

Image Data Extraction  
Design Report  
28th March 2017  
Version 1.0

## 1 Summary

The product aims to solve the problem of reading text from the images. As we all are aware of the fact that in today's scenario how important data is and searching of relevant information is. Therefore, we are presenting a software which converts the text in the image a into text document and then operations can be performed on that text.

## 2 Algorithmic Design

In the following section we had described the algorithmic design of our product.

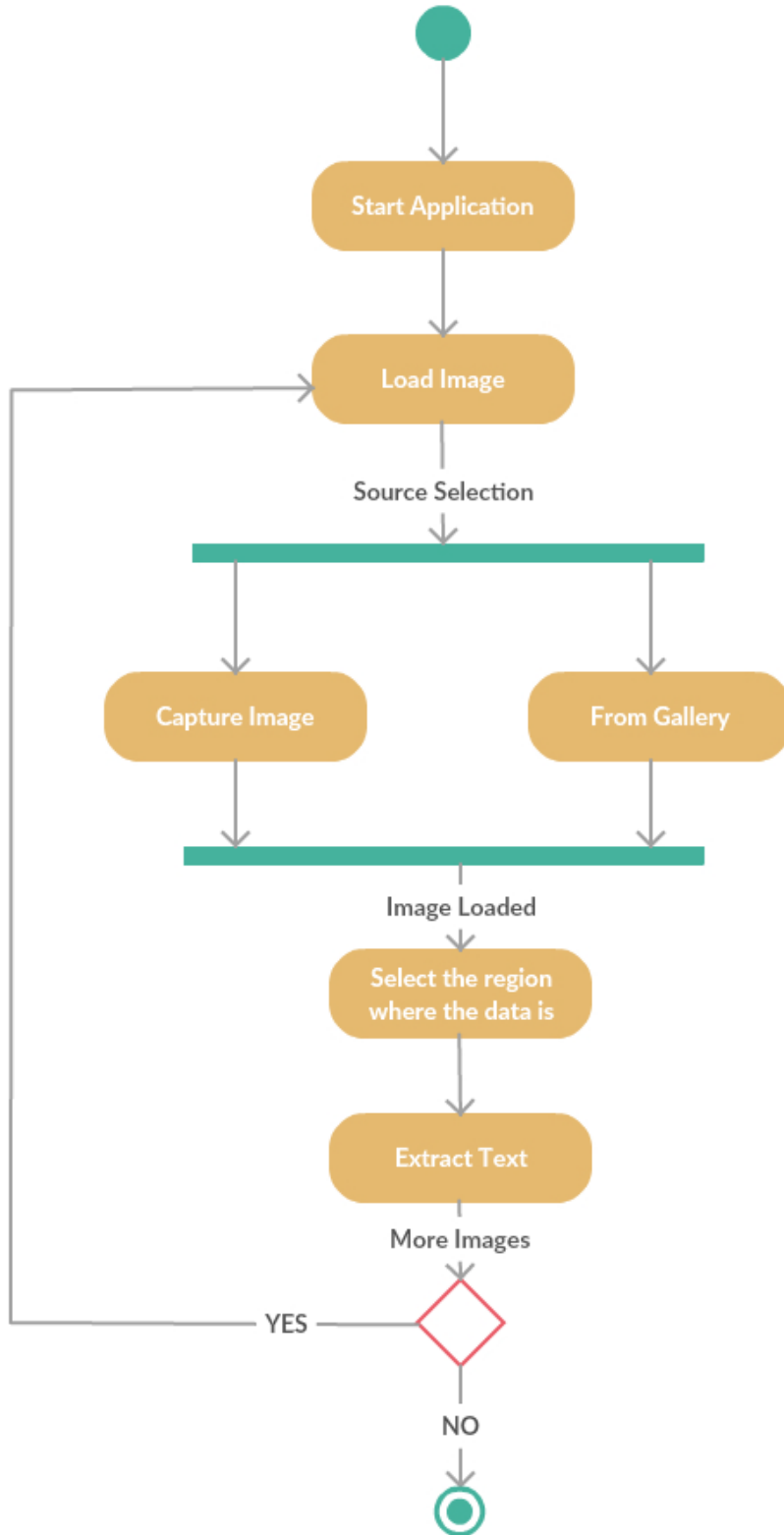
### 2.1 DATA FLOW DIAGRAM

In this section we had shown how the image(data) is processed in our application. The input image goes through several processes and finally results in a text document containing the extracted text from the image.



