Scene Character Reconstruction through Medial Axis

Shangxuan Tian, Palaiahnakote Shivakumara, Trung Quy Phan, and Chew Lim Tan
Department of Computer Science, School of Computing, National University of Singapore
Computing 1, 13 Computing Drive, Singapore 117417
Email: {tians, shiva, phanquyt, tancl}@comp.nus.edu.sg

Abstract—Character shape reconstruction for the scene character is challenging and interesting because scene character usually suffers from uneven illumination, complex background, perspective distortion. To address such ill conditions, we propose to utilize Histogram Gradient Division (HGD) and Reverse Gradient Orientation (RGO) to select Candidate Text Pixels (CTPs) for a given input character. Ring Radius Transform is applied on each pixel in a CTP image to obtain radius map where each pixel is assigned a value which is the radius to the nearest CTP. Candidate medial axis pixels are those having maximum radius values in their neighborhoods. We find such pixels on horizontal, vertical, principal diagonal and secondary diagonal directions to determine the respective medial axis pixels. The union of all medial axis pixels at each pixel location is considered as a candidate medial axis pixel of the character. Then color difference and $k$-means clustering are employed to eliminate false candidate medial axis. The potential medial axis values are used to reconstruct the shape of the character. The method is tested on 1025 characters of complex foreground and background from ICDAR 2003 dataset in terms of shape reconstruction and recognition rate. Experimental results demonstrate the effectiveness of our proposed method for complex foreground and background characters in terms of character recognition rate and reconstruction error.

I. INTRODUCTION

Text recognition in natural scene images has received increasing attention in recent years due to its effectiveness in scene understanding. It has found applications in automatic vehicle navigation, assistance system for the visually impaired and image tagging and retrieval. An experimental study in [1] shows that human beings tend to be attracted by text when they are presented with images containing text and other objects. This study further discusses the importance of text recognition. Therefore, capturing the text in images is a very important and worthwhile task.

However, because of the intrinsic characteristics of natural scenes, classical Optical Character Recognition (OCR) systems have become ineffective in dealing with scene texts directly. Scene characters are usually of complex foreground and background in terms of color, texture and uneven illumination. Such conditions are detrimental to OCR engines resulting in really low accuracy in terms of character recognition. Fig. 1 shows some samples for which OCR does not recognize correctly due to complex background and foreground while our proposed method recognizes correctly. Therefore, achieving good recognition rate for scene characters using publicly available OCR is very challenging and still an elusive goal for researchers.

The main reason for poor recognition rate for scene and video texts with OCR is the loss of text information through binarization using the current binarization methods. In view of this, Shivakumara et al. [3] have proposed the ring radius transform to reconstruct shapes of characters to improve character recognition rate for video and scene characters. However, the method works well for filling horizontal and vertical gaps but not other kinds of gaps. In addition, shape may not be preserved when there is a big gap in the character’s edge image. Later in [4], iterative-midpoint-method is proposed to fill gaps and preserve shapes of the characters without assuming the direction of gaps. However, this method is good when the character’s contour has limited small gaps. Thus it fails for characters having multiple and large gaps. Thus poor recognition rate for scene characters is reported in the paper. The main focus of these methods [3], [4] is to reconstruct shape of video characters but not scene characters. In addition, the robustness of the construction algorithms depends on Sobel and Canny edge maps of the input images. These factors motivated us to propose a new character reconstruction method through medial axis which can take care of multiple gaps, large gaps and character shapes to achieve better recognition rate for scene characters without depending on edge maps of characters.

Text binarization is a familiar topic for document analysis community. The methods [5], [6], [7], [8] are mainly based on thresholding techniques. These techniques achieve excellent performance in document image binarization because the foreground and background of the document are generally monotonous. However, when such techniques are applied to scene texts, the performance is far from satisfactory because of the variety of colors, different textures and varying illumination. To deal with problems of thresholding techniques over scene character images, several binarization approaches [9], [10], [11] have been proposed. A variant of Niblack’s
binarization algorithm [7] is employed in [9] which chooses the local window size adaptively to binarize detected scene texts. In [10], the authors propose a binarization method on single character using $k$-means clustering in terms of color information to get $2^k - 2$ binarized characters. And the one with maximum character-likeness measured by a Support Vector Machine is taken as the result. The Markov Random Field (MRF) model is adopted for binarization in [11] where an auto-seeding technique is proposed to first select foreground and background pixel seeds and then MRF is used to segment text and non-text regions. Since most of the previous methods are based on plain background, it is hard to expect those methods to work on scene character images having cluttered background. Hence, in this paper, we propose a new method to deal with this problem using candidate text pixels and medial axis. Experimental results show our approach outperforms state-of-the-art methods under complicated conditions with various characters.

