Scene Text Binarisation : Slide-Otsu and Kapur-Sahoo-Wong methods Improvization

Sagar Kumar Agrawal (M22AIE212)
Udayan Ghosh (M21AIE265)

Abstract

Scene text binarization is a difficult task due to text differences in the background noise, and uneven lighting in Natural Scene pictures. In our work we try to improve upon the existing Otsu's and Kapur Sahoo Wong's binarization method that leverages class variances of image histogram of the image to classify the pixel into background and foreground based. We propose a modified Otsu's method that improves upon the existing method by incorporating Gaussian Blur to reduce noise, enhances the contrast of the image using Histogram Equalization and next apply sharpening. Finally, it applies the Otsu's binarization using a sliding window technique such that binarization is performed on local context of image instead of the global context. And Kapur Sahoo Wong method maximizes entropy of image's histogram to get optimum threshold. Here we applied Gaussian Blur to reduce noise next applied Non-Local Mean Two Dimensional Histogram Equalization and finally applied Kapur Sahoo Wong's algorithm to get optimum threshold.

1. Introduction

Scene text recognition is a challenging problem. Text in natural scenes can appear in a variety of fonts and styles, making it difficult to recognise the characters, may be affected by variations in lighting conditions, such as shadows, glare, and reflection, may be captured from different angles, resulting in perspective distortion, may appear on a complex background. In our project we tested with two approaches that tries to further improve the performance of existing algorithms to get threshold.

Sagar Kumar Agarwal performed gray scaling of images further performed Gaussian Blur for noise reduction and Histogram Equalization for contrast enhancement sharpening the images and finally applied Slide-Otsu method to get optimum thresold value for binarization.

Udayan Ghosh performed gray scaling of images further performed Gaussian Blur for noise reduction and Non-Local Mean Two Dimensional Histogram Equalization to maximize entropy in the images and finally applied Kapur Sahoo Wong algorithm to get optimum thresold value for binarization.

1.1. Method 1 Image Enhancement + Slide Otsu :

Otsu is a popular thresholding technique that calculates the global threshold to classify the pixels either as foreground or background. This methods iteratively tests the multiple thresholds and exhaustively search for the threshold that minimises the intra-class variances.

$$\sigma_w^2(t) = w_0(t) * \sigma_0^2(t) + w_1(t) * \sigma_1^2(t) \tag{1}$$

$$\omega_0(t) = \sum_{i=0}^{t-1} p(i) \tag{2}$$

$$\omega_1(t) = \sum_{i=t}^{L-1} p(i) \tag{3}$$

Weights ω_0 and ω_1 are the probabilities of the two classes separated by a threshold t ,and σ_0^2 and σ_1^2 are variances of these two classes.

In our project we propose ***Slide-Otsu*** technique. In an image the pixels in a local region have high correlation compared with the pixels that are located far away. The classical Otsu uses calculates the global threshold value. Thus a single threshold is calculated. In our method we experimented with a sliding window, **(size approximately equals to imagewidth/35)**, which uses a horizontal and vertical stride equal to window width size and we compute the local thresholds.

In our approach we initially start with a chain of image enhancement techniques, Gaussian Blur , Image Histogram Equalisation , Image Sharpening to enhance the image quality and then apply ***Slide-Otsu*** on enhanced image.

1.2. Method 2 Image Enhancement + Kapur Sahoo Wong Algorithm

Kapur Sahoo Wong algorithm is a Maximum entropy method for accurately segmenting image. This method

works by maximizing entropy of image's histogram. This algorithm was developed for gray image segmentation which determines optimum threshold values for maximizing the entropy of histogram.

The algorithm by Kapur, Sahoo and Wong considers the foreground and background images as two distinct sources with the following distributions

$$b : p(i) = \frac{p_i}{P_t}, 0 \leq i \leq t \tag{4}$$

$$w : p(i) = \frac{p_i}{1 - P_t}, t + 1 \leq i \leq 255 \tag{5}$$

Then one calculates the entropy of the two sources

$$H_b(t) = -\sum_{i=0}^{t} p(i)log(p(i)) \tag{6}$$

$$H_w(t) = -\sum_{i=t+1}^{255} p(i)log(p(i)) \tag{7}$$

where p(i) is determined by Equations 4 and 5. After, one determines the optimal threshold that maximizes the sum of the two entropies.

$$H(t) = H_b(t) + H_w(t) \tag{8}$$

In our project we propose Non-Local Mean Two Dimensional Histogram Equalization on Kapur Sahoo Wong algorithm which uses the technique of maximizing entropy of image's histogram.

