morphological structuring element has achieving result from this approach is not possible.
Though I have found approach [I] convincing but implanting K mean clustering algorithm is not possible for mobile application keeping in mind the computation power of Morden mobile devices. I have to further investigate on this issue. Again [I] is based on wavelet transform approach which is not required for the objective of this application as we have discussed previously.
Results gain from [M] is not impressive as [M] can use approach [I] for improving their results.

CHAPTER 3

Purposed Method

Keeping in mind the task of text extraction and our discussed in previous chapters. I have decided to use two subroutines for the task, User will provide the information about the image, and if the image has been taken from the distance then extra pre-processing for region extraction will be performed as suggested in [G]. Else this step will be escaped because the background of images in sign boards are generally content and image will only contain region of interest, hence transforming the image using DTC or wavelet transform followed by constructing high pass filter will be futile.
As a base I will be following [D] but based on our previous discussion I will implement few steps with efficient alternatives and also add new steps like Skew correction and Quantization.

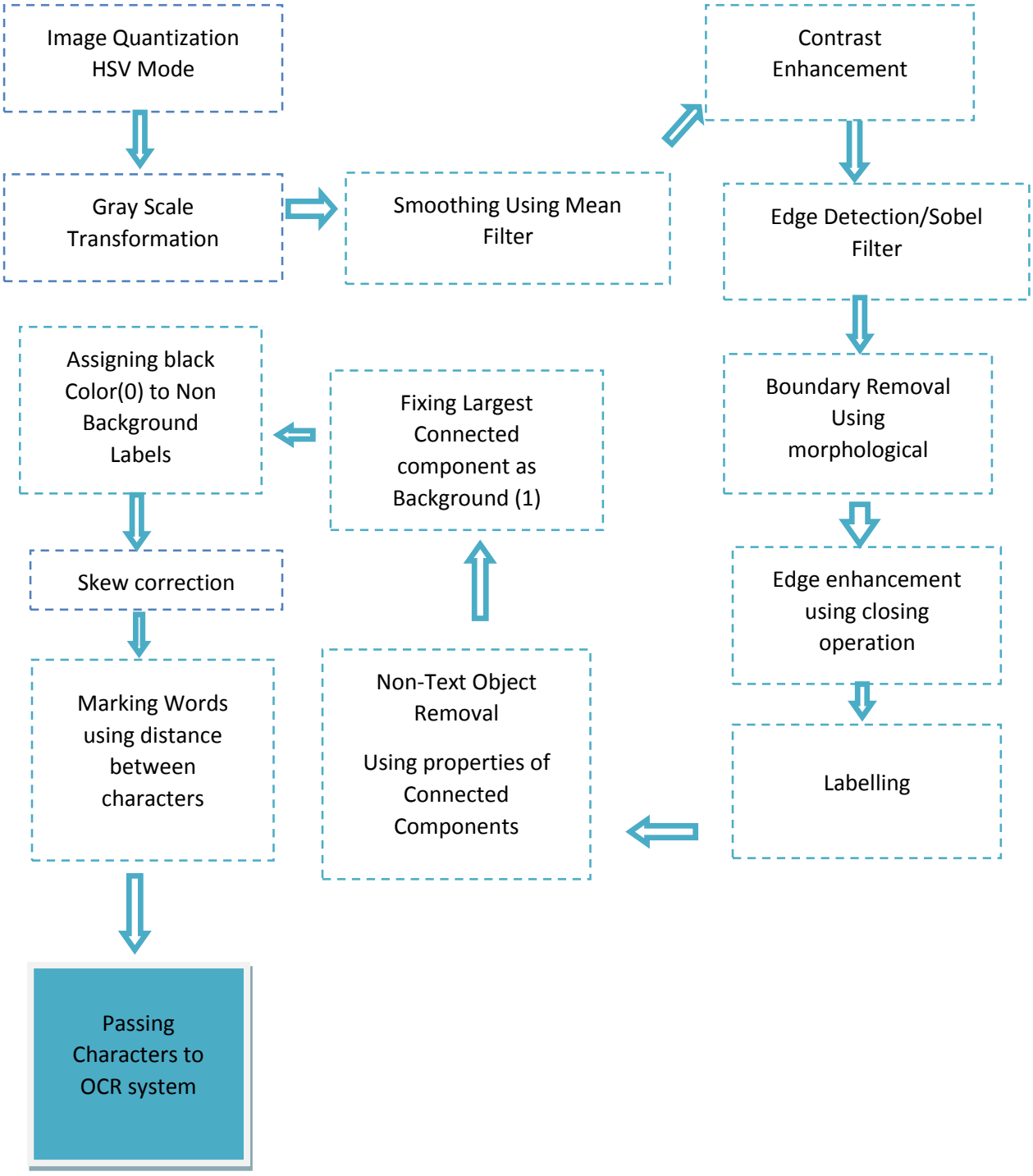
High Level Design: -

Pre-Processing : In this part of the procesedure task like smoothing , contrast Inhacemnt and qunitzation will be performed. Objective of this step is create ideal conditions for edge detection.

