NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

CZ4003 Computer Vision

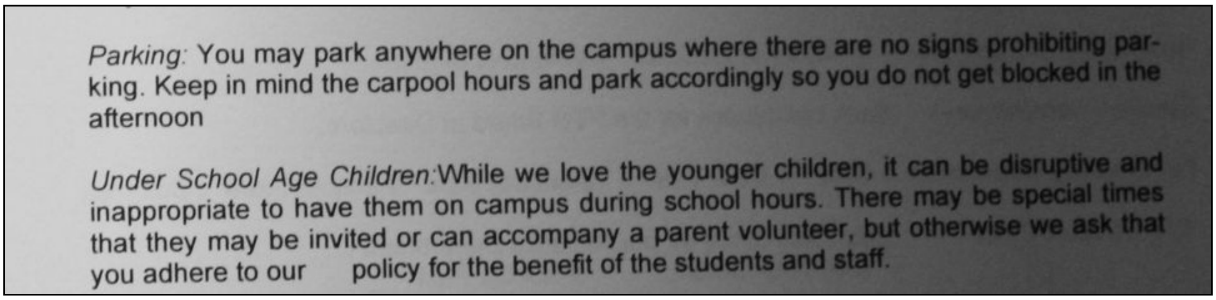
Project: Text Image Segmentation for Optimal Optical character recognition with Tesseract

Name: Lim Wai Leong

U1821194K

1. Implement the Ostu global thresholding algorithm for binarizing the sample text images and feed the binarized images to the OCR software to evaluate the OCR accuracy. Discuss any problems with the Otsu global thresholding algorithm.

sample01:



Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon

*Under School Age Children:*While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.

Results after performing otsu thresholding on the images:

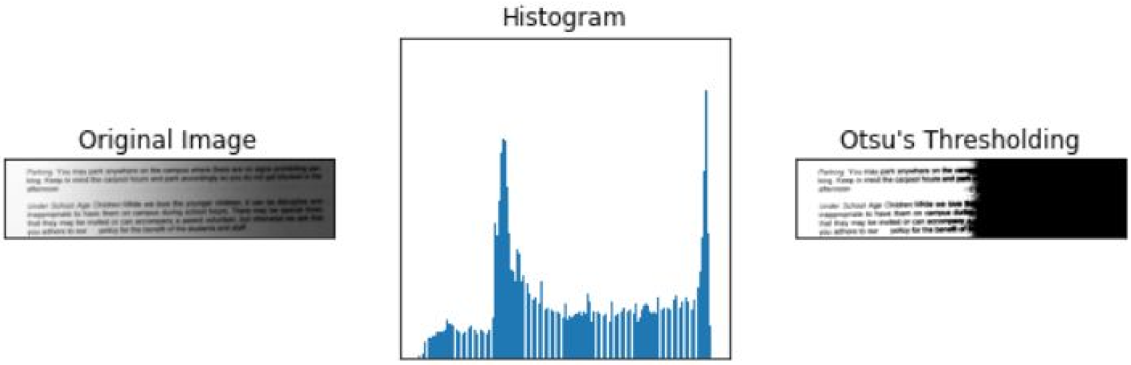
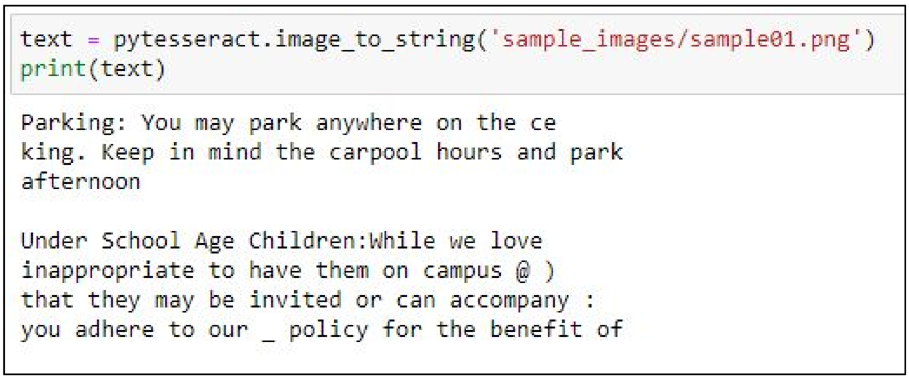


Figure 1: Otsu's Thresholding result on original image

Performance of Otsu's thresholding was not very good. It recognises the entire right half of the image as a background image and colours it black. The contrast in the image is too large and thus performance is not good.
Otsu's method avoids choosing an arbitrary value as a threshold and tries to determine it automatically. As seen in the histogram, there are 2 peaks of light and dark pixels, and as a result the determined threshold is not good.
Image pre-processing has to be done in order to improve the results.

