**Team 7**

# Restaurant Menu Expert

## Mentor TA

Adithya Arun

## Team Members


| | |
|---|---|
| Tirth Pandit | ( 2019201017 - Mtech CSE ) |
| Jay Krishna | ( 2019201019 - Mtech CSE ) |
| Smit Khanpara | ( 2019201021 - Mtech CSE ) |
| Pratik Tiwari | ( 2019201023 - Mtech CSE ) |
| Dharmesh Gusai | ( 2019201025 - Mtech CSE ) |


**GitHub Repo Link :** https://github.com/Digital-Image-Processing-IIITH/project-team-7.git

# Introduction

- Its a System that translates the text in English restaurant menus into pictures of the dishes.
- Takes photo of a menu that users take with a camera as an input
- It then processes it to retrieve the corresponding pictures of the dishes from database
- Finally displays the pictures of Dishes with its names to users

## Application

Helps Non-English Speakers To get visualization of dishes in the menu

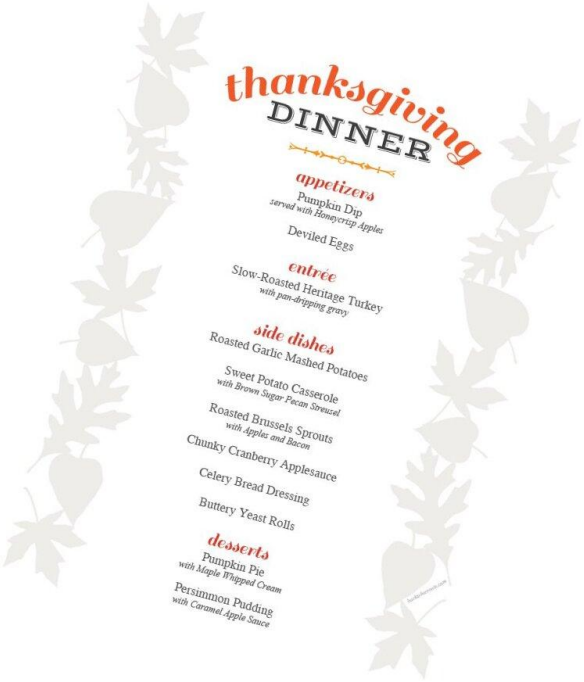
# Features Of the Pipeline

- Pipeline Corrects input image rotation by performing featureless rotation
  Which reduces search space and noise by identifying the ROI.

- Pipeline does post-processing with edit-distance error
  Improve accuracy from the Tesseract OCR engine.

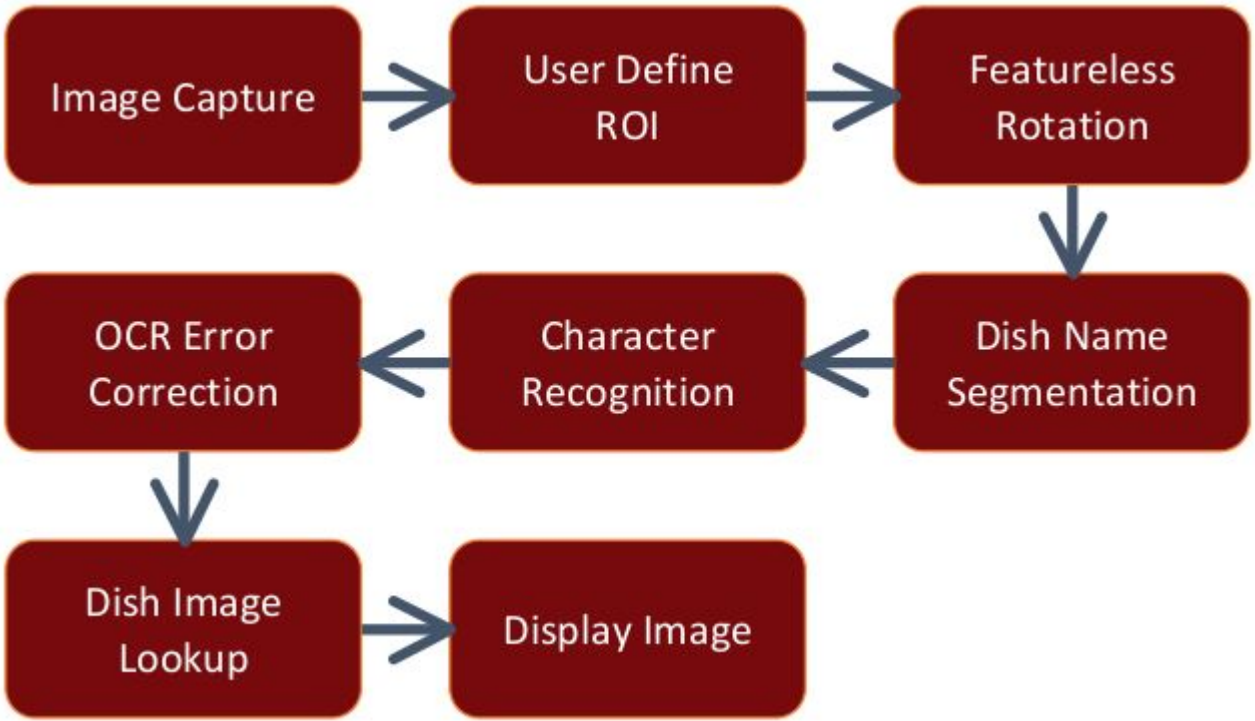
- System can attain 50% higher in accuracy rate than the plain OCR recognition without our pipeline.

