Performance Analysis of Optical Character Recognition Using Adaptive Binarization of Degraded Document Images

M. Manimaraboopathy, M. Anto Bennet*, A. Priya, S. Vijayalakshmi, D. Hemavathy
Department of Electronics and Communication Engineering, VELTECH, Chennai-600 062
*Corresponding author: E-Mail: bennetmab@gmail.com

ABSTRACT

The planned OCR formula to retrieve the text within the scanned document pictures. Here the text detection formula supported two machine learning classifiers: one permits generating candidate word regions and therefore the different filters out non-text ones. The extract connected elements (CCs) in pictures by victimization the maximally stable extremal region formula. In CC cluster adaboost classifiers are accustomed confirm whether or not the region contains text or not. Then victimization binarization methodology, the grey image is converted into binary image. The binarization outcomes are subject to OCR and therefore the corresponding results evaluated with relevance character and word accuracy. As additional and additional text documents are scanned quick and correct. Extra performance metrics of the proportion rates of broken and uncomprehensible text, false alarms, ground noise, character enlargement and merging. This effectiveness of the planned methodology is additionally confirmed by tests carried on realistic document pictures. For planned formula MATLAB version thirteen package is employed.

KEY WORDS: Maximally Stable Extremal Regions (MSER), optical character recognition (OCR).

1. INTRODUCTION

Document image binarization could be a very important pre-processing step to document image analysis and recognition. Also, it area unit typically thought of as an important stage in OCR code systems since the results of consequent steps is very captivated with its effectiveness. Often the explanation why document image binarization has been a subject of intensive analysis throughout the last decades. The excellence sweetening methods, then at abstraction filtering methods that sharpen edges and exclude voluminous of the image blur. The image shown inside the Fig.1 is that the input image given for binarization. (Detector standardization is typically the first step of the image sweetening chain, but this was mentioned earlier as a district of the device modelling). For simplicity, assume that the images have associate eight-bit dynamic range; i.e., there are twenty eight = 256 accomplishable gray levels, that the gray levels inside the image area unit getting to be inside the vary 0–255, with zero being black and 255 being white. color footage have three arrays of numbers sometimes representing the red, green, and blue footage that are combined to relinquish the whole spectrum of colours. The image shown inside the Fig.2, is that the binary result. Specialize in method single-band footage, i.e., black and white footage.

The image is split into segments, denoted by (S), of fastened size. In every phase, the frequency of black pixels is calculated. The chosen segments type areas by connecting neighboring segments in regard to their original position within the image. The row-by-row labeling algorithmic rule is employed for scanning the document by the window. The parameter k within the formula determines the sensitivity of the detection technique. The upper the k, the less segments are going to be detected. The image shown within the Fig.3 is that the sample for noise. Noise space may be having two sort foreground and background. In the user was aided by code to form the bottom truth for machine-printed pictures, by merging and rending clusters within the character clump stage, still as by adding and removing character models to degraded character instances within the character matching stage. The image shown in Fig.4, is that the original image. However, the aforesaid procedure may be applied solely on machine-printed document while not many alternative font varieties.

Figure.1. Input image
Figure.2. Binarization result
Figure.3. Sample noise image
Figure.4. Original image
Figure.5. Ground truth
In the projected methodology, the event of ground truth plays a significant role, since it aids towards the automation of the analysis procedure. It consists of two distinct stages, specifically Skeletoned Ground Truth (SG) stage and calculated Ground Truth (EG) stage. The image shown in Fig.5 is that the bottom truth image. Transcript mapping (or text alignment) techniques unit of measurement utilised so as to map the correct text information to a segmentation result created automatically. Usually, these techniques unit of measurement really useful thus on automatically manufacture benchmarking data sets. The image shown in Fig.6, sample for skeletonized ground truth image. They are within the main supported hidden Markov models (HMMs) and dynamic time distortion (DTW) and within the main focus on the alignment of piece of writing pictures with the corresponding transcription on word level. Associate in Nursing economical transcript mapping technique to ease the event of document image segmentation ground truth that options text-image alignment in text line, word and character level.