Code and result of text extraction on the original image:



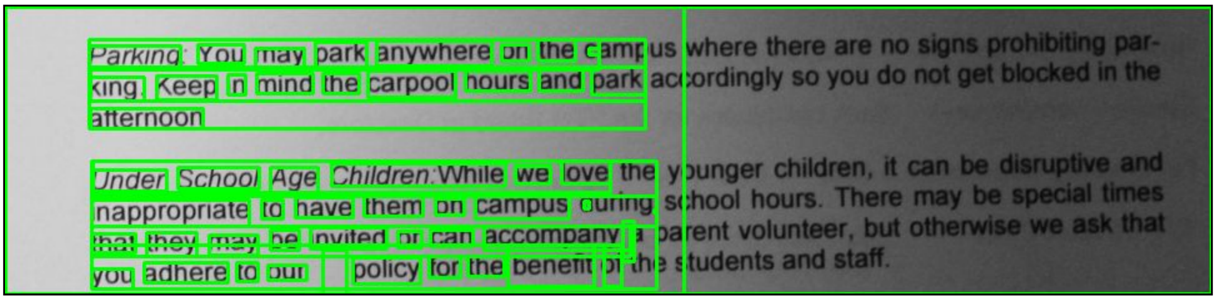
```
text = pytesseract.image_to_string('sample_images/sample01.png')
print(text)
```

```
Parking: You may park anywhere on the ce
king. Keep in mind the carpool hours and park
afternoon

Under School Age Children:While we love
inappropriate to have them on campus @ )
that they may be invited or can accompany :
you adhere to our _ policy for the benefit of
```

Figure 2: Text extraction on original image

To determine what tesseract was doing, bounding boxes of each text recognised was drawn on the image.



Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon

Under School Age Children:While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.

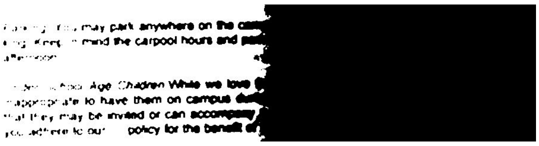
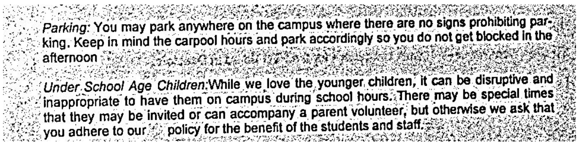
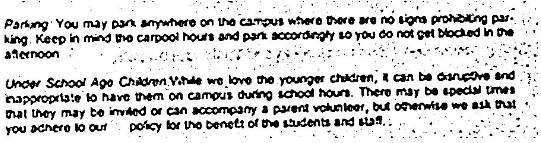
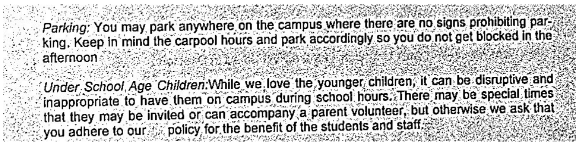
Figure 3: Bounding Box Recognition on original image