# Example



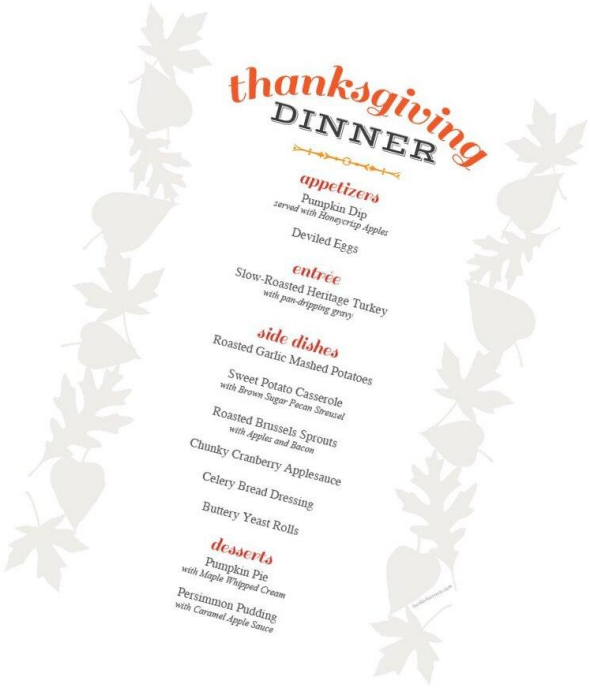
Source : https://web.stanford.edu/class/ee368/Project_Autumn_1516/Reports/Wang_Chen_Lang.pdf

# Proposed Pipeline




Source : https://web.stanford.edu/class/ee368/Project_Autumn_1516/Reports/Wang_Chen_Lang.pdf

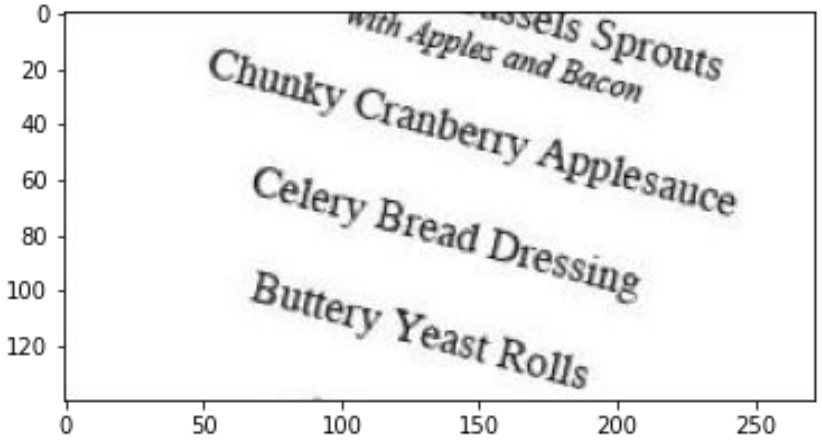
# Original Image




Rotation Angle : 15

# Image Capture

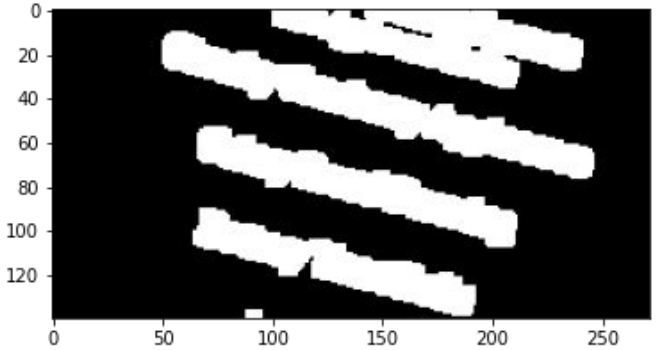
- User takes a photo of the menu with a camera.
- We provided an interface to Select ROI from captured image
- We assume user takes photo such that other variables that may result from the environment such as motion blur and out out-of-focus image don't effects the input image.