Region Sementation: In this part of the using sobel edge detector, edges will be detected and based on our discussion in previous chapter, using [c] we will remove unwanted edges (Edges pertaining to boudry of the image). Then using morphological operations like closing and skeletanization all the edges will be enhaced. Objective of this step is creating ideal conditions for labeling.

Post Processing: Perfroming labeling and removing non-text element. Objective of this step is creating binary image which has a clean background and Image.

Low Level Design: -



Details of the low level Design.

1. Image Quantization HSV Mode: - As per our observations in Assignment 5, by converting image in HSV model and performing quantization will element the effect of shadows and reflection, as using HSV model we can separate the intensity from image. So result for image taken during morning and afternoon of the same sign board will be same.
2. Gray Scale Transformation:- Image will be converted to gray scale.
3. Smoothing Using Mean Filter: - smoothing using mean filter before edge detection will give better results than Gaussian as Gaussian will blur the image.
4. Contrast Enhancement: - As per our finding during previous Homework's, for the time being I will go with adaptive histogram technique for this step as it has given us best results.
5. Edge Detection/Sobel Filter: - Edge detection is a crucial step in this process, as we are targeting to completely closed bodies of the words. As for sobel edge detection, unlike canny, we don't have to set parameters, so I will use Sobel Edge Detection.
6. Boundary Removal Using morphological: - As per our previous discussion, Using technique described in [C] I will remove tilted long lines along with long horizontal and vertical lines (non-character elements). Using adequate structuring element this step will be performed.
7. Edge enhancement using closing operation: - before libelling process, I will performing closing operation as it is important to have all edged bounded so that characters can be assigned different labelling numbers
8. Labelling: - As Discussed in [F], I will use a fast forward-backward propagation algorithm, as it is faster than traditional two scan algorithm for labelling. By labelling we will give distinct number to the all pixel of a particular character.
9. Non-Text Object Removal using properties of Connected Components: - As per hour discussion in previous chapter and mentioned conditions we will remove the non-text characters, based on the shapes. Geometric constrains discussed into [O] will be used to differentiate between text and non-text elements.
10. Fix the Largest Connected component as Background (1)
11. Assigning black Colour (0) to Non Background Labels
12. Skew correction: - as discussed in [K] we can measure overall skew using information of all the connected component, as it says "For obtaining the orientation, ellipse is fitted over the connected component regions. The angle of major axis of the ellipse is measured with respect to the horizontal axis and this angle is considered as the orientation angle, after calculation of the orientation angles of all the regions, average orientation of these regions is taken as the estimated skew angle of the text image.
13. Marking Words using distance between characters
14. A simple OCR can be created using templates, each extracted character, after normalization and skeletanization can be correlated with the skeletonized templates. The templates dataset will consist of all famous founts. From Matlab help I lean using Corr2, correlation coefficient can be computed, Here A and B are MxN matries, so A can be skeletonized templet matrix and B can be extracted character after

skeletonized and normalized to the size B.

$$r = \frac{\sum_m \sum_n (A_{mn} - \overline{A})(B_{mn} - \overline{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \overline{A})^2\right)\left(\sum_m \sum_n (B_{mn} - \overline{B})^2\right)}}$$

As correlation is the measurement of likelihood of two vectors hence the highest the templet for the highest correlation will be messaged that template will be assigned for the character.

Algorithm for labelling suggested in [F]

```
Definitions:   A: i/p NxM binary image ,    B: o/p NxM labeled image
Algorithm:
      begin
            % Assign temporary labels to nonzero pixels of the image.
            for i=1 to N, j=1 to M, do
                if A(i,j) =1, then sequentially assign unique integer number to B(i,j).
            end do
            set max_temp_label= the maximum assigned integer label.

            % Propagate maximum label per connected component.
            do while at least one pixel value of B(i,j) changes
                % back propagation row-wise
                for i=N to 2, j=1 to M, do
                    if both B(i,j) and B(i-1,j) > 0, then
                      set both B(i,j) and B(i-1,j) to maximum(B(i,j),B(i-1,j))
                end do
                % back propagation column-wise
                for i=1 to N, j=M to 2, do
                    if both B(i,j) and B(i,j-1) > 0, then
                      set both B(i,j) and B(i,j-1) to maximum(B(i,j),B(i,j-1)
                end do
                % forward propagation row-wise
                for i=1 to N-1, j=1 to M, do
                    if both B(i,j) and B(i+1,j) > 0, then
                      set both B(i,j) and B(i+1,j) to maximum(B(i,j),B(i+1,j)
                end do
                % forward propagation column-wise
                for i=1 to N, j=1 to M-1, do
                    if both B(i,j) and B(i,j+1) > 0, then
                      set both B(i,j) and B(i,j+1) to maximum(B(i,j),B(i,j+1))
                end do
            end while

            % Sequential relabeling
            for k=1 to max_temp_label, do
                initialize array_label(k)=0
            end do
            max_label=0
            for i=1 to N, j=1 to M, do
                if A(i,j) =1, then
                  if array_label(B(i,j)) = 0, then
                    increment max-label
                    array_label(B(i,j))=max_label
                    B(i,j)=max_label
                  else
                    B(i,j)=array_label(B(i,j))
            end do
```

Conclusion: - labelling and edge detection are the two critical operations in the suggested approach, the failure of systems depends on the failure of these two operations. Quantization step should not divide the background into two different parts.

