

Digitising of Bills

Team Bulldozers

Manan Bhandari (201431166)

Toshad Salwekar (201401167)

Shivang Nagaria (201401146)

Introduction

- Converting physical information into digital data has always been a challenge to technology.
- Keeping physical information organized is difficult. Digital data can be managed very easily.
- This approach extracts useful information from images of bill and stores in digital form.

Approach

Pre-processing

- Image Alignment
- Cropping out bill image and removing other objects
- Image restoration

Processing

- MSER and OCR method
- Different approach for handwritten and printed bills

Post-processing

- Extracting the information and saving it in a file

Pre-processing

Obtaining the bill from the image

- Cropping out irrelevant objects
- Image Alignment

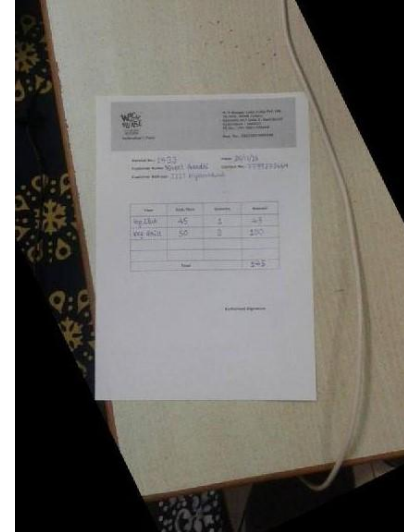
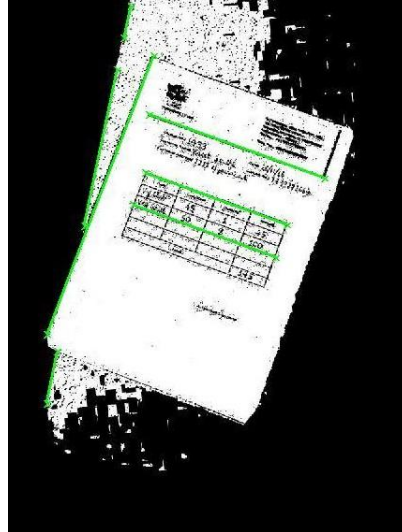
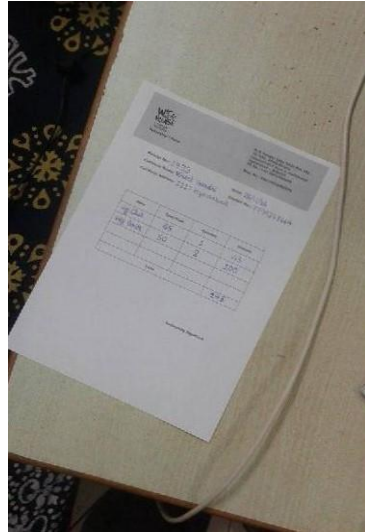
Cropping out irrelevant objects

The receipt image could contain other unwanted objects too. Here, we make use of the fact that the receipt has a white background and that it is written on a rectangular sheet and located near the centre of the image.



Image Alignment

As our solution needs image of a bill, we need to first align the bill accordingly so that text detection can be improved. We've used Hough Transform for this purpose.