**ROI Selection**

# Otsu Thresholding & Dilation

**OTSU Thresholding**



**Disk Dilation**

**Structural Element**
Disk with radius 5

**Creates Connected Components**

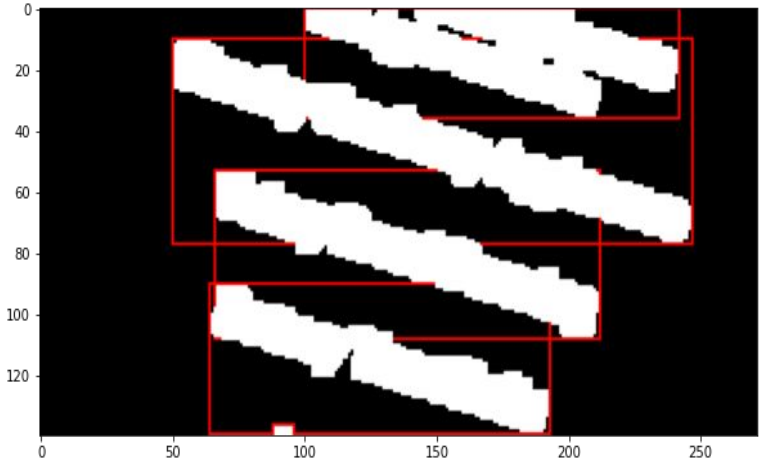
# FeatureLess Rotation

1. It automatically finds out the rotation angle based on the inherent structure of menu where most texts are horizontally aligned

2. First global otsu thresholding and then dilation which creates connected components and then we can create bounding boxes around them

3. With these bounding boxes of the connected components, we calculate the mean aspect ratios between the width and the height of the bounding boxes at every 10° from -90° to 90° with subsequent refinement of every 1° to pinpoint the angle of rotation.

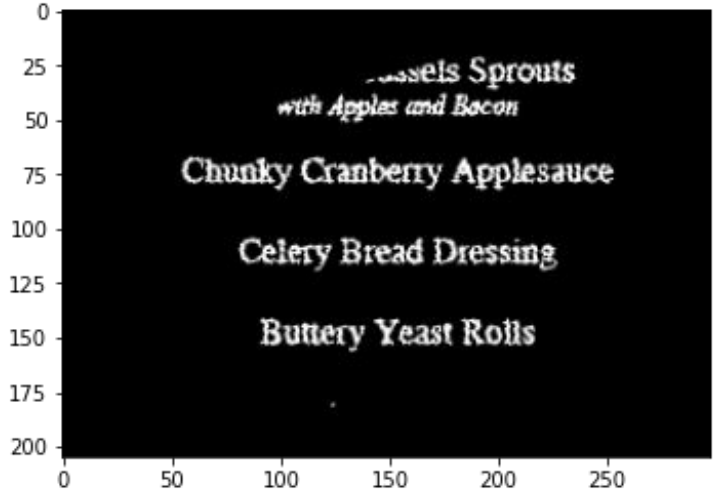
4. From these set of rotations, we automatically find the largest mean aspect ratios, as that suggests that most bounding boxes are horizontal at that particular angle



**Bounding Boxes**

**Rotation Angle**

**15**



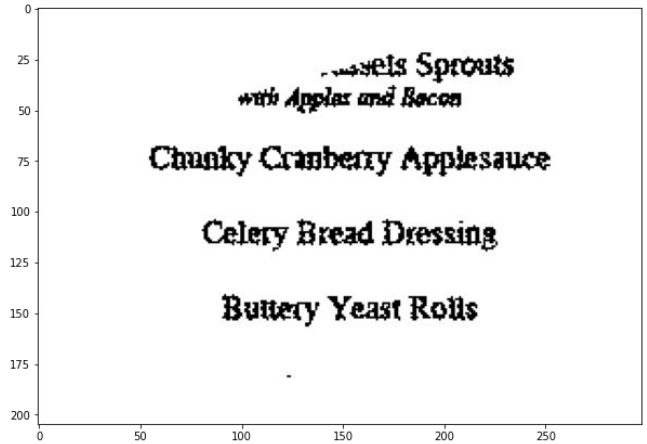

**Featureless Rotation**

Rotating Original Image with angle that resulted from Featureless Rotation Technique

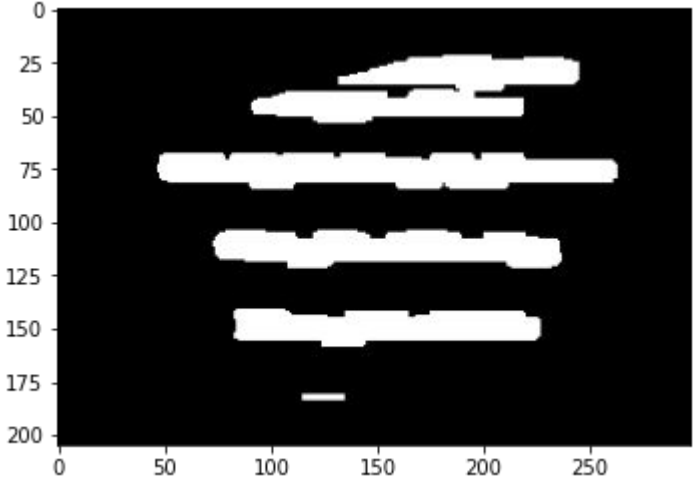
# Dish Name Segmentation

1. OTSU Thresholding on Rotated Image
2. Line Dilation
3. Merging the Bounding Boxes
4. Cropping out each bounding boxes from the rotated image