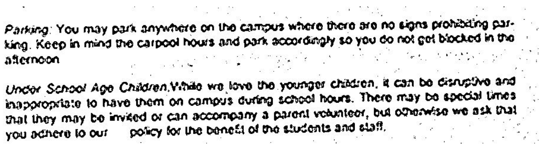
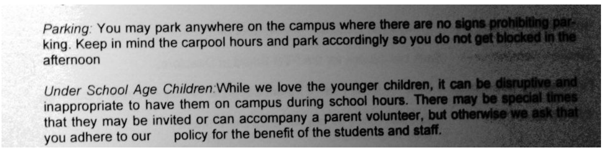
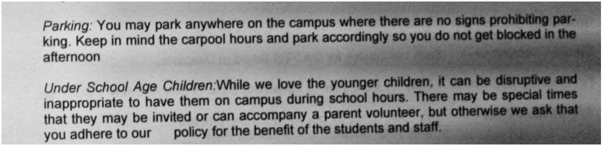
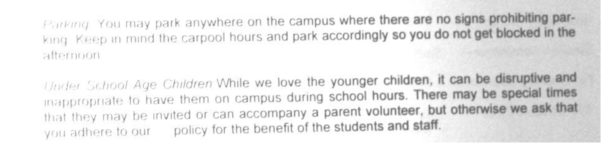
Conclusions

- Tesseract internally applies the Otsu algorithm (confirmed in Tesseract docs)
  - No text at all is recognised in the right side
  - **Need to preprocess image**
- Many additional bounding boxes, which do not even fit any text
  - **Need to refine bounding box selection**

Image Preprocessing

First, a few different methods and approaches were tested and evaluated

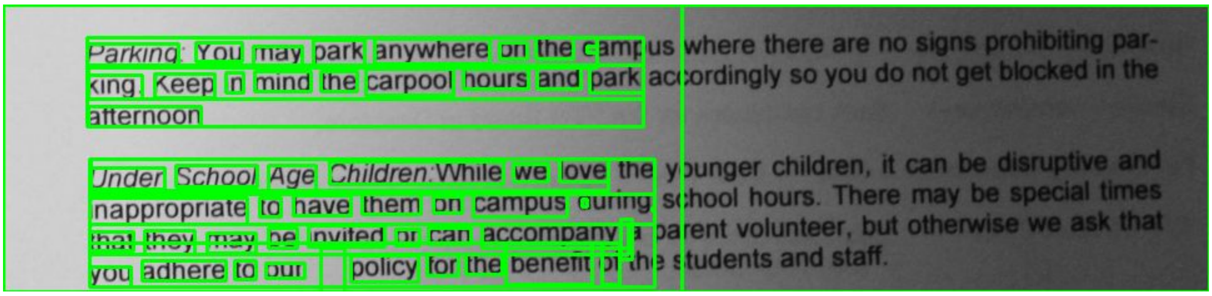
Approach	Evaluation
 Global Thresholding	Not good, similar to Otsu's thresholding, more than half the image is completely lost ~40% Accuracy
 Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon Under School Age Children:While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. Adaptive Mean Thresholding on raw image  Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon Under School Age Children:While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. Adaptive Mean Thresholding with median filter applied	Contrast is evened out, but details in the image are lost. Running text_to_string produces no meaningful output. Due to additional noise in the image, median filtering was used to remove some of the noise. However, only gibberish was produced by tesseract.
 Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon Under School Age Children:While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. Adaptive Gaussian Thresholding on raw image	Similar results to adaptive mean thresholding.

 Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon Under School Age Children:While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. Adaptive Gaussian Thresholding with median filter applied	
 Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon Under School Age Children:While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. Histogram equalization  Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon Under School Age Children:While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. CHALE Histogram equalization  Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon Under School Age Children While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. Increasing brightness	Histogram Equalization extended the readable range of text, but even after tuning with the CHALE object, the words in the right corner still could not be read. When increasing brightness, it was either too dim to read the words on the left, or too bright that words on the left would be obscured.

Bounding Box Selection

Using `pytesseracts image_to_data` function, the confidence level and content of the bounding boxes could be observed. After testing with various images, an appropriate level of 70% confidence was chosen. Boxes with null and blank content were also omitted.

