Report: Su et al. Binarization Implementation

Giuseppe Tripodi

May 16, 2022

Abstract

The purpose of this report is to describe the implementation of the binarization technique described in the article [SLT10]. Much of the design choices will then be described, with references to the rationale behind those choices. At the end, the texts carried out with which the proposed solution was compared with the Otsu technique will be described, with reference to the advantages and disadvantages of both methods.

1 Introduction

Image binarization is a special preprocessing step that is performed to separate the foreground from the background in images. This process is often performed in different preprocessing steps of different document analysis mechanisms, such as OCR, Writer Identification, etc.

The conversion of an image to black-white format, though it reduces some of the information in the image itself, is useful to better distinguish the elements in the image, in addition binary images decreases the computational load and increases efficiency of the following preprocessing step.

Su et al's binarization technique [SLT10] is different from the other proposed techniques; it makes use of a contrast image, by which it allows for the determination of pixels with higher contrast that are then used in the actual pixel classification. This method was constructed for the analysis of historical documents.

2 Implementation

The implementation provided is similar with that described in the main article. I will then proceed to describe the most relevant aspects of the code.

2.1 Binarization Function

In the file *binarization_function.py* the functions for binarization have been implemented. Most of the functions invoke the methodologies described in the article.

The construction of the image contrast is self-explanatory; the values corresponding to the maximum and minimum intensity values are calculated first, after which the corresponding value is calculated for each pixel. This phase of the implementation lower the image background and brightness variation.

Of note is the implementation of the high-contrast image pixels. Since the constructed contrast image has a biomodal pattern, this part was implemented using Otsu global thresholding.

I used the function provided by OpenCV:

```
def high_contrast_image(self):
    gray = self.contrast_image
    blurred = cv2.GaussianBlur(gray, (7,7), 0)

    # applying Otsu threshdolding tecnique
```

```
ret, thresh1 = cv2.threshold(blurred, 0, 255, cv2.THRESH_BINARY + cv2.THRESH_OTSU
self.high_contrast_img = thresh1
#The high contrast pixel value is 1, because are the pixel above the threshold
self.high_contrast_pixel_value = 255
```

First, I applied a Gaussian blur to obtain a smooth image without noise and to remove some parts of the high-frequency image that may spoil the segmentation.

To the resulting image, I applied the Otsu thresholding method, which automatically estimates the threshold value and applies it to the image. Note that I also used *cv2.THRESH_BINARY* because I wanted to set all pixels larger than T to the output value. I have chosen as output value 255.

Regarding the text pixel classification, my implementation is very reminiscent of the one described in the paper. Of note, however, is the choice of model parameters, the size of the neighborhood window and the minimum number of the high contrast image pixels N_{min} within the neighborhood window.

In the paper, these parameters are estimated from the data; however, I felt it was more functional for assignment purposes to set these parameters a priori. Therefore, I performed numerous tests with different combinations of the parameters and eventually chose the values that guaranteed me better results.

2.2 Measures Function

The implementation of F-measure and PF-measure follows the basic algorithm. The implementation of PSNR is also very clear and self-explanatory.

On the other hand, the implementation of Distance-Reciprocal Distortion Measure is noteworthy. This method measures the distortion of the processed image compared with the original image using a weighted matrix. I have implemented this metric following the information in the paper [LKS04]. I started by calculating the weight matrix following the basic formula, after which I calculated the number of non-uniform blocks of size 8x8. After that, I calculated for each k the value DRD_k that correspond to the weighted sum of the pixels in the block B_k of the original image that differ from the flipped pixel $g[(x,y)_k]$ in the processed image.

3 Testing and methods valutation

I have compared Su et al. method with the Otsu's global thresholding method. These two methods though similar have some substantial differences that characterize them.

The *Su et al.* method on average gives better results, the images obtained are in fact less noisy and clearer. This is due to the fact that the method suppresses noise very well, as it suppresses the image background contrast during normalization. This allows only a few high contrast pixels to be selected in the background, and thus most of the background pixels will be excluded from the classification.

However, this technique also has some disadvantages and limitations. When the text strokes on the black side are darker than those on the front side, the method cannot correctly classify the two types of character strokes. In addition, the method is highly dependent on the high contrast pixels, an incorrect recognition of them in fact may preclude the whole classification. In regard to this, I believe that the recognition of high contrast pixels was one of the issues that prevented my implementation from achieving the same results provided in the paper.

On the other hand, Otsu's technique turns out to be much faster than the proposed method and still yields decent results when the image quality is poor anyway.

The disadvantages of this method, however, are several: first, the technique works well only in situations where the illumination is uniform and there is no noise present; in addition, the method fails if there are local variations in illumination. This method is also suitable for use in images with a biomodal distribution of gray values.

The tests performed basically confirm and highlight this substantial difference between the two techniques.

Method	Otsu	Su et al
F-Measure	0.71	0.57
p-F-Measure	0.12	0.17
PSNR	32.89	26.41
DRD	8.24e-05	5.77e-05

Table 1:

As anticipated, the two techniques were compared using the following measures:

- F-measure: is a measure of a test's accuracy, it is based on Precision and Recall
- p-F-measure: it is similar to the F-measure, the only difference is in this case we use a different Recall called pRecall that compare the input image with the skeletonized Ground Truth.
- PSNR (Peak signal-to-noise ratio): is a similarity measure, the measures how close the resultant image is to the ground truth image. The higher the value, the higher is the similarity of two images
- DRD (Distance Reciprocal Distortion Metric) [LKS04]: This method measures the distortion of a processed image compared with the original image using a weighted matrix, with each of its weights determined by the reciprocal of a distance measured from the center pixel. This metric measure the distortion, so a smaller value means less distortion.

The results obtained on the images provided under review are show in Table 1.

One can immediately see the difference in accuracy between the two methods, with the implemented method being significantly better than the Otsu method. The PSNR also highlights how the result obtained by the Su et al method is more similar to the GT image than the result of the Otsu technique.

One result to consider in the conducted experiment is that the first proposed solution has a greater distortion (DRD) than the solution provided by Otsu. I guess this is due to the bias introduced by the choice of parameter N_{min}.

References

[LKS04] Haiping Lu, Alex Kot, and Y.Q. Shi. Distance-reciprocal distortion measure for binary document images. *Signal Processing Letters, IEEE*, 11:228 – 231, 03 2004.

[SLT10] Bolan Su, Shijian Lu, and Chew Lim Tan. Binarization of historical document images using the local maximum and minimum. pages 159–166, 06 2010.