# OTSU Thresholding & Line Dilation




**OTSU Thresholding**

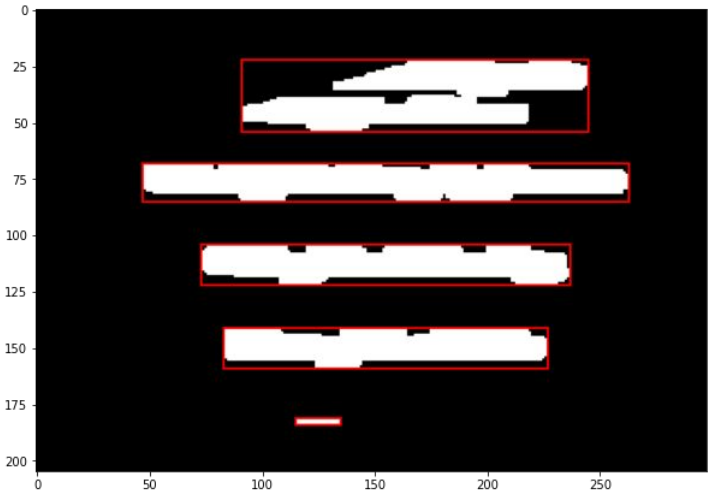


**Line Dilation**
Structural element : (10,2) rectangle

# Merging Bounding Boxes

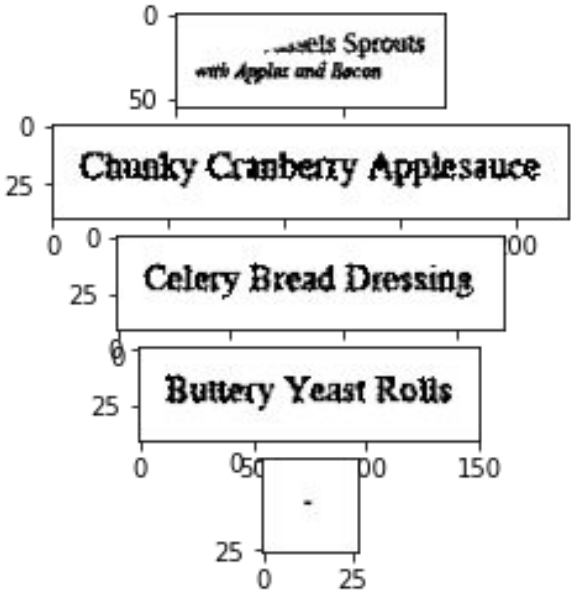
- There Maybe bounding boxes which only consist on one word of dish ,while other word is bounded by other bounding boxes

- So to create single bounding box consisting whole dish name , these bounding boxes which are almost in same line in horizontal orientation are merged



# Segmenting Dish Names from bounding boxes

Crop Out each Bounding Box , and then give it to the OCR engine



# OCR (Optical Character Recognition )

- MATLAB implementation of Tesseract algorithm, an open source OCR engine initially developed at HP Labs and currently managed by Google.

- Libraries Available for python “Pytesseract”
  https://pypi.org/project/pytesseract/

```
[['cwsvels Sprouts', 97, 20, 141, 35],
 ['Chunky Cranberry Applesauce', 53, 66, 203, 20],
 ['Celery Bread Dressing', 79, 102, 151, 21],
 ['Buttery Yeast Rois', 89, 139, 131, 21],
 ['', 121, 179, 7, 6]]
```