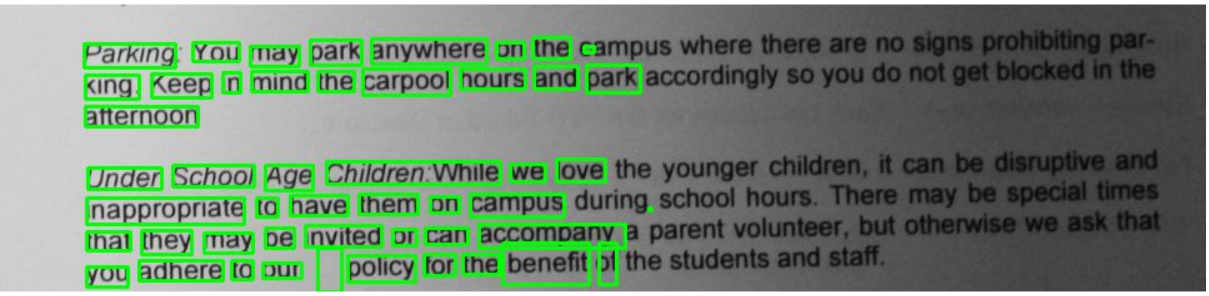
Before filtering:



Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon

Under School Age Children:While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.

After filtering:



Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon

Under School Age Children:While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.

A much cleaner result was obtained.

Development of algorithm

Main issue is that either bright parts are oversaturated or the dark parts remain too dark. Split the image based on brightness level and then perform brightness adjustment separately before recombining and reading the text.

1. Splitting the image

Global filtering is used first to find the bright and dark regions in the image. A very large median filter is used to make the regions more distinct before splitting the image at the boundary.

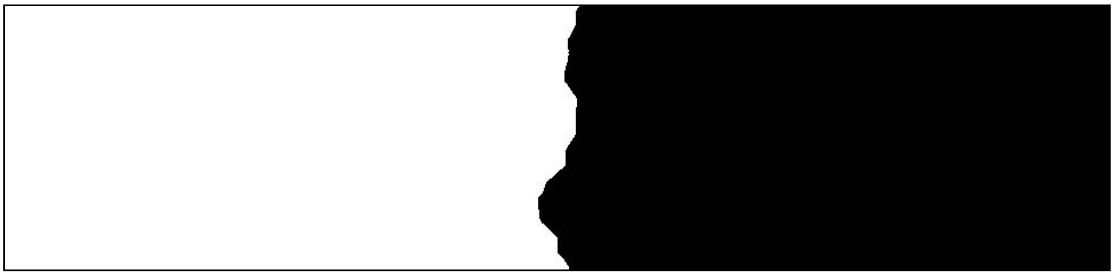
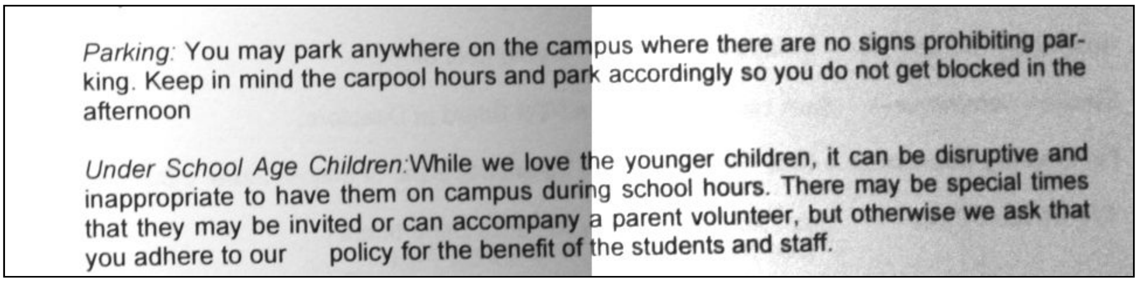


Figure 4: Median filter result

2. Brightness adjustment

The brightness is increased adaptively for each individual image. The average brightness is calculated, and then the image is increased in brightness on a ratio based on that brightness. Afterwards the 2 images are recombined.



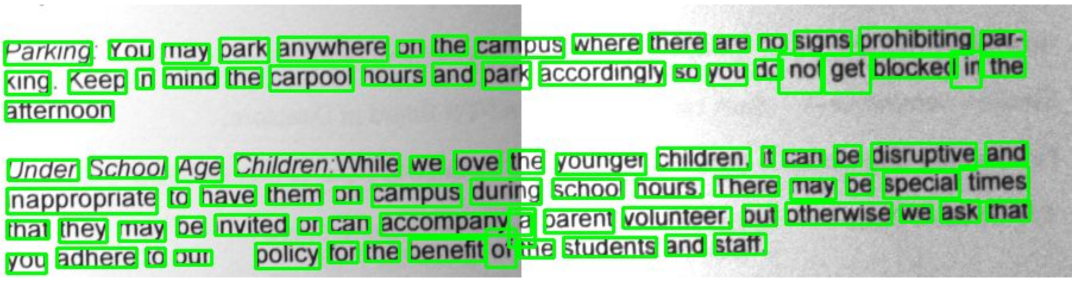
Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon

*Under School Age Children:*While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.

