A New Approach to Picture Quality-independent Robust Text Binarization using Differential Binning

The Overview of the steps involved

1. **Foreground – Background Pixels split**: The first step involves identifying and splitting the pixels of the image into sets of candidate foreground and candidate background pixels. We used a differential binning approach which gives different sets of candidate foreground and background splits. This is the critical step as the next set of steps depends on the fact that there is at least one split where the text pixels fall into one set of pixels and with very little background pixels in the same set.
2. **Homogeneous region detection:** Step 1 gives several sets of candidate foreground and background pixels. This step involves identifying clusters of connected pixels which represent a homogenous region in the image. A homogeneous region is defined as a set of connected pixels in an image whose deviation in RGB values is less than the deviation from the surrounding pixels of that cluster.
3. **Text, non-Text Separation:** The Homogeneous regions extracted in the form of connected set of pixels from Step 2 are now classified into text and non-text using a classifier model. Regions classified as Non-Text are removed and candidate Text regions are retained for combination
4. **Probabilistic Combination:** Step 3 gives us multiple sets of candidate Text pixels across multiple bin sizes used in the differential binning step. We identify the probability of a pixel being a text pixel based on its occurrence across the different bin sizes. We label Pixels in the final binarized image based on a probability outcome
5. **Foreground – Background Pixels split: High Frequency Binning**

The first target is to identify sets of foreground and background pixels and split them. We use two-level overlapped binning and Adjacent Bin Recombination

1. **Two-Level Overlapped Binning**

For a given Bin Size s, we divide the entire range of RGB Euclidian distances into two levels of Bins

*For a given Bin size s:*

**Level 1** *(Bin number 1 to k)*

*= {( 0 to s-1), (s to 2\*s – 1), (2\*s to 3\*s – 1) ,*

*……*

*(k-1\*s to k\*s) }*

**Level 2** (Bin number k+1 to 2\*k -1)

*= {(0.5\*s to 1.5\*s – 1), (1.5\*s to 2.5\*s – 1),*

*…...*

*((k-1.5)\*s to (k-0.5)\*s) }*

*Where where is the maximum value possible for a point.*

Each Bin is represented as a binary image which is the kth bin for a given bin size S and ()ij is the value of the ith row and jth column of

= 1 *if*

()ij  = 1 *if*

= 0 otherwise

PIXEL-VALUE(i,j) =

R(i,j) = Red Value of pixel at ith row and jth column of image

G(i,j) = Green Value of pixel at ith row and jth column of image

B(i,j) = Blue Value of pixel at ith row and jth column of image

For a given bin size s, there are total bins where ,each represented by a binary image where the first images are the Level 1 bins and the next images are Level 2 bins.

In order to achieve robustness and quality independence, we use *Decremented Interval Binning*– using multiple bin sizes with the difference between corresponding bin sizes decreasing the higher we go. At this stage, let it be clear that when we say foreground or background pixel, it is for convenience and we do not label any specific set of pixels as foreground or background. Any one of the sets can correspond to either foreground or a background the next steps will do the work of eliminating the sets of background pixels.



Fig 1. Recombined Bins Flowchart. The Bin Size denotes the size of the original Bins and the Range is the new size after the Adjacent Bin Recombination step

The lines connect subset bins to their superset bins. For example, Bin (374 to 406) is a subset of Bin (367 to 420) and is connected by a line. Later, this relationship will be used to determine stable regions and extract features for text non-text separation.

1. **Adjacent Bin Recombination**

Consider a Homogenous region Ri with PIXEL-VALUE in the range of 30 to 60. We want all homogenous regions that are less than size s to fall into at least one of the bins of size s. If suppose bin size is 32 we want it to fall entirely in one of the bin. However, Let’s look at the pixel values that the first few bins of both the levels in two-level binning accumulates

Level 1

{(0 to 31), (32 to 63), (64 to 95), (96 to 127)…. }

Level 2

{(16 to 47), (48 to 79), (80 to 111) …}

We see in both the levels, the region Ri fails to fall into any of the bins despite having a pixel value deviation less than the particular bin size.

