AN IMPROVED BINARIZATION METHOD USING INTER- AND INTRA-BLOCK FEATURES FOR NATURAL IMAGES

Jufeng Yang, Kai Wang, Jiao Jiao, Jing Xu

College of Information Technical Science, Nankai University, China
{yangjufeng,wangk}@nankai.edu.cn

ABSTRACT

Binarization of natural images is important for text location and content-based analysis. In this work, a new adaptive method is introduced. It is able to improve the binarization results on the degraded images, such as the complex background, the non-uniform illumination, the variations of text font, size, color, and line orientation. The presented method contains three main stages. Firstly, original threshold of each pixel is calculated to produce some candidate blocks. Secondly, the new inter- and intra-block features are extracted from the candidates based on the characteristics of text. Finally, each block is scored from 0 to 1 using the mentioned features. The blocks with low scores are considered as subcomponents of background. After extensive experiments, our method demonstrated superior performance against two well-known techniques on the ICDAR 2005 competition dataset.

Index Terms—Binarization, natural images, inter- and intra-block features

1. INTRODUCTION

Binarization is an important preprocessing step in many image-related applications. For the images which contain text regions, binarization may be defined as the process of separating the text pixels from the background [1].

In current work, most researchers focus on document image binarization. Sahoo et al. [2] compared more than 20 thresholding algorithms using uniformity or shape measures, and demonstrated the effectiveness of the methods proposed by Otsu [3] is yield adequate binarization result. Sezgin and Sankur [4] made a categorized survey of image thresholding methods. They selected a subset of 40 bilevel image thresholding methods, which had been implemented and for which the thresholding formulas had been expressed in a streamlined fashion. Quantitative evaluation scores had been obtained using a database of 40 document images.

The existing methods can be classified into two categories: global binarization and local binarization. In global methods, a single threshold is estimated and applied to the whole image. While in local methods, the threshold is determined for each pixel based on its own gray value and the gray value of its neighborhood. Hence, the local approaches have also been called adaptive binarization algorithms [5, 6]. Pratikakis and Gatos organized DIBCO 2011, which represent the state of the art for document image binarization [7].

Natural image binarization is receiving intensive attention in recent years, because it is important for many hot applications such as content-based image retrieval, robotic navigation in urban environments, natural understanding and virtual reality. Although document image binarization has been successfully applied to many practical systems, fast and accurate binarization for natural images is still a challenge due to the variations of text font, size, color and alignment orientation, and it is often affected by complex background, illumination changes, image distortion and degrading. Some researchers have reported promising results. Kita and Wakahara [8] used k-means clustering and support vector machines to select a single binarized image with maximum degree of “character-likeness” as an optimal binarization result. Mishra et al. [9] represented the pixels random variables in an MRF and minimize the energy function to find the optimal binarization. However, the two approaches are only effective on a “single-character” or “single-region” image. It is difficult to segment text accurately in a whole natural image since there are many non-text blocks which are easily confused with text when analyzed individually.

To overcome the above difficulty, we present a novel method using inter- and intra-block features for the binarization of natural images. We use an existing adaptive algorithm to form a threshold surface. Then the connected pixels are divided into candidate blocks. Four intra-block features and one inter-block feature are extracted and evaluation of each block is executed. After that, we classify the candidates into either foreground text or background.

The outline of the paper is as follows. We formulate our binarization process and discuss the chosen block features in Section 2, where the famous Sauvola’s method [10] used to generate candidate blocks is also introduced. Then results and conclusions are presented in Section 3 and 4, respectively.
2. BINARIZATION USING BLOCK FEATURES

The flowchart of our binarization method is shown in Fig.1. The input of the system is a gray-scale image, and we use Sauvola’s method to calculate an original threshold surface of the natural image and generate some candidate blocks (Section 2.1). Subsequently, four intra-block features, containing width-height, black pixels density, hollow pixels density and stroke width variance, are extracted from the blocks (Section 2.2). We also extract the array model feature, which is named inter-block feature, with prior knowledge of text position and scale (Section 2.3). As a next step, the candidate blocks are scored from 0 to 1 with the resulting features. Finally, we evaluate each block and classify it into either text or background at the output of our system (Section 2.4).