**Figure 6. Sample for skeletonized ground truth image**

By concluding this noise in background and foreground can be quantitatively measured by using the garbou filter. Different type of noise can be detected and the ground truth images used for matching the input with grayscale image. Finally the degraded document can be detected.

**Literature Review:** Lamirjoy (2011), projected techniques used here were mathematical logic, Self-organized neural networks. The tactic done here was used the image color values and additional native spatial choices extracted at intervals the neighbourhood of the pixels. Every image and native choices values feed a kohonen self-organised feature map neural network classifier. Associate in nursing light-weight document merely modified but desires any spatial knowledge. Low and Maxenchuk (2010), projected techniques used here were Bi-modal distribution, optimization. The tactic done here was a powerfully bi-modal image with sleek regions in every the foreground and background, whereas effort sharp discontinuities at the perimeters. Then bi-modal and average that live desired properties in text image but it provides low-resolution image. Moghaddam and Cheriet (2010), projected techniques used here were Sauvola’s binarization technique, Automation and optimum selection of binarization technique. the tactic done here was tried to match binarization algorithms by victimization the exactness and recall analysis of the resultant words at intervals the foreground or by evaluating their result on end-to-end character or word recognition performance in a) very complete archive document recognition system utilizing OCR. The tactic provides quality footage but this model is not over parameterized. Matsui (2010), projected techniques used here are accommodative filtering, restoration, scanning. The tactic done here was that the correction of show through of the written documents. It offers rate improvement in thresholding and successfully eliminating show through from back side but the efficiency is not much better Anto Bennet (2012), projected techniques are MLIR where it'll notice the information that expressed in any language. This paper measuring the effectiveness of MLIR system, this explaining the measures that ar oft used for document retrieval. Here the efficiency to guage the performance of the system is not plenty of. Anto Bennet (2015), projected techniques used here were Binarization formula, Hybrid formula. the tactic done here was binarization techniques focus either on finding Associate in Nursing acceptable world threshold or adapting associate a neighborhood threshold for each space therefore on get obviate smear, strains, uneven illumination etc. Here, a hybrid approach is provided that initial applies a world thresholding technique then, identifies the image aras that area unit plenty of most likely to still contain noise. The method is economical but formula used is Anto Bennet (2016), projected classifying info victimization Boosting formula performs supervised learning that's understood as machine learning meta-algorithm. Boosting ways in which are sometimes used to notice objects or persons in videoconference, security system, etc. as Associate in nursing approximation of provision regression, or exaggerated with arithmetic enhancements of calculation of weight coefficients. This paper provides an honest survey of the literature on mining with rare classes and rare cases victimization Boosting techniques that shows original approach to classification and its variants. Fully completely different analysis metrics on rarity mining are mentioned throughout this paper, but error reduction is not economical.

**2. PROPOSED METHOD**

CC-based ways in which use a bottom-up approach by grouping very little elements into successively larger elements until all regions are famous among the image. A geometrical analysis is needed to merge the text elements exploitation the arrangement of the elements so on filtrate non-text elements and mark the boundaries of the text regions. A CC-based methodology might section temperament into multiple CCs, significantly among the cases of polychrome text strings and low-resolution and shouting video photos. Further, the performance of a CC-based methodology is severely littered with half grouping, sort of a projection profile analysis or text line selection. In addition, several threshold values are needed to filtrate the non-text elements, and these threshold values are addicted to the image/video info. A block-matching formula exploitation the mean absolute distinction criterion is employed to estimate the motion. Blocks incomprehensible throughout following are discarded. Their primary focus is on caption text, like pre-title sequences, credit titles, and motility sequences, that exhibit an improved distinction with
the background. This makes it straightforward to use the excellence distinction between the boundary of the detected elements and their background among the filtering stage. Finally, a geometrical analysis, along with the dimension, height, and ratio, is utilized to filter any non-text elements. Supported experiments exploitation 2247 frames, their formula extracted eighty six to one hundred pc of all the caption text. A binarized character with missing foreground pixels(false negatives) from the contour that do not have a control on the character topology, compared to a binarized character that the shortage of an analogous amount of foreground pixels alters the character topology achieves: a) equal performance once the everyday measures of Recall or PSNR ar used, as a result of an analogous amount of false negative pixels, b) higher performance once the distance-based measures MPM and DRD ar used, as a results of those measures apply lower group action getting ready to very cheap truth image.