Figure 5: Brightness adjustment result

3. Drawing bounding boxes and extracting text

The bounding boxes are drawn with the same filtering



Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon

*Under School Age Children:*While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.

Figure 6: Drawing of bounding boxes

Resulting text with the same filters is as follows:

Parking: You may park anywhere on the campus where there are no signs prohibiting par- king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon Under School Age Children:While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.

100% accuracy is achieved

Application on image sample02

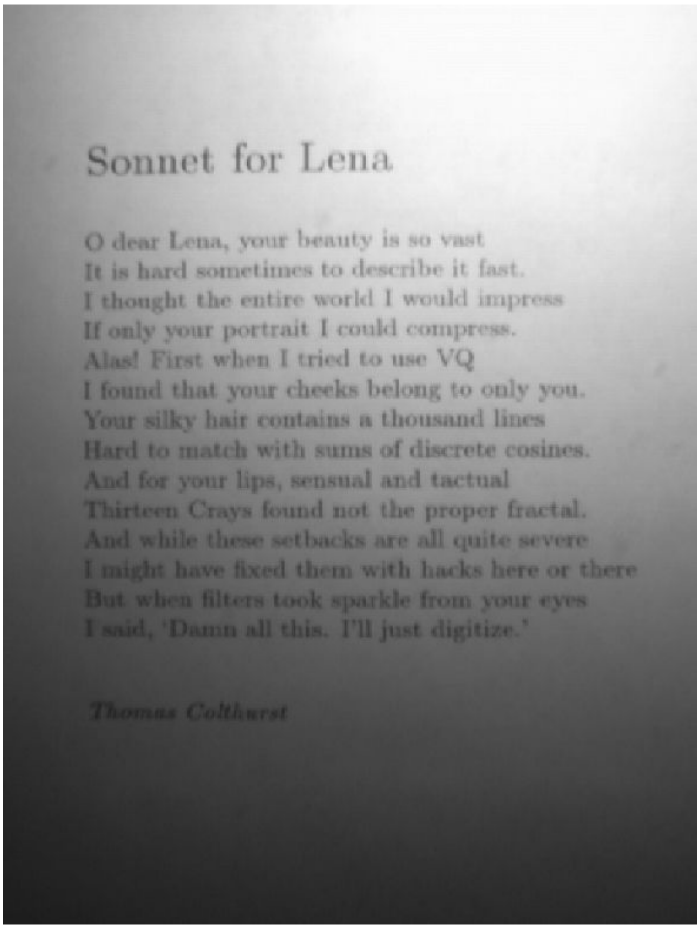
Image has even worse contrast than sample01, and text is slightly blurry.

Results of directly applying the `image_to_string` are as follows:

```
Sonnet for Lena

when I tried to use V

ur cheeks belong to or
```



Sonnet for Lena

O dear Lena, your beauty is so vast
It is hard sometimes to describe it fast.
I thought the entire world I would impress
If only your portrait I could compress.
Alas! First when I tried to use VQ
I found that your cheeks belong to only you.
Your silky hair contains a thousand lines
Hard to match with sums of discrete cosines.
And for your lips, sensual and tactual
Thirteen Crays found not the proper fractal.
And while these setbacks are all quite severe
I might have fixed them with hacks here or there
But when filters took sparkle from your eyes
I said, 'Damn all this. I'll just digitize.'