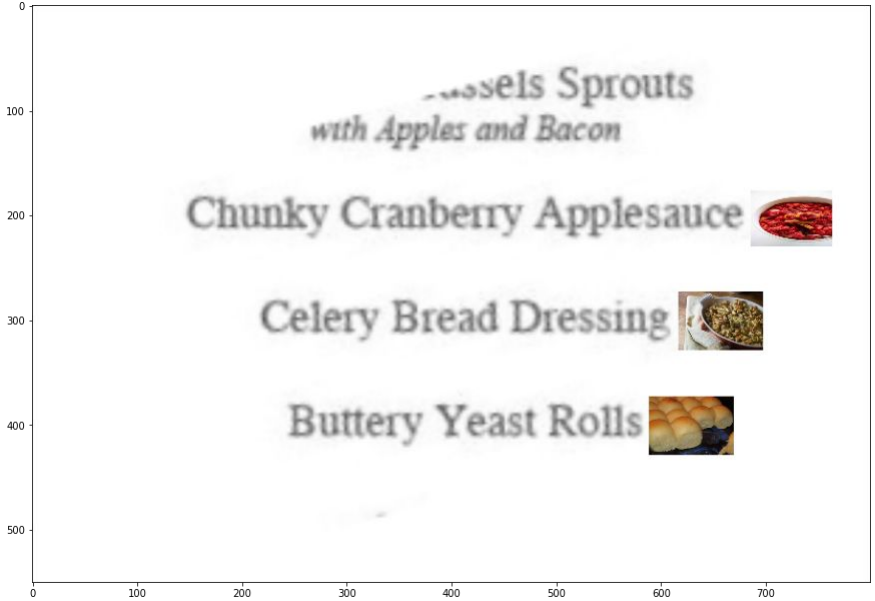
# OCR Correction Algorithm

- OCR May generate Spelling Errors
- Algorithm of finding minimum edit distance given acceptable number of character mismatch, our system is able to correct the results from the OCR engine, which increases the overall success rate of image lookup.
- Such as shortest Levenshtein distance, to resolve mismatched characters within a word

```
['cwsvels Sprouts', 91, 22, 154, 32, None]
['Chunky Cranberry Applesauce', 47, 68, 216, 17, 'chunky cranberry applesauce']
['Celery Bread Dressing', 73, 104, 164, 18, 'celery bread dressing']
['Buttery Yeast Rois', 83, 141, 144, 18, 'buttery yeast rolls']
['', 115, 181, 20, 3, None]
```

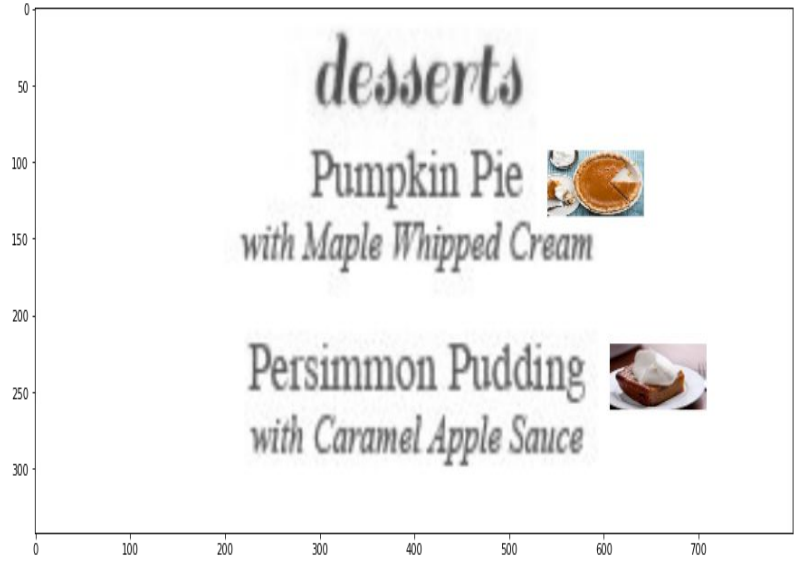
# Final Output Generation

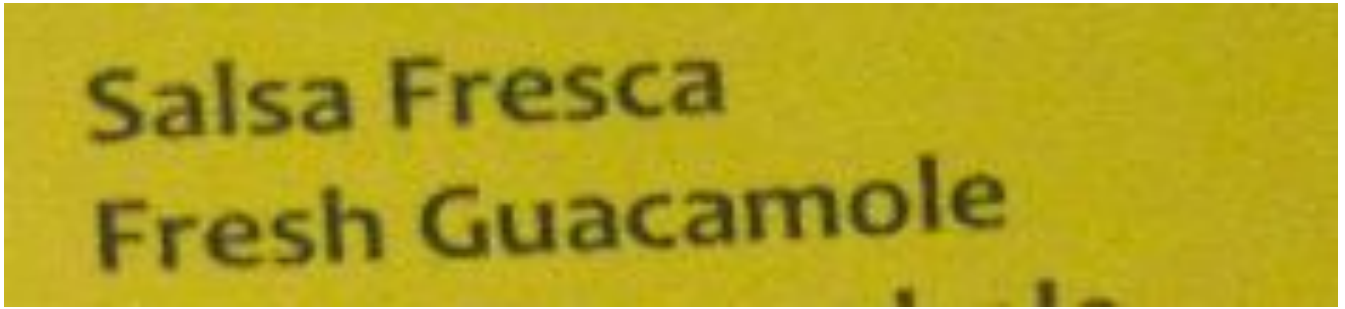
- Original Image is Rotated with angle resulted from the Featureless rotation
- Menu Image is scaled out ,to properly fit the dish images
- Dish Images are then resized according to their bounding boxes and pasted beside the Bounding box of the dish name