2.1. Generating Candidate Blocks

As the proposed approach involves the computation of thresholds using Sauvola’s method [10], we begin by giving a brief summary of that method. Given gray-scale image, Sauvola’s method first performed a rapid classification of the local contents of a page to background, foreground and text pixels near the threshold surface may be misclassified. Since the purpose of binarization is to separate text pixels from background, in our experiments, we set R=128 and k=0.5 to generate candidate blocks, where most of the foreground pixels are reserved. At this step, we obtain a rough estimation of foreground regions. Our intention is to proceed to an initial segmentation of foreground and background regions that will provide us a superset of the correct set of foreground pixels [6]. Then we extract inter- and intra-block features to find out the misclassified blocks in the candidates.

2.2. Extracting Intra-block Features

The candidate set generated with Sauvola’s method consist of both text and non-text blocks, between which some characteristics named intra-block features are different. In our experiments, we use four intra-block features, containing width-height, black pixels density, hollow pixels density and stroke width variance.

2.2.1. Width-height feature (WH)

The width-height feature is considered to be a basic shape feature used in character classification and recognition. The block B’s WH feature is defined by Eq. (2).

\[ f_1(B) = \text{WH}(B) = \begin{cases} 1, & \text{if WH}_{\text{Low}} < W(B) < WH_{\text{High}} \\ 0, & \text{otherwise} \end{cases} \]

where \( W(B) \) and \( H(B) \) are the width and height of \( B \). The parameters \( WH_{\text{Low}} \) and \( WH_{\text{High}} \) represent the left and right boundary respectively.

2.2.2. Black pixels density (BD)

The black pixels density is another important feature. We consider block as text when its BD feature is appropriate.

\[ f_2(B) = \text{BD}(B) = \begin{cases} 1, & \text{if BD}_{\text{Low}} < BPN(B) \times H(B) < BD_{\text{High}} \\ 0, & \text{otherwise} \end{cases} \]

where \( BPN(B) \) is the black pixels number of the block. The parameters \( BD_{\text{Low}} \) and \( BD_{\text{High}} \) represent the left and right boundary respectively.

2.2.3. Hollow pixels density (HD)

We define hollow as the white pixel areas which are surrounded by black pixels. Generally speaking, there are few hollows in text region and they are always central. If a block has a number of small hollows, it is likely to be a subcomponent of the background. Fig.2 shows the hollow pixels density feature of both text block and non-text block. The HD feature is defined by Eq. (4).
Fig. 2. Example of hollow pixels density. (a) hollows in text block. (b) hollows in non-text block.

2.2.4. Stroke width variance (SWV)

We define counter pixels of block $B$ as the ones having one or more white pixels in their 4-neighbor. As Fig. 3. shows, for each counter pixel, we search its opposite pixel on the block’s border. Then the distance between the two pixels, considered as the current stroke width, is calculated. When a counter pixel has more than one opposite pixel in different directions, we choose the minimum value of $d_i$ as stroke width.

Since a text block is likely to have more even stroke width than a non-text block, we calculate the stroke width variance feature to determine whether the block is a text component. If it is supposed that we have found $t$ pairs of pixels on the block’s border, Eq. (5) presents the SWV feature.

$$f_4 = SWV(B) = \begin{cases} 1, & \text{if } \frac{1}{t} \sum_{i=1}^{t} SW_i - \frac{1}{t} \sum_{j=1}^{t} SW_j < SWV_{\text{high}} \, \text{otherwise} \\ 0, & \text{otherwise} \end{cases}$$

where $SW_i$ is stroke width estimated with the $i^{th}$ pair of counter pixels and $SWV_{\text{high}}$ represents the SWV’s upper limit of a text block.

Fig. 3. Estimation of stroke width

2.3. Extracting Inter-block Features

In natural images, the text content always have a systematic arrangement, while non-text content (or noise) are listless. Since the spatial relationship of the candidate blocks is important for classification, we divide the arrangement into two categories, line model and curve model as Fig. 4 shows. Based on the categorization, regardless of block height, the algorithm of extracting array model feature (AM) is as follows.