**Candidate Generation:** For the generation of candidates, extract CCs in pictures and partition the extracted CCs into clusters, wherever the agglomeration rule relies on a contiguity relation classifier. During this section, initial CC extraction methodology shown. Then (i) to make coaching samples (ii) to coach the classifier, and (iii) to use that classifier in our CC agglomeration methodology.

**MSER Algorithm:** The core of the vision system are Maximally Stable Extremal Regions, or MSERs, introduced by Matas etc. all for gray-scale pictures and later ex-tended to paint as Maximally Stable Color Regions, or MSCR. Details regarding MSER and MSCR principles square measure given severally. The most usage of MSER detection is for wide-baseline image matching primarily as a result of its variance and high repeatability. To match 2 pictures of an equivalent scene (taken from totally different viewpoints), MSERs are extracted from each pictures and so fitly delineated victimization (usually in-variant) descriptor. As a result of MSER extraction is very repeatable, the bulk of the regions ought to be detected in each pictures.

**CC Extraction:** For any 2 nodes s and t in a very graph, their connected parts square measure either identical or disjoint begin BFS from some node s. this offers one part of the graph choose any presently undiscovered node u begin another BFS. This offers another part. Continue during this manner till all nodes square measure explored aimless graphs: s-t property, Directed graphs.

**Mutual Connectivity:** If u and v square measure reciprocally approachable, and v and w square measure reciprocally approachable, then u and w square measure reciprocally approachable. If u and v square measure reciprocally approachable, and v and w square measure reciprocally approachable, then u and w square measure reciprocally approachable. To travel from u to w, we are able to go via v. to travel from w to u, we are able to once move to via v.

**Strong Connectivity of Digraphs:** A Directed graph is powerfully connected if each 2 nodes, u and v square measure reciprocally approachable from one another analogous to property in aimless graph. The sturdy part containing s is that the reciprocally approachable from one another. Analogous to connected part containing s in aimless graphs. For any nodu n R, u and s square measure reciprocally approachable at any 2 nodes u and v in R square measure reciprocally approachable and u and s square measure reciprocally approachable; s and v square measure reciprocally reachable. Hence, u and v square measure reciprocally approachable. The image shown in Fig.8, is that the original image and text extracted from the first image. For any node u in R, u and s square measure reciprocally approachable any 2 nodes u and v in R square measure reciprocally approachable u and v square measure reciprocally approachable and s and v square measure reciprocally reachable. Hence, u and v square measure reciprocally approachable for any 2 nodes u and v, the sturdy parts of u and v square measure either identical or disjoint.

![Figure 8](image.png)

**Figure 8. Original image and extraction of text from original image**

Text in images can exhibit many variations with respect to the following properties:

**Size:** Although the text size can vary a lot, assumptions can be made depending on the application domain.

**Alignment:** The characters in the caption text appear in clusters and usually lie horizontally, although sometimes they can appear as non-planar texts as a result of special effects. This does not apply to scene text, which can have various perspective distortions. Scene text can be aligned in any direction and can have geometric distortions.

**Inter-character distance:** characters in a text line have a uniform distance between them.

**Color:** The characters in a text line tend to have the same or similar colors. This property makes it possible to use a connected component-based approach for text detection. Most of the research reported till date has concentrated on finding text strings of a single color. However, video images and other complex color documents can contain text strings with more than two colors for effective visualization, i.e., different colors within one word.

**Motion:** The same characters usually exist in consecutive frames in a video with or without movement. This property is used in text tracking and enhancement. Caption text usually moves in a uniform way: horizontally or vertically. Scene text can have arbitrary motion due to camera or object movement.
Edge: Most caption and scene text is designed to be easily read, thereby resulting in strong edges at the boundaries of text and background.