---

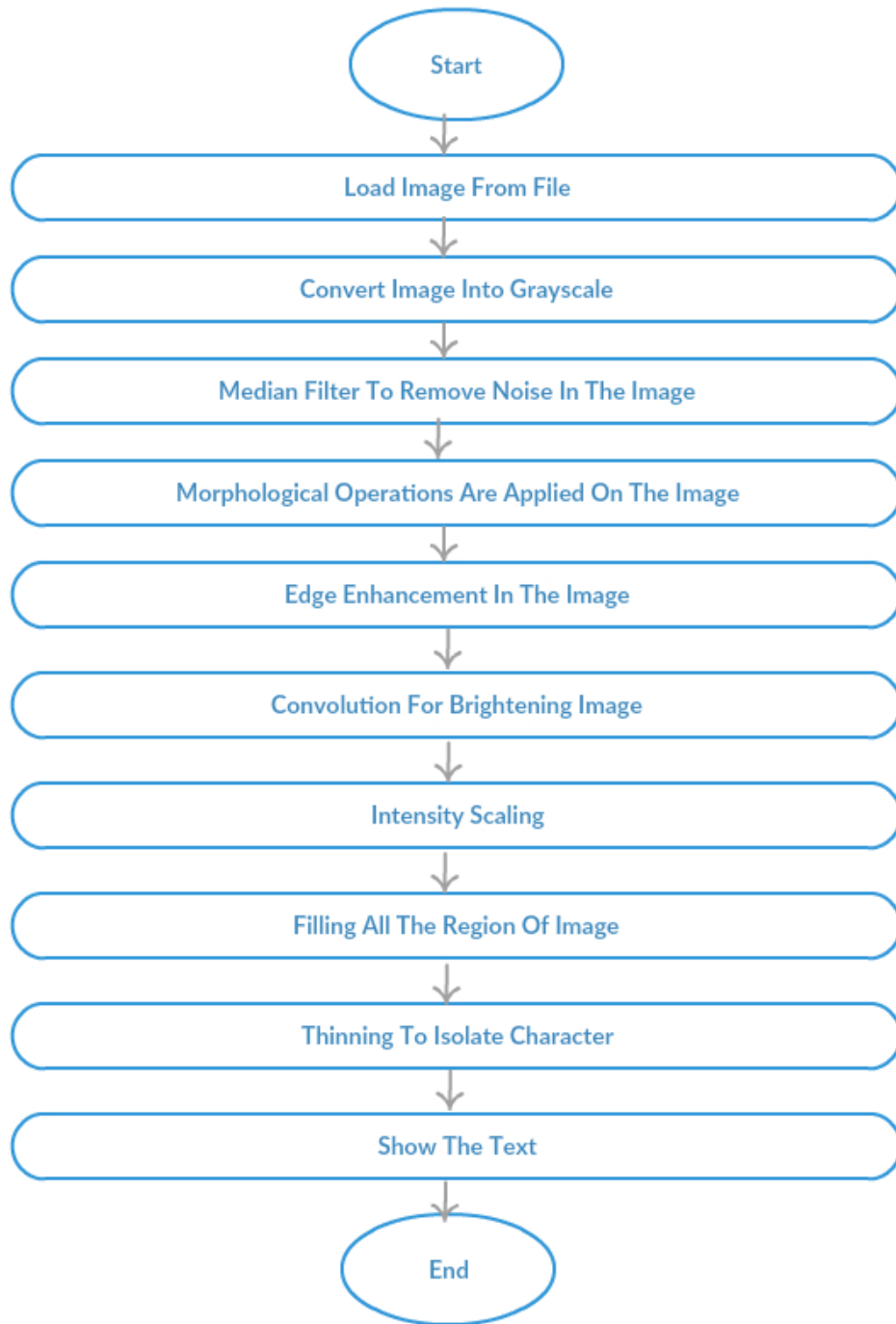
## 2.2 ACTIVITY DIAGRAM



---

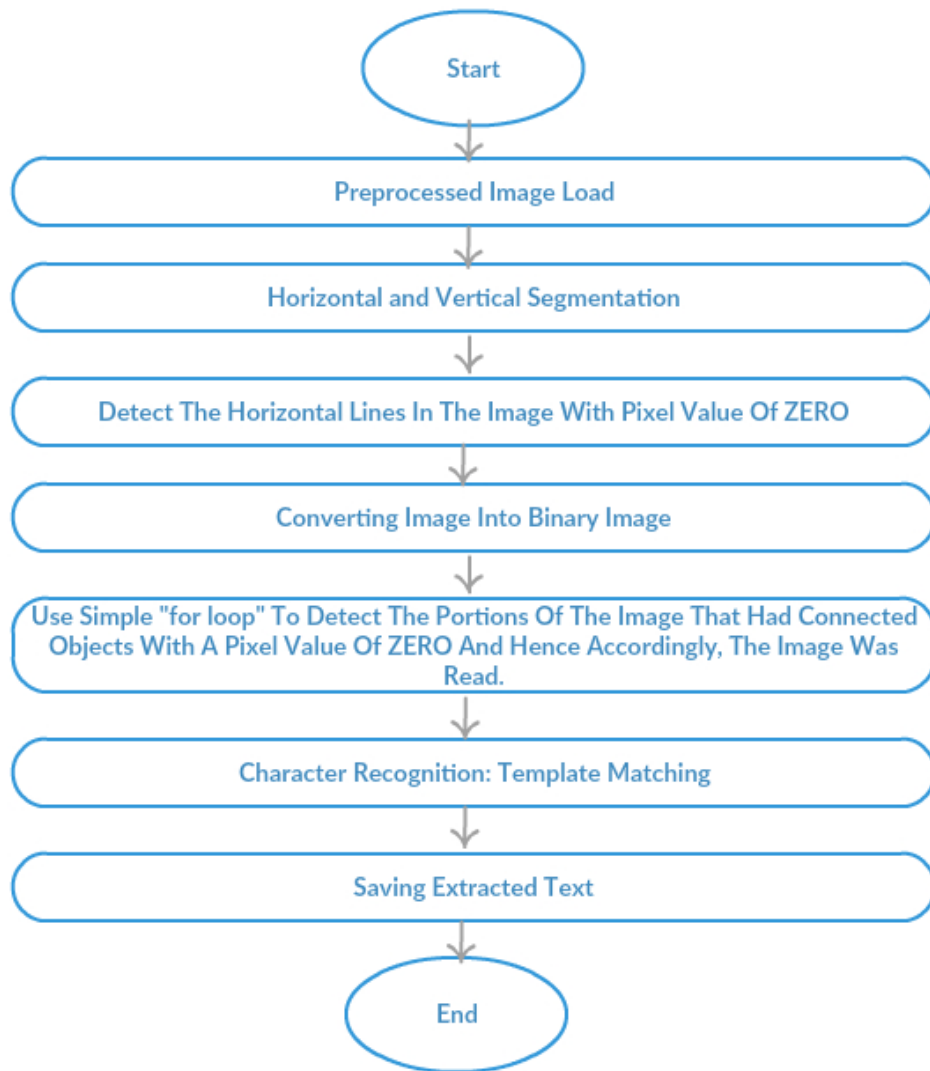
### 2.3 Algorithm

Preprocessing of an image. This is done for the better efficiency of the algorithm. In preprocessing the end data is the image in which the text region is known.