Initially we apply Gaussian blur to remove noise from the picture. Then Non-Local Mean Two Dimensional Histogram Equalization and finally apply Kapur Sahoo Wong algorithm.

2. Related Work

Valley-seeking is the most commonly used method for extracting objects from a picture, but pay-level histograms are not always bimodal. So Different algorithms use different techniques to detect the object in the image and thus the find the optimum threshold values. Like Kapur Sahoo Wong uses entropy maximizing histogram to get the optimum value of the threshold value and Otsu method try to find global minima or maxima for the whole image to finalize the threshold value. Those are not always the optimum threshold values as a result binarized images are not always optimum. Here we are try to improvise the above mentioned methods of binarization.

3. Methodology

In this section we present the methodologies for the two methods proposed :

3.1. Method 1 Image Enhancement + Slide Otsu :

1. Pre-processing the input image using Gaussian Blur to reduce noise.
2. Enhancing the contrast of the image using Histogram Equalization.
3. Sharpening the image to enhance the high frequency information in image
4. Execute ***Slide - Otsu*** on after image enhancement steps

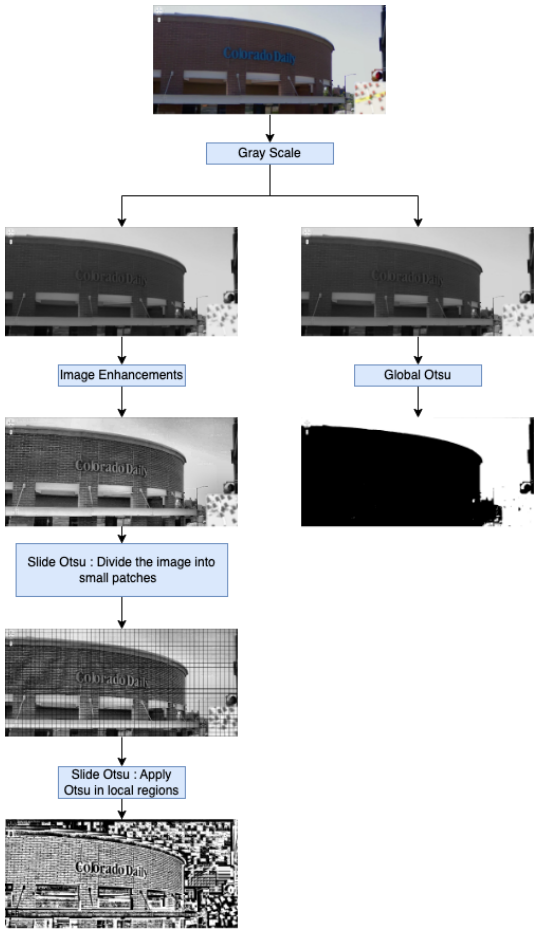

Figure 1. Slide Otsu Vs Otsu

3.2. Method 2 Image Enhancement + Kapur Sahoo Wong Algorithm:

1. Pre-processing the input image using Gaussian Blur to reduce noise.
2. Enhancing the contrast of the image using Non-Local Mean Two Dimensional Histogram Equalization.
3. Execute ***Kapur Sahoo Wong*** on image after enhancement steps

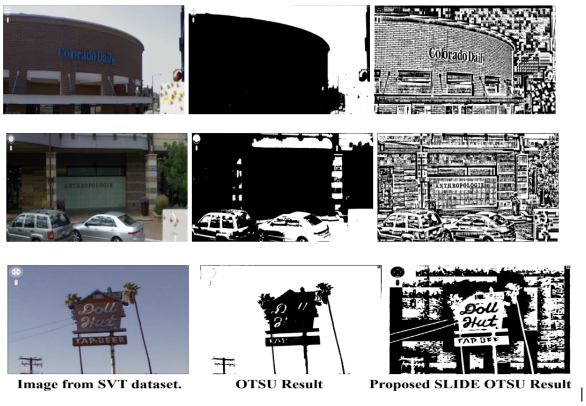

Figure 2. Results comparison

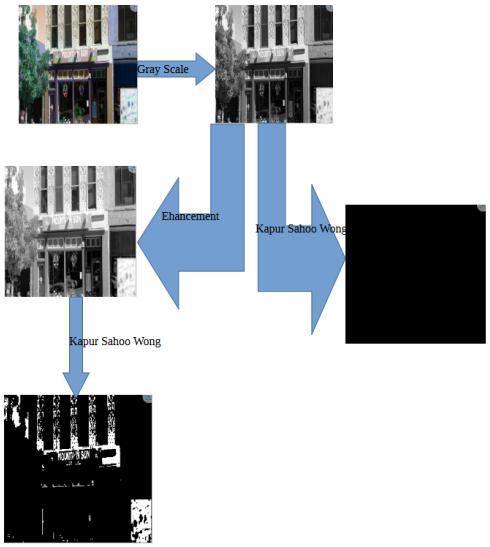

Figure 3. Enhanced Kapur Sahoo Wong vs Kapur Sahoo Wong

Figure 4. Results comparison

4. Model Performance and Evaluation

In this section we focus upon the evaluation of the performances of the proposed methods and their comparison with the original methods :