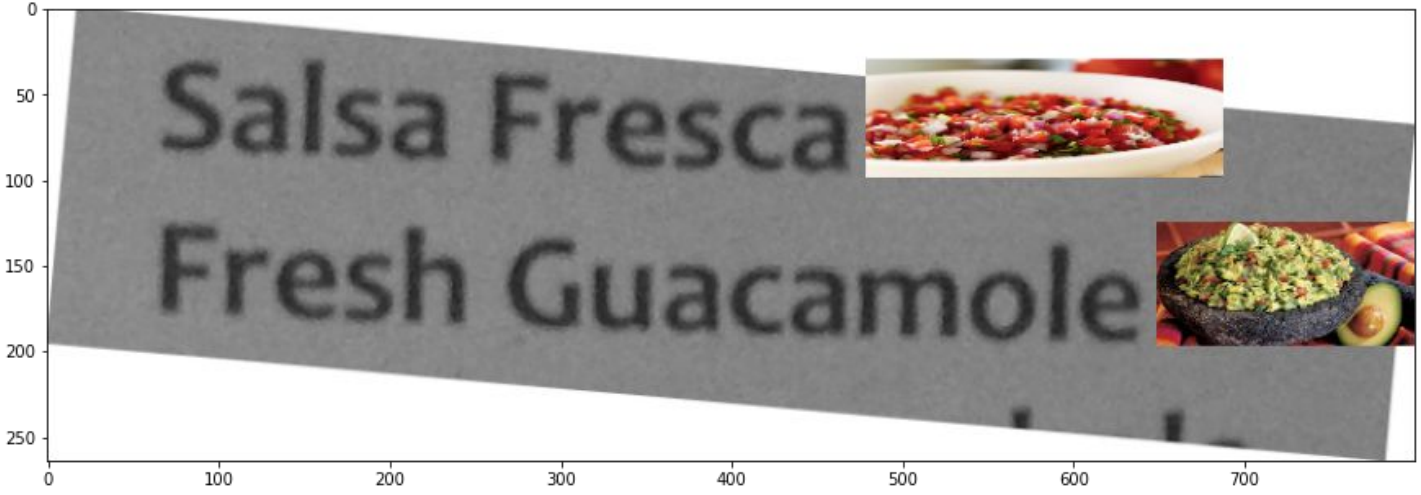


# Examples

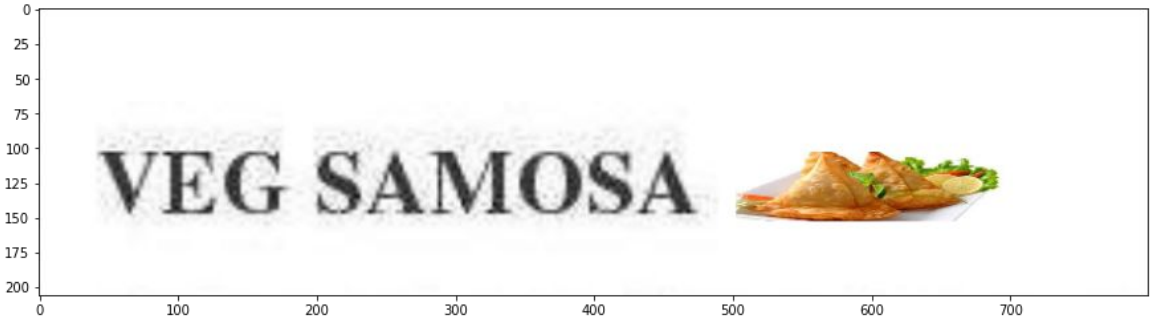
desserts
Pumpkin Pie
with Maple Whipped Cream
Persimmon Pudding
with Caramel Apple Sauce


desserts
Pumpkin Pie
with Maple Whipped Cream
Persimmon Pudding
with Caramel Apple Sauce
0
50
100
150
200
250
300
0
100
200
300
400
500
600
700