To solve this problem, we create a new set of Recombined Bins called as Δ bins

**LEVEL 1**

**LEVEL 2**

Where, +1

*Note: + operator while adding bins is pixel-wise logical OR operator.*

Considering the previous example, Let us see the first few bins for bin size 32

LEVEL 1

{ ( 0 to 47),(32 to 79),(64 to 111) ..…}

LEVEL 2

{(16 to 63),(48 to 95) , (80 to 127)….}

Region Ri falls in the first bin of Level 2.

We state the following lemma and it’s proof for any Region Ri. A region is a set of points on an image that is connected

Lemma : *If the difference between the pixel with the highest PIXEL-VALUE and lowest PIXEL-VALUE in Region Ri is D and let S be any bin size that is selected such that then if , ()ij = 1 for one value of positive integer k*

Proof: Let the minimum pixel value be for pixel (xa,ya) and maximum be for pixel (xb,yb) such that .

Let +1

Then,

then from the definition of ()ij ,

Now

Since,

Thus,

Also,

Thus,

Since is an integer

Case 1:

In that case, both and . For any ,

Since is an integer

Since is arbitrary , this implies that

for all

ij

Now,

Since, ij for all

ij  = 1 for all

Case 2 :

We have to consider two different sub cases depending on value of

Sub Case 1 :

Let us consider bin

Now,

…..*condition of subcase*

Thus,

Thus,

From Definition of ()ij

()XbYb = 1

Bin k has values ranging from and Bin m has the values

To

*Inference 1: Bin k and Bin m combined has all the pixel values P such that*

The pixel ranges Bin k and Bin m captures are overlapping. Since Pmin is in bin k and Pmax in bin m ,for any the following holds

From definition of

From Inference 1 and the previous inequality we thus prove

ij  = 1 for all

Subcase 2 :

, Since

Thus,

+1

Let

From definition of ()ij  , ()XaYb = 1

Bin k+1 has values ranging from and Bin m has the values

To

*Inference 2: Bin k+1 and Bin m combined has all the pixel values P such that*

The pixel ranges Bin k+1 and Bin m captures are overlapping. Since Pmin is in bin m and Pmax in bin k+1,for any the following holds

From definition of

From Inference 2 and the previous inequality we thus prove

ij  = 1 for all

*Thus for all the cases the lemma gets proved.*

During the proof we draw two important inferences, each of which gives us the range for a bin. The range of the delta bins is

The Lemma tells us that if there is a text region and if the range of PIXEL-VALUES in the text is D, all the pixels will occur as a part of a connected component in a bin of size D(or more). Moreover, the bin where it will occur,if as close to D as possible(best if it is D) then the chances of background pixels occurring in the same bin is also very less assuming that the deviation within a Text(or a homogenous region) is much less than the deviation between the Text pixels and the background pixels immediately surrounding the Text region. We find this to be true in nearly all the scene text images we come across.

Thus, the bins which are represented as binary images capture several homogenous regions for several different bin sizes. A connected component in one binary image of bin size s will have the same or more pixels for the next bin size. next

1. **Homogenous Region Detection: Identifying stable regions using ensemble of classifiers on connected components**

Step 1 gives several sets of candidate foreground and background connected components. All of the components do not correspond to homogenous regions in the original image. Bin sizes smaller than the deviation in a homogenous region will break up the connected components into pieces. Fig 1 shows how the size of the connected component increases with increasing bin size. For a particular connected component, after a particular bin size, the size will stop increasing as rapidly and rate of increase of size of connected component will be minimum.

This property is used to extract un-fragmented homogenous regions. We compare rate of increase of the size of connected components when the range of pixel values in a binary image increases. We describe stable homogenous regions as regions which has