- MSER and OCR methods
- Different methods for handwritten and printed bills

Processing

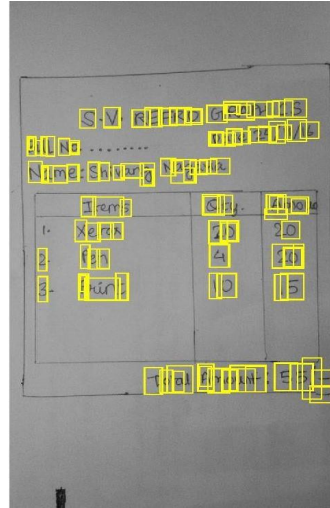
Text Extraction

MSER and OCR Methods

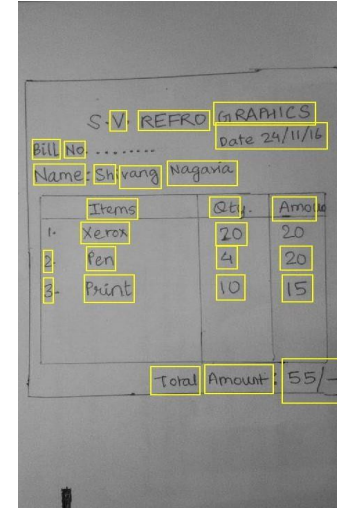
Recognizing text in images is useful in many computer vision applications such as image search and robot navigation. For detecting the letters and hence the words written on the bill, be it printed or handwritten, we have used MSER & OCR method.



Recognising each character



Making Bounding Boxes



Merging them to get each word

Different methods for handwritten and printed bills

Generally all the information like product name, quantity, amount etc are written in 2 ways on the bill : either in tabulated form or normal form with some space between product and their cost.



Receipt No.: 1433 Date: 17/9/2016
Customer Name: Dhriti Contact No.: 9581248002
Customer Address: IIIT Hyderabad

Item	Cost/Item	Quantity	Amount
Sandwich Pack	30	300	9000
Total			9000

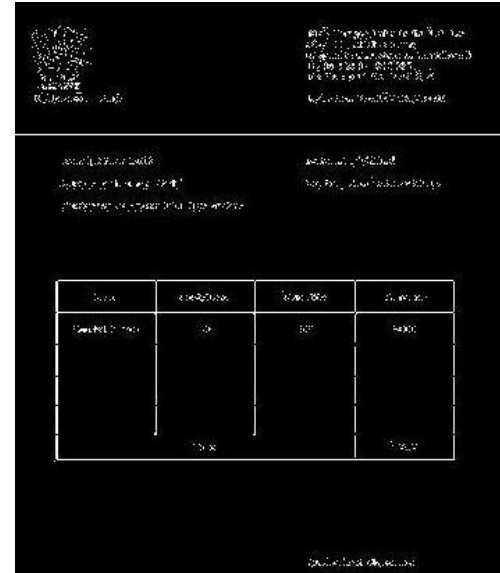
Authorized Signature



Receipt No.: 1433 Date: 17/9/2016
Customer Name: Dhriti Contact No.: 9581248002
Customer Address: IIIT Hyderabad

Item	Cost/Item	Quantity	Amount
Sandwich Pack	30	300	9000
Total			9000

Authorized Signature



Post-processing

Finally saving data

As the information is obtained when the code is running, it is simultaneously being saved in a text file. By this, we get the basic as well as the important information.

Example 1

Receipt No.: 1433 Date: 17/9/2016
Customer Name: Dhriti Contact No.: 9581248002
Customer Address: IIIT Hyderabad

Item	Cost/Item	Quantity	Amount
Sandwich Pack	30	300	9000
Total			9000



Receipt No.: 1433 Date: 17/9/2016
Customer Name: Dhriti Contact No.: 9581248002
Customer Address: IIIT Hyderabad

Item	Cost/Item	Quantity	Amount
Sandwich Pack	30	300	9000
Total			9000



Receipt No.: 1433 Date: 17/9/2016
Customer Name: Dhriti Contact No.: 9581248002
Customer Address: IIIT Hyderabad

Item	Cost/Item	Quantity	Amount
Sandwich Pack	30	300	9000
Total			9000



Receipt No.: 1433 Date: 17/9/2016
Customer Name: Dhriti Contact No.: 9581248002
Customer Address: IIIT Hyderabad

Item	Cost/Item	Quantity	Amount
Sandwich Pack	30	300	9000
Total			9000

Example 1

Receipt No.: 1433 Date: 17/9/2016
Customer Name: Dhriti Contact No.: 9581248002
Customer Address: IIIT Hyderabad