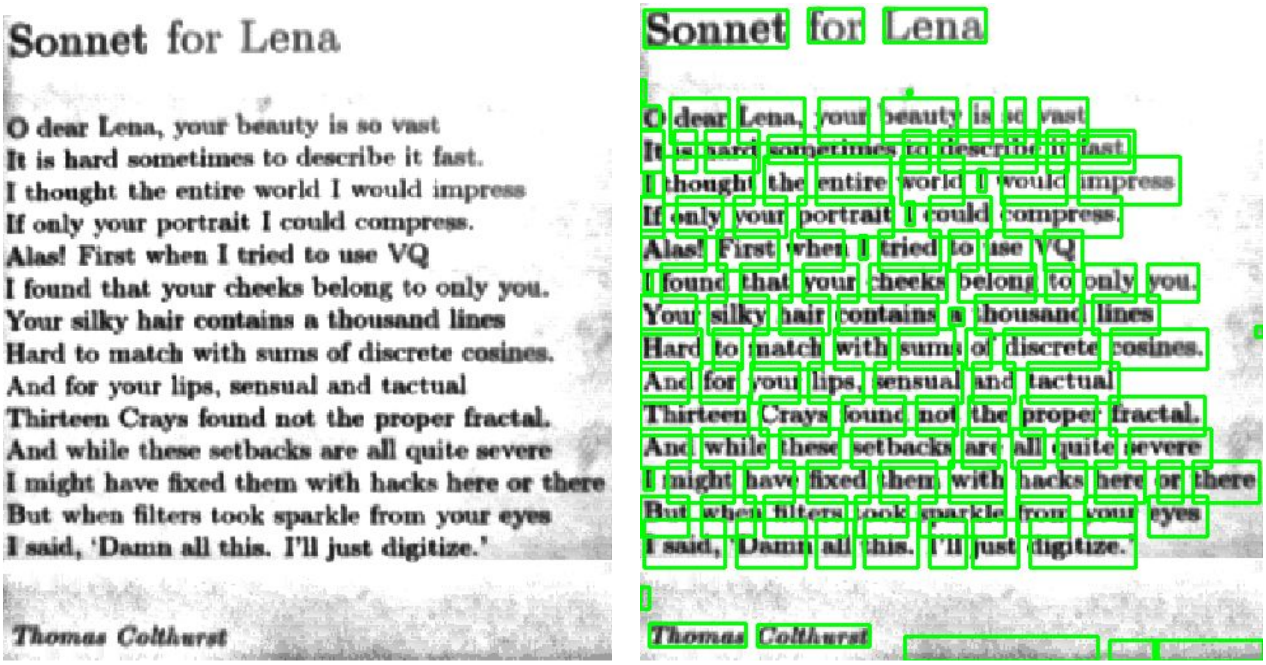
Thomas Colthurst

Figure 7: sample02.png

Some slight modifications to the algorithm applied to the first algorithm are done to improve the results on the 2nd image.

1. Manual cropping of the text was done
    a. Extreme contrast in the top corner and bottom made the brightness adjustments very bad even after splitting the image
2. Image was rotated before and after splitting
    a. Splitting works on left and right boundary, just a workaround to fit the image for the algorithm
3. Histogram equalization to help sharpen the image text
4. Bounding box confidence reduced to compensate for some of the unclear blurry text

After running the image through the algorithm the results are as shown:



Sonnet for Lena

O dear Lena, your beauty is so vast
It is hard sometimes to describe it fast.
I thought the entire world I would impress
If only your portrait I could compress.
Alas! First when I tried to use VQ
I found that your cheeks belong to only you.
Your silky hair contains a thousand lines
Hard to match with sums of discrete cosines.
And for your lips, sensual and tactual
Thirteen Crays found not the proper fractal.
And while these setbacks are all quite severe
I might have fixed them with hacks here or there
But when filters took sparkle from your eyes
I said, 'Damn all this. I'll just digitize.'

Thomas Colthurst

Sonnet for Lena

O dear Lena, your beauty is so vast
It is hard sometimes to describe it fast.
I thought the entire world I would impress
If only your portrait I could compress.
Alas! First when I tried to use VQ
I found that your cheeks belong to only you.
Your silky hair contains a thousand lines
Hard to match with sums of discrete cosines.
And for your lips, sensual and tactual
Thirteen Crays found not the proper fractal.
And while these setbacks are all quite severe
I might have fixed them with hacks here or there
But when filters took sparkle from your eyes
I said, 'Damn all this. I'll just digitize.'