---

After preprocessing, from the known text region we isolate characters and recognise them using Template Matching. Then the recognised text is stored.

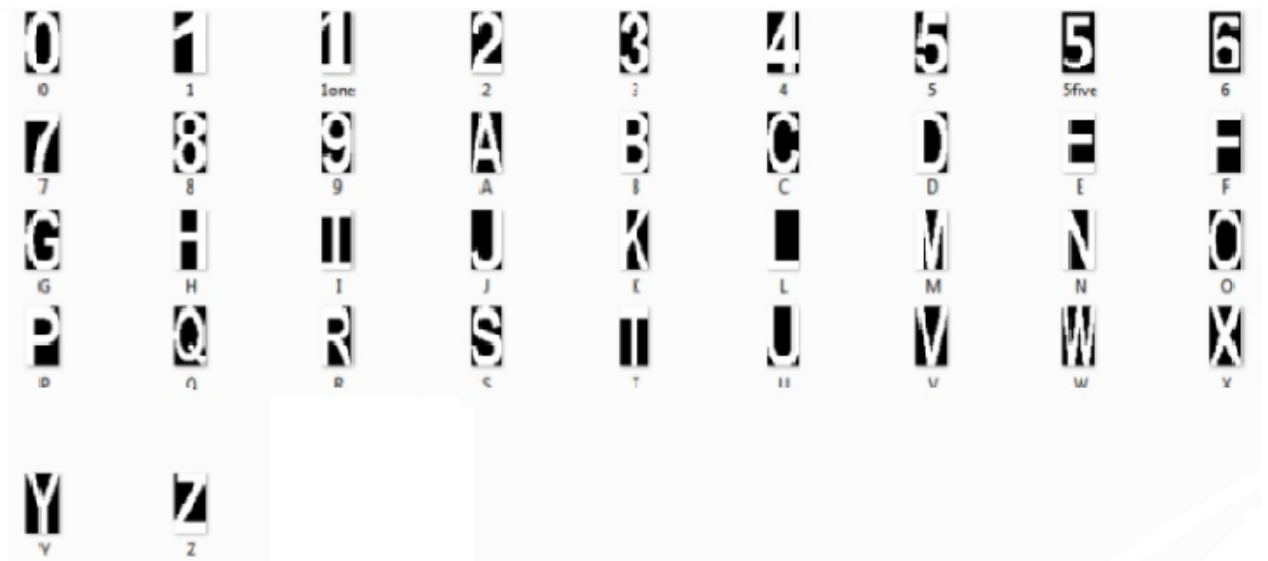


---

Following are the example of templates used for recognising the text from the image.

# TEMPLATE MATCHING

The used templates are given in the figure below:



## 3 Frequently Asked Questions

### 3.1 What is the function of your application?

Our application extracts the text from an image and stores it in a text document.

### 3.2 What is the benefit of Image Data Extraction?

Image Data Extraction is the best application that provides instantaneous data from images.

---

### 3.3 How can we load images in the application?

Images can be loaded in two ways:

1. load image from gallery or/and,
2. capture image

### 3.4 Does your application extracts the hand written text?

No, our application doesn't extract the hand written text.

### 3.5 Does image should only contain text or it can contain other objects?

No, it can contain other non-text objects in the image also.

## 4 Comments

For further information, suggestion or complaint contact:

Project Code: <https://github.com/kanishk1225/Image-Data-Extraction>

Github profile: <https://github.com/kanishk1225>

Mail ID: [kanishk125.ucs2015@iitr.ac.in](mailto:kanishk125.ucs2015@iitr.ac.in)

Gitter Chat Group: <https://gitter.im/Image-Data-Extraction/Lobby#>

## 5 Team Members

1. Kanishk Goyal (15114035)
2. Siraz Shaikh (15114065)
3. Sunil Yadav (14114064)
4. Ujjawal (15114074)