Ethical considerations: - After musing upon all the use of the possible uses of application, I have arrived to the conclusion as of now, there is no ethical consideration associated with the approach.

Testing and Data

Testing of Region Segmentation step will be performed on the below database:
http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/

Is there a business case for this work: - This entire project I developing as a business product.

References:

[A]Persian Cursive Script Recognition Mahmoud Reza Hashemi, Omid Fatemi, and Reza Safavi* Department of Electrical Engineering, University of Ottawa errail:hashemi@marge.genie.uottawa.ca; * Iran Communication Industries Inc.
[B]Automatic Processing of Arabic Text Ziad Osman, Electrical Engineering, Beirut Arab University, Beirut, Lebanon E-mail: zosman@bau.edu.lb Lama Hamandi, Electrical and Computer Engineering, American University of Beirut, Lebanon E-mail: lh13@aub.edu.lb Rached Zantout, College of Computer and Info Sciences, Prince Sultan University, Riyadh, KSA E-mail: rzantout@psu.edu.sa Fadi N. Sibai, Computer Systems Design, College of IT, UAE University, Al Ain, UAE E-mail: fadi.sibai@uaeu.ac.ae
[C] Devanagari and Bangla Text Extraction from Natural Scene Images U. Bhattacharya, S. K. Parui and S. Mondal Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata – 108, India {ujjwal, swapan, srikanta_t}@isical.ac.in
[D] Text Extraction from Images Captured via Mobile and Digital Devices Jian Yuan, Yi Zhang, Kok Kiong Tan, Tong Heng Lee Department of Electrical and Computer Engineering National University of Singapore. Email: {g0801794, zhangyi, kktan, eleleeth}@nus.edu.sg
[E] Morphological Text Extraction from Images Yassin M. Y. Hasan and Lina J. Karam
[F] R. Kasturi and M. M. Trivedi, Images Analysis Applications. New
York: Marcel Dekker, 1990.
[G] Text Region Extraction from Low Resolution Natural Scene Images using Texture Features S. A. Angadi Department of Computer Science & Engineering Basaveshwar Engineering College Bagalkot, Karnataka, India M. M. Kodabagi Department of Computer Science & Engineering Basaveshwar Engineering College Bagalkot, Karnataka, India
[H]MULTISCALE EDGE-BASED TEXT EXTRACTION FROM COMPLEX IMAGES
Xiaoqing Liu and Jagath Samarabandu the University of Western Ontario
Department of Electrical & Computer Engineering London, Ontario, N6A 5B9, Canada
xliu65@uwo.ca, jagath@uwo.ca
[I]TEXT EXTRACTION ALGORITHM UNDER BACKGROUND IMAGE USING WAVELET TRANSFORMS XIAO-WEI ZHANG, XIONG-BO ZHENG, ZHI-JUAN WENG School of Science, Harbin Engineering University, Harbin, 150001, china E-MAIL: zhangxiaowei@hrbeu.edu.cn
[J] Text Detection and Recognition in RealWorld Images. Raid Saabni Moti Zwilling

Faculty of Engineering, Tel-Aviv University. Triangle R&D Center, Kafr Qarea, Israelsaabni@cs.bgu.ac.ilSchool of Business Administration,Academic Center of Law Business,Ramat Gan, Israel
[K] **Skew Detection in Binary Image Documents Based on Image Dilation and Region labeling Approach**B. V. Dhandra[1], V. S. Malemath[1]*, Mallikarjun H. [1], Ravindra Hegadi[1]*[1]Post-Graduate Department of Studies & Research in Computer Science,*
Gulbarga University, Gulbarga-585106 Karnataka, INDIA
[L] Connected Component Based Approach for Text Extraction from Color Image Kamrul Hasan Talukder Computer Science & Engineering Discipline Khulna University, Khulna Bangladesh. k.h.t@alumni.nus.edu.sg Tania Mallick Computer Science & Engineering Discipline Khulna University,
[M] Extraction of Texts from a Scene Image using Morphology Based Approach Mohammad ShorifUddin, Madeena Sultana, Tanzila Rahman, and Umme Sayma Busra Dept. of Computer Science and Engineering Jahangirnagar University Savar, Dhaka, Bangladesh
[N] Text Detection and Removal from Image using Inpainting with Smoothing Priyanka Deelip Wagh Dept. of Computer Engineering SES's R. C. Patel Institute of Technology Shirpur (MH), India. D. R. Patil Dept. of Computer Engineering SES's R. C. Patel Institute of Technology Shirpur(MH).
[O] Scene Text Detection via Stroke Width Yao Li and Huchuan Lu School of Information and Communication Engineering, Dalian University of Technology, China lhchuan@dlut.edu.cn
[X] https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers
[AA] https://en.wikipedia.org/wiki/Devanagari
[Y] http://www.mathworks.com/help/vision/examples/automatically-detect-and-recognize-text-in-natural-images.html?refresh=true