Item	Cost/Item	Quantity	Amount
Sandwich Pack	30	300	9000
Total			9000

Detected Text

Receipt No.: 1433 Date: 17/9/2016
Customer Name: Dhriti Contact No.: 9581248002
Customer Address: IIIT Hyderabad

Item	Cost/Item	Quantity	Amount
Sandwich Pack	30	300	9000
Total			9000

Result - Notepad

File Edit Format View Help

Receipt No. : 1433 Date : 17/9/2016 Name : Dhriti Sandwich Pack 300 9000

Example 2

Rustom's
-- Prelim Receipt --
THIS IS NOT A FINAL BILL

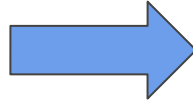
R02

Rustoms1
Prem
10-05-2015 02:35 pm

RUSTOMS1
Patron Count :1

	Rs.
Guest 1	
1x BHEEDA PAR EEDU	200.00
1x BOTTLED WATER	50.00
1x KHEEMA PATTICE	295.00
1x KOLMI NO PATIO	395.00
1x LAGAN NU CUSTARD	225.00
1x MALAI NA PARATHA	80.00
1x PAATRA NI MACCHI	295.00
1x RUSTOM'S GINGER FIZZ	120.00
Sub Total	1,660.00
Service Charge	166.00
Service Tax 4.944%	90.14
VAT AB 20%	26.40
Vat 12.5%	211.75
Total	Rs.2,154.29

-- Thank You For Your Order --
facebook.com/rustombsbhoru



Result1 - Notepad

File Edit Format View Help

Rs.

Guest 1 :

ix BHEEDA PAR EEDU 200 00
ix BUTTLED NATE - 50.00 ,
ix KHEEMA PATTI 295.00
ix KULMI NO PAT 395.00
ix LAGAN NU BUS ARD 225.00
ix MALAI NA PARATHA 80.00
ix PAATRA NI MACCHI 295.00 |
ix RUSTOM'S GINGER FIZZ 120.00

Further Improvements

Number of true principle lines detected can be increased if we also consider their length.

Template matching for each character and number, and then use bounding boxes to club the letters to form the words.

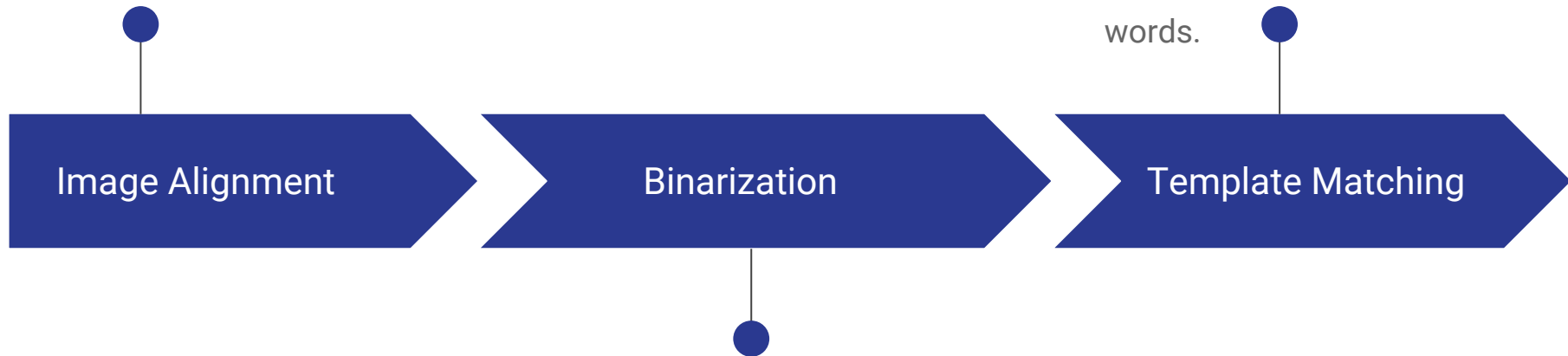


Image Alignment

Binarization

Template Matching

Adaptive thresholding can be used in place of local thresholding for better results.



Thank You