Compression: Many digital images are recorded, transferred, and processed compressed format.

CC grouping or Clustering: The main aim of CC grouping is to group adjacent characters detected in the previous steps into separated meaningful words and further reject false positives. Based on the observation that characters in the same word usually share similar properties, such as intensity, size, stroke width etc., this valuable information can be utilized in CC grouping. For CC grouping AdaBoost classifier is used which is use full in finding the adjacency relationship from CC. bounding box of c_i and denote its width and height as w_i and h_i, respectively. Given a pair (c_i, c_j) ∈ C×C (i not equal to j), the horizontal distance, horizontal overlap, and vertical overlap between two boxes are denoted as d_{ij}, h_{ij}, and v_{ij} respectively.

AdaBoost Algorithm: This paper presents an algorithmic rule for detective work and reading text in town scenes. This text includes conventional forms corresponding to street signs, hospital signs, and bus numbers also as a lot of variable forms corresponding to search signs, house numbers, and billboards. The info of town pictures were taken in part by ordinarily sighted viewers and part by blind volunteers UN agency were amid sighted guides mistreatment automatic camera settings and tiny sensible data of wherever the text was situated within the image. The databases are labeled to alter U.S.A. to coach a part of our algorithmic rule and to guage the algorithmic rule performance. The negative examples were obtained by a bootstrap method the same as Drucker. 1st elite negative examples by arbitrarily sampling from windows within the image dataset. When coaching with these samples, applied the AdaBoost algorithmic rule to classify all windows within the coaching pictures (at a variety of sizes). Those misclassified as text was then used as negative examples for preparation AdaBoost. The image regions most simply confused with text we're vegetation, repetitive structures corresponding to railings or building facades, and a few probability patterns. The image shown in Fig.9 is that the positive examples used for coaching AdaBoost. The previous section delineated the weak classifiers used for coaching AdaBoost.

Figure 9. Positive examples used for training AdaBoost.

Text Reading: Then applied business OCR software system to the extended text regions (produced by AdaBoost followed by extension and binarization). This was used each to browse the text and to discard false positive text regions. Overall, the AdaBoost robust classifier (plus extension/ binarization) detected ninety seven.2 % of the visible text in take a look at dataset (text that would be detected by a usually lynx-eyed viewer). For typical samples of the text that AdaBoost fails to find.

Most of those errors correspond to text that is blurred or badly unambiguous. Others occur as a result of don't train AdaBoost to detection vertical text or individual letters. (The coaching examples were horizontal segments typically containing 2 or 3 letters/digits). For the 286 extended text regions properly detected by the AdaBoost robust classifier (plus extension/binarization), then obtained an accurate reading rate of 93.0 xmas (proportion of words properly read). This needed a preprocessing stage to scale the text region. The seven xmas errors area unit caused by little textareas. For text which will browse with success and for text that can't browse.

After CC agglomeration, we've a group of cluster, normalizing these clusters corresponding regions for the reliable text/non-text classification. Additionally there are a unit vital variations between text and face stimuli as a result of the spatial variation per constituent of text pictures is way bigger than for faces. Face expression, equivalent to eyes, are a unit in more or less a similar spatial position for any face and have similar look.

Geometric Normalization: Given w_i ∈ W, first localize its corresponding region. Even though text boxes can experience perspective distortions, approximating the shape of text boxes with parallelograms whose left and right sides are parallel to y-axis. This approximation alleviates difficulties in estimating text boxes having a high degree of freedom (DOF): only have to find a skew and four boundary supporting points. To estimate the skew of a given word candidate w_k, build two sets:

\[ T_k = \{ t(c_i) | c_i \in w_k \} \]
\[ B_k = \{ b(c_i) | c_i \in w_k \} \]