Step1: Supposed that the candidate blocks set $B_c = \{B_1, B_2, \cdots, B_m\}$, we search a new blocks set $\tilde{B}_c$ for each $B_i \in B_c$. The search is within a circle whose center is set to the center of $B_i$ and radius $R=2 \times \text{HEIGHT}(B_i)$.

Step2: For each $\tilde{B}_i \in \tilde{B}_c$, we find out other related blocks which are arranged in the direction fixed by $\vec{d}_i$ and $\vec{d}_j$. Repeat this step until all blocks which fit well with the line model are found.

Step3: For each block $B_k$ not fitting line model, we search a new blocks set $\tilde{D}_k$. The search is within a circle whose center is set to the center of $B_k$ and radius $R=5 \times \text{HEIGHT}(B_k)$. For each $\tilde{B}_i \in \tilde{B}_c$, we find out other related blocks which are arranged at the angle fixed by $\vec{d}_i$ and $\vec{d}_j$. Repeat this step until all blocks which fit well with the curve model are found.

Step4: Supposed that the subset $B_s = \{B_1, B_2, \cdots, B_p\}$ fit either line model or curve model, we set the AM features of these blocks as Eq. (6).

$$f_5 = AM(B_1) = AM(B_2) = \cdots = AM(B_p) = p$$

2.4. Blocks Evaluation and Filtering

Since the mentioned features have been extracted from the candidate blocks, we score each block from 0 to $s$. Then the result whether a block is text or background can be calculated as Eq. (7).

$$\lambda_{B_i} = \begin{cases} 1, & \text{if } \sum_{j=1}^{5} f_j > \Theta_{\text{Text}} \\ 0, & \text{otherwise} \end{cases}$$

where $f_i$ is the $i^{th}$ feature and $\Theta_{\text{Text}}$ represents the boundary of the block $B$. If the score of $B$ reaches the boundary, we consider that this block belongs to text.
Fig. 5. Binarization result of an image from ICDAR 2005 dataset. (a) original image. (b) result with Sauvola’s method in [10]. (c) result with Multi-scale method in [12]. (d) result with proposed method.

Table 1. Binarization results

<table>
<thead>
<tr>
<th>Method</th>
<th>R(%)</th>
<th>P(%)</th>
<th>F-measure(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sauvola’s method [10]</td>
<td>48.32</td>
<td>40.71</td>
<td>44.19</td>
</tr>
<tr>
<td>Multi-scale method [12]</td>
<td>46.32</td>
<td>43.92</td>
<td>45.14</td>
</tr>
<tr>
<td>Proposed method</td>
<td>55.34</td>
<td>66.21</td>
<td>60.29</td>
</tr>
</tbody>
</table>

3. EXPERIMENTAL RESULTS

In this section, the performance of the proposed binarization method is evaluated against the ICDAR’05 dataset [11], which consists of 509 degraded natural images. These images were captured using digital camera in indoor and outdoor conditions. The measures used to compare proposed method and Sauvola’s method [10] and another famous multi-scale method [12] are the well-known recall (R), precision (P), and F-measure. Table 1 and Fig. 5 show that the proposed method gives better performance.

\[
R = \frac{TP}{TP + FN} \quad P = \frac{TP}{TP + FP} \quad F = \frac{\beta^2 \times R \times P}{\beta^2 \times (R + P)}
\]

where \(TP\), \(FP\), \(TN\), and \(FN\) denote the true positive, false positive, true negative, and false negative values, respectively. The parameter \(\beta\) is set to 1.

4. CONCLUSIONS

In this paper, we introduce a new approach to the binarization of natural images. Based on the characteristics of text region, proper techniques are used to extract inter- and intra-block features. The main advantage of the proposed method is that it prevents lots of confusing blocks while preserving most of the text regions. Experimental results on the ICDAR 2005 dataset show that the proposed method provides better performance than existing methods.

5. ACKNOWLEDGMENTS

This work is supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China under grant no.20100031120042, the Fundamental Research Funds for the Central Universities under grant no.65010201, 65012131 and the Tianjin Natural Science Foundation of China under grant no.12JCYBJC10100. The authors also would like to thank Jiaofeng Li and Mingda Li for their valuable feedback and suggestions.

6. REFERENCES