Thomas Colthurst

Figure 8: Image results after processing

Text results are:

Sonnet for Lena j . © dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world | would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes: Hsaid, ‘Damn all this. I'll just digitize.” : Thomas Colthurst } a Fi

The text was read with close to 100% accuracy. Noise in the image contributed to some additional random words being read, however increasing the threshold would lead to a few words being left unread.

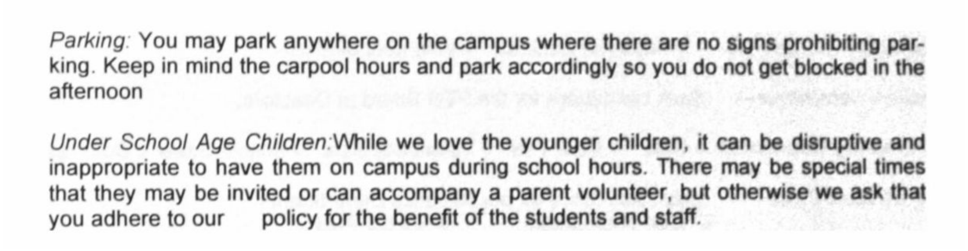
Improvement for more robust character recognition algorithms

Tesseract only works well for well structured, clean images, such as scanned documents or softcopy documents. Tesseract works very poorly when text is irregularly shaped, slanted or has noise in the image.

A simple improvement would be to change the way bounding boxes are drawn. Tesseract bounding boxes are defined by their two corners, a width and a height. This limits the way boxes can be drawn especially for slanted text. Drawing boxes with 4 corners instead would be an improvement. This implementation can be seen in other OCR algorithms such as CLOVA.

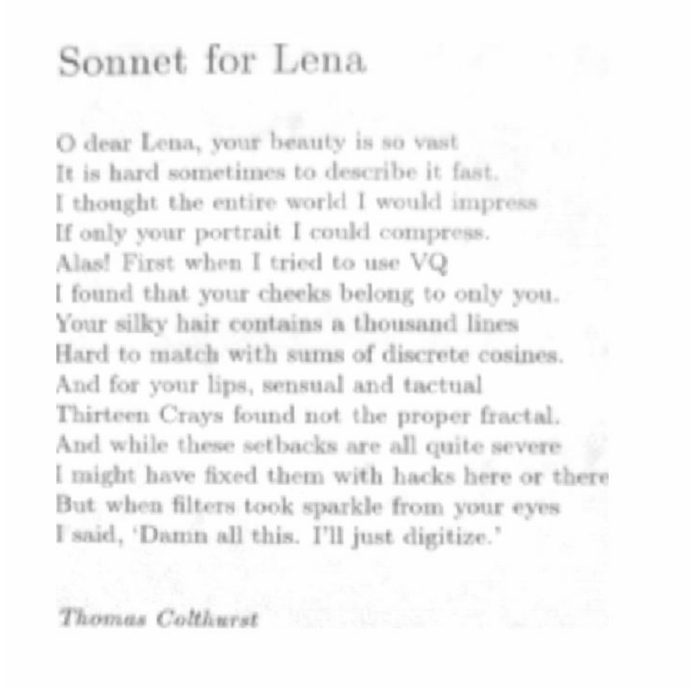
Running some tests confirmed that even text slanted by 10 degrees would be unrecognisable by tesseract.

A better contrast improvement can also be executed on the image. A more advanced background cleaning would lead to better results as shown below.



Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon

Under School Age Children: While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.



Sonnet for Lena

O dear Lena, your beauty is so vast
It is hard sometimes to describe it fast.
I thought the entire world I would impress
If only your portrait I could compress.
Alas! First when I tried to use VQ
I found that your cheeks belong to only you.
Your silky hair contains a thousand lines
Hard to match with sums of discrete cosines.
And for your lips, sensual and tactual
Thirteen Crays found not the proper fractal.
And while these setbacks are all quite severe
I might have fixed them with hacks here or there
But when filters took sparkle from your eyes
I said, 'Damn all this. I'll just digitize.'

Thomas Colthurst

The results from these images were much better for tesseract, with near 100% accuracy without any other image pre-processing.