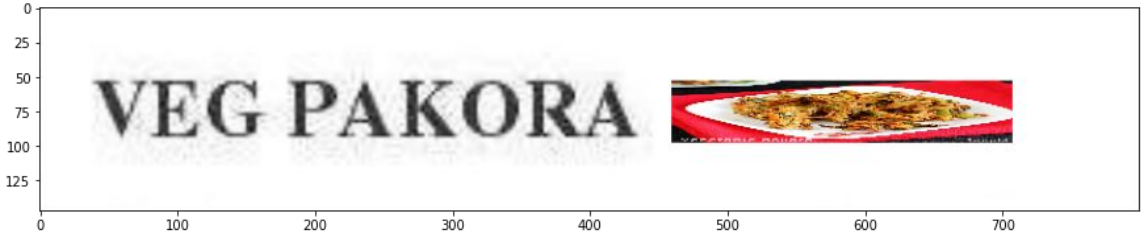

Salsa Fresca
Fresh Guacamole


Salsa Fresca
Fresh Guacamole
0
50
100
150
200
250
0
100
200
300
400
500
600
700

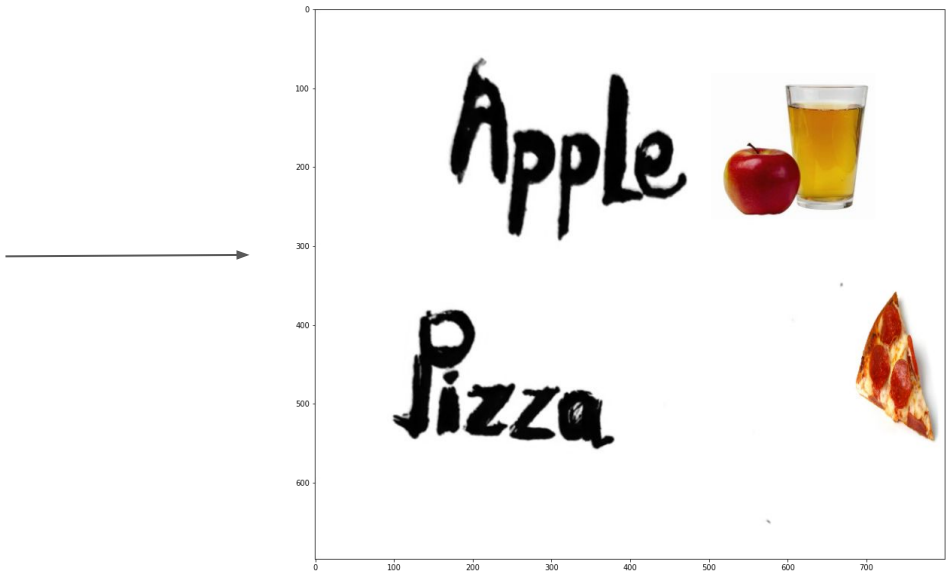
VEG SAMOSA

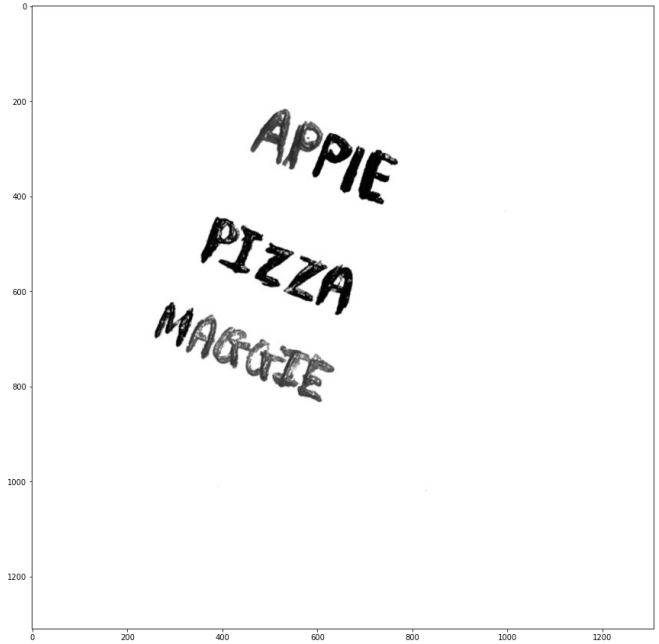

VEG SAMOSA
0
25
50
75
100
125
150
175
200
0
100
200
300
400
500
600
700

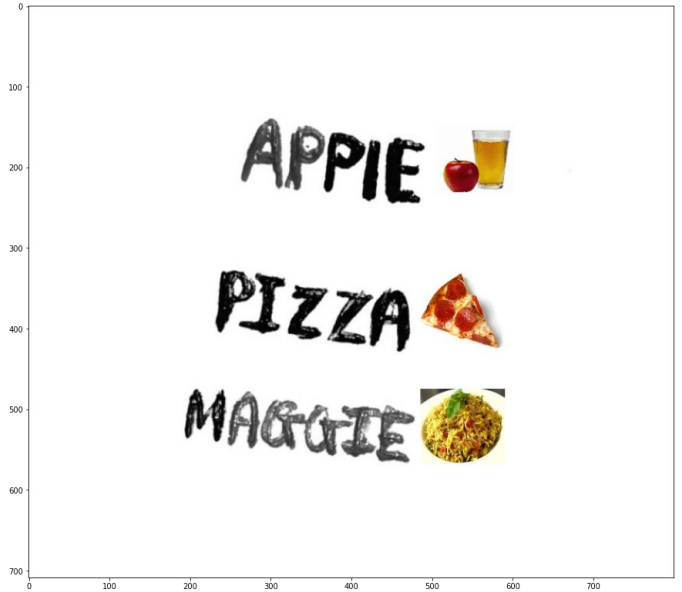
VEG PAKORA



Apple
Pizza


0
100
200
300
400
500
600
0
100
200
300
400
500
600
700
Apple
Pizza

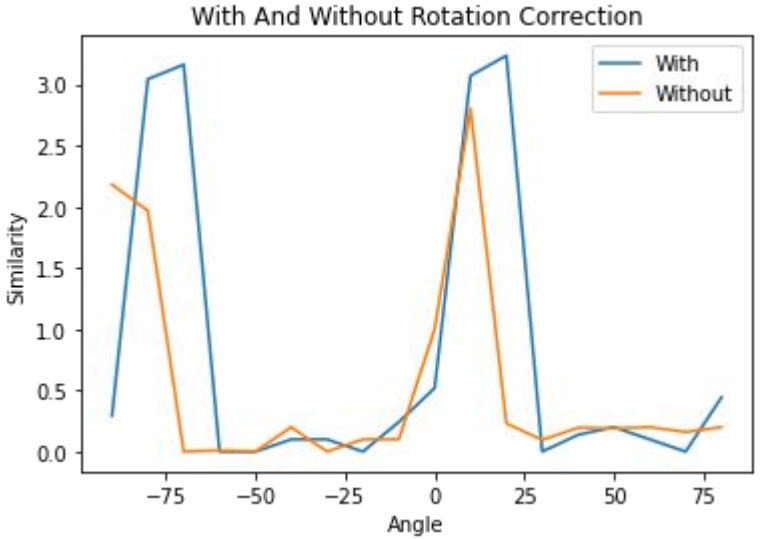

APPIE
PIZZA
MAGGIE


APPIE
PIZZA
MAGGIE

# Experimental Results

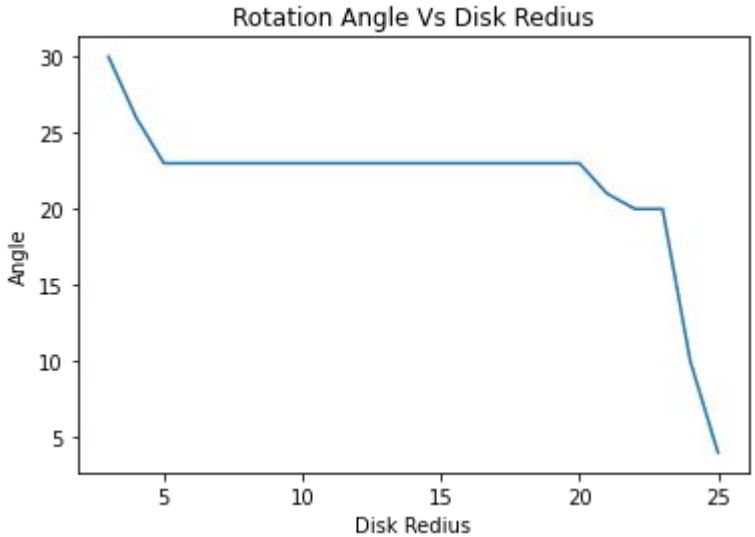
## 1) OCR Correctness with and without Rotation Correction

- OCR with rotation Correction Technique gives good result compare to the without rotation correction.
- Low Performance of OCR with featureless rotation is due to the resolution drop due to rotation



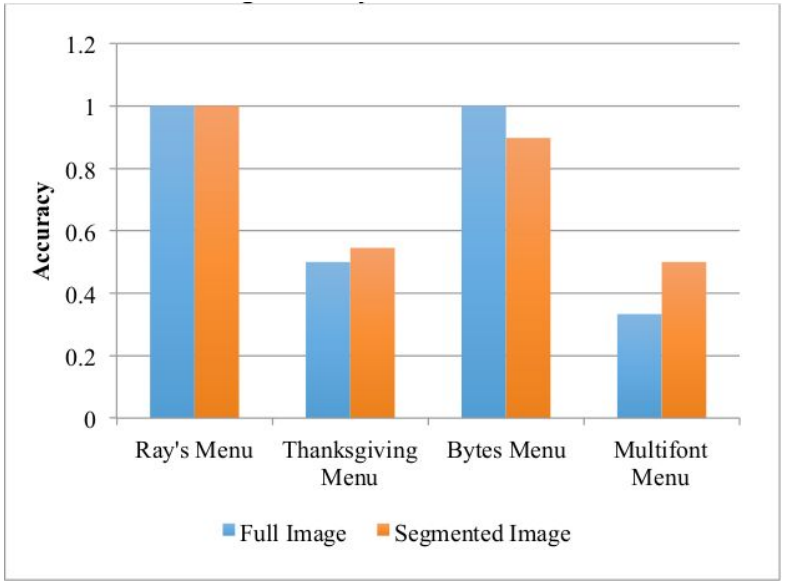
## 2. Featureless Rotation Angle generated from Different Disk Radius

- Without any dilation at all, the identified angle of rotation is bit different from other angles
    - This is because characters are now considered as one connected component
    - hence the assumption that aspect ratio is the largest when rotated the image back to the actual angle is no longer valid.
- When the structural element diameter is too big the identified angle of rotation is incorrect as well.
    - This is because all the words in the menu and the border of the menu is grouped into one single connected component and hence the aspect ratio does not represent the bounding box around a single English word




**Original Rotation Angle**：23

## 3. Passing Full image and Segmented Image into OCR




Source : https://web.stanford.edu/class/ee368/Project_Autumn_1516/Reports/Wang_Chen_Lang.pdf

# Issues with the pipeline

1) **OCR Accuracy Depends on the Font Styles and Fonts Sizes**

   Some Fancy font styles create inaccuracies in finding correct rotation angle which leads to the incorrect OCR Output

   Very less line space may create large connected components and that may cause error in rotation angle detection

2) **Noise , Intensity Gradient in the image heavily affects the Accuracy**

3) **Choosing the Thresholding Method**

4) **High allowed maximum edit distance in the OCR correction method may lead to wrong dish name selection.**

## Division Of Work

**Tirth Pandit** : Featureless Rotation , Experimental Analysis

**Jay Krishna** : Dish Name Segmentation , Experimental Analysis

**Smit Khanpara** : Final Output Generation , Examples

**Pratik Tiwari** : OCR Correction Algorithm , Examples

**Dharmesh Gusai** : Merging Bounding Boxes , Database Generation ,