Where t (c_i) and b (c_i) are the top-center point and the bottom center point of a bounding box of c_i, respectively illustration of B_k. For every pair in B_k and T_k, the slope of a line connecting the pair is discretized into one of 32 levels in [−\pi/8, \pi/8], and each pair votes for the skew angle. After voting, the most common angle is considered as a skew. Then, perform geometric normalization by applying an affine mapping that transforms the corresponding region to a rectangle.
Binarization: Given geometrically normalized pictures, binary pictures area unit to be thought of. However, acting the binarization singly by estimating text and background colours. It is as a result of (i) the MSER results might miss some character parts and/or yield rip-roaring regions (mainly because of the blur) and (ii) have to be compelled to store the purpose data of all CC for the MSER based mostly binarization think about the typical color of CC because the text color and think about the typical color of a complete block because the background color. Then, acquire a binary price of every constituent by scrutiny the distances to the calculable text color and also the calculable background color. 12 norms in RGB area unit used.

Text/Non-text Classification: Developing a text/non-text classifier that rejects non-text blocks among normalized images. In this classification, do not adopt sophisticated techniques such as cascade structures, since the number of samples to be classified is usually small. However, one challenge in this approach is the variable aspect ratio.

Feature Extraction from a Square Block: For the feature extraction, divide a square block into four horizontal and four vertical blocks and extract the features. The image shown in Fig.11 for the feature extraction, by horizontal and vertical blocks.

For a horizontal block $H_i$ ($i = 1, 2, 3, 4$), consider

<table>
<thead>
<tr>
<th>$H_1$</th>
<th>$H_2$</th>
<th>$H_3$</th>
<th>$H_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_1$</td>
<td>$V_2$</td>
<td>$V_3$</td>
<td>$V_4$</td>
</tr>
</tbody>
</table>

Figure 11. For the feature extraction, by horizontal and vertical blocks

Multilayer Perceptron Learning: For the coaching, normalized pictures. For this goal, the algorithmic program is employed (i.e., candidate generation and normalisation algorithms) to the coaching pictures. Shows the multilayer perceptron with 2 hidden layers. Then manually classified them into text and non-text. Discarded some pictures showing poor binarization results, and picked up 676 text block pictures and 863 non-text block pictures. However, it's been found that additional negative samples square measure required for the reliable rejection of non-text parts and picked up additional negative samples by applying constant procedure to pictures that don't contain any text three, 568 text pictures. Multi-layer perceptron is trained for the classification of sq. patches. One hidden layer is employed consisting of twenty nodes and set the output price to +1 for text samples and zero otherwise. To assist the educational, input options square measure normalized. Finally, text is alone detected by filtering the non-text areas.

Image Enhancement: Feature vectors are extracted to measure useful information from the decomposed sub images. Many feature vectors have been use document image binarization. Most of them were applied to printed documents with clean (white) backgrounds but did not work well for degraded images. Three feature vectors are proposed in this paper, which focus on handwritten document image with messy background and faded writing.

OCR: OCR addresses the matter of reading optically processed characters and has become one in every of the foremost prospering applications of technology within the field of pattern recognition and AI. Most optical character recognition (OCR) systems are designed to remodel text pictures to legible text codes, however perform poorly once text is embedded into complicated background thanks to background interferences and low frequency of incidence of text. An Adaboost text detection formula supported machine learning techniques is projected. To be precise this methodology 2 classifiers are used: one classifier was designed to get candidates and also the different classifier was for the filtering of non-text candidates. MSER methodology to take advantage of multi-channel info. Optical character recognition that is employed to match the image with the bottom truth image. Finally original pictures are retrieved from the degraded image. During this methodology yields the progressive performance each in speed and accuracy.
3. CONCLUSION

In this paper an Adaboost scene text detection algorithmic rule supported machine learning techniques is projected. To be precise this methodology 2 classifiers are used: one classifier was designed to come up with candidates and therefore the different classifier was for the filtering of non-text candidates. MSER methodology to take advantage of multi-channel data. Optical character recognition that is employed to match the image with the bottom truth image. Finally original pictures are retrieved from the degraded image. During this methodology yields the progressive performance each in speed and accuracy. During this methodology is intended to handle the text detection downside in document pictures, wherever English alphabets are placed horizontally. This methodology ought to be modified so as to notice Asian scripts and texts of absolute orientations. The overall framework ought to be extended.

REFERENCES