4.1. Method 1 Image Enhancement + Slide Otsu :

Evaluation Steps :

1. Run the algorithm to convert the Scene Text images to Binary images
2. Use off-the shelf OCR technique, for our project we used **Keras - OCR** to extract text from binary images
3. Calculate similarity score ratio measure the similarity of extracted text with the ground truth and define a similarity score

   For evaluating the quality of extraction we have used fuzzy logic for string matching. We have calculated the partial ratio raw score which is a measure of the strings similarity.The partial ratio helps us to perform substring matching. This takes the shortest string and compares it with all the substrings of the same length. This helps us to identify if the extracted result is closer to the ground truth.

   For our project we have tested our algorithms on **322+ images from SVT dataset**. It is a dataset that was harvested from Google Street View. In **Table 1.** we compared Similarity Score Distribution of Otsu vs Slide Otsu. In our results we can see that with the proposed Slide Otsu method, the average similarity with the ground truth has improved by **5 %** from **57.27 %**

to **62.1 %** that too with a lesser standard deviation as compared to the Original Otsu.

	similarity score otsu	**similarity score slide otsu**
Image count	322	322
Mean	57.27 %	62.10 %
Std. Dev.	26.44%	21.12%
25th Percentile	38.00 %	48.25 %
50th Percentile	58.00 %	60.00 %
75th Percentile	75.00 %	76.75 %
Max	100.00 %	100.00 %

Table 1. Comparison of Similarity Score between Extarcted Text and Ground truth : Otsu vs Slide Otsu.

4.2. Method 2 Image Enhancement + Kapur Sahoo Wong Algorithm :

Evaluation Steps :

1. Run the algorithm to convert the Scene Text images to Binary images
2. Use off-the shelf OCR technique, for our project we used **Keras - OCR** to extract text from binary images
3. Calculate similarity score ratio measure the similarity of extracted text with the ground truth and define a similarity score

   For evaluating the quality of extraction we have used fuzzy logic for string matching. We have calculated the partial ratio raw score which is a measure of the strings similarity.The partial ratio helps us to perform substring matching. This takes the shortest string and compares it with all the substrings of the same length. This helps us to identify if the extracted result is closer to the ground truth.

   For our project we have tested our algorithms on **349 images from SVT dataset**. It is a dataset that was harvested from Google Street View. In **Table 2.** we compared Similarity Score Distribution of Kapur Sahoo Wong vs Modified Kapur Sahoo Wong. In our results we can see that with the proposed Modified Kapur Sahoo Wong method, the average similarity with the ground truth is quite similar but max has improved by **17 %** from **33 %** to **50 %**.

4.3. Links

GitHub Link: Link
YouTube Video Link: Link

4.4. References

[1–6].

	similarity score Kapur Sahoo Wong	**similarity score Modified Kapur Sahoo Wong**
Image count	349	349
Mean	4.12 %	4.40 %
Std. Dev.	6.56%	6.64%
75th Percentile	9.00 %	9.00 %
Max	33.33 %	50.00 %

Table 2. Similarity Score Distribution Comparison : Kapur Sahoo Wong vs Modified Kapur Sahoo Wong.

References

[1] A.K. Bhandari, A. Kumar, and G.K. Singh. Modified artificial bee colony based computationally efficient multilevel thresholding for satellite image segmentation using kapur's, otsu and tsallis functions. 4

[2] João Marcelo Monte da Silva, Rafael Dueire Lins, Fernando Mário Junqueira Martins, and Rosita Wachenchauzer. A new and efficient algorithm to binarize document images removing back-to-front interference. 4

[3] J.N. Kapur, P.K. Sahoo, and A.K.C. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. 4

[4] manuelaguadomtz. Kapur threshold calculation. 4

[5] N. Otsu. A threshold selection methodfrom gray level histograms. 4

[6] Abhay Sharma, Rekha Chaturvedi, Sandeep Kumar, and Umesh Kumar Dwivedi. Multi-level image thresholding based on kapur and tsallis entropy using firefly algorithm. 4