II. CHARACTER SHAPE RECONSTRUCTION

Inspired by the work proposed in [12] where it is shown that Histogram Gradient Division (HGD) on local gradient is useful for obtaining dominant text pixels without losing much text pixels for video text with complex background, we extend the division operation to diagonal direction to obtain dominant pixels for scene characters in this work. In addition, we propose Reverse Gradient Orientation (RGO) to study the local gradient information to obtain dominant text pixels. Then union operation is applied on dominant pixels by HGD and RGO which outputs Candidate Text Pixels (CTPs) that may scattered without connecting to each other. Ring Radius Transform (RRT) [3] is applied on CTP image to obtain a radius map. Then probable Medial Axes (MA) are extracted based on the radius map and false branches are eliminated via color information. Iterative-midpoint-method (IMM) [4] is applied to connect small gaps of the medial axis. The resultant medial axis has preserved the character’s shape and we propose a novel way to reconstruct a solid binary character directly from the medial axis. The pipeline of the proposed method is shown in Fig. 2. The method consists of three subsections which deal with candidate text pixels selection, medial axis extraction and character shape reconstruction, respectively.

A. Candidate Text Pixels Selection

Scene texts appear to have consistent foreground which usually does not vary greatly in terms of color and texture such that humans can locate and recognize them. In contrast, no constrains are given to background, which means background can appear in any way. Given this assumption, we propose gradient division and gradient orientation to obtain two kinds of dominant text pixels of the input character and union operation is applied on them to get the candidate text pixels. The candidate text pixels are supposed to lie densely on the contours of the character and may scatter within the background region when it is very cluttered.

1) Dominant Text Pixels using Histogram Gradient Division: Gradient information is extremely useful in character shape reconstruction because it gives lower gradient values in the text region and higher values along the contours and cluttered background of the characters. Based on this fact, we propose to use Histogram Gradient Division (HGD) to find the dominant text pixels of the characters.

Given an input image, we first compute different kinds of gradient as shown in Fig. 3. The horizontal gradient is computed via $[-1,0,1]$ while a transpose is used to get the vertical gradient. Principal diagonal gradient is computed via $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ while $\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ is used to get the secondary diagonal gradient. After obtaining the gradients, the character is divided as illustrated in Fig. 3 to get 8 subregions. Then, we use $k$-means to get two clusters according to the gradient magnitude in each subregion. Between the two clusters, the pixels belonging to the cluster that has the larger mean magnitude are kept, while the rest are discarded because we expect the candidate text pixels to have larger gradient values. Then a gradient magnitude histogram of the retained pixels is computed and we take the pixels whose gradient values correspond to the peak in the histogram as the dominant pixels. Fig. 4(b) shows the dominant pixels obtained using gradient division of the input character in Fig. 4(a).

2) Dominant Text Pixels using Reverse Gradient Orientation: As studied in [13], each edge pixel lying on one side of the stroke of a scene character is likely to have a corresponding pixel with opposite gradient orientation lying on the other side of the same stroke. Inspired by this assumption, we propose Reverse Gradient Orientation (RGO) to capture those pixel pairs that have opposite gradient orientation as well as large gradient magnitudes. First, gradient orientation is computed at each pixel and quantized into integer orientation bins ranging from $-180^\circ$ to $+180^\circ$ based on the horizontal and vertical gradient computed in the above HGD section. Note that here we use the same orientation map in the eight subregions whereas in the gradient domain we compute gradient differently when we divide the image into different subregions. Then a gradient map is obtained using $L2$ norm with respect to the horizontal and vertical gradient magnitudes.

As to each of the eight subregions, we take the following steps. We first use $k$-means to get two clusters in terms of gradient values. The one whose center’s value is larger is noted as the high-mean cluster. For each of the pixels in each orientation bin (say bin $x$), we check whether there exists a pixel in the opposite orientation bin (i.e. bin $-x$). If yes, we further check whether both of them belong to the high-mean cluster we get previously. Mark this pair of pixels as dominant pixels if the above conditions are satisfied. The rationale for this step is that we expect those candidate text pixel pairs to have opposite gradient orientations and large

![Fig. 2: Pipeline of the proposed method.](image-url)
The reconstructed character is already a binary character with a jagged boundary. In order to refine the character shape, we use k-means clustering with respect to color feature of those pixels in the original image that correspond to text in the binary image. The cluster that has the larger number of pixels is kept. In this way, we can eliminate some background pixels obtained previously. Till this step, we get an initial medial axis as in Fig. 5(e). The initial medial axis may have some false branches that appear in the background. We eliminate those false branches by using the color difference with the text regions via k-means cluster. After this step, there are still some noisy medial axes which, in most cases, have the radius values that do not appear frequently. Therefore, we find the dominant radius with the highest frequency and keep those radii that are within a small range from the dominant radius and discard those isolated small medial axes. So far, we successfully obtain the medial axis we want as in Fig. 5(f).

The medial axis we get may not be connected due to varying illumination, fancy fonts and so on. We perfect it using iterative-midpoint-method [4] to fill the gaps to get Fig. 5(g).
Proposed IMM Sobel 0.078 0.154 0.178 0.188

Canny and Sobel edge detections are conducted on the original image obtained and the ground truth in [15]. At the same time, we perform Canny edge detection on the binary character image to recover the shape of the input characters in terms of edge map.

A. Qualitative Measure for Shape Reconstruction

Fig. 6: Shape reconstruction and comparison: (a) Original input character. (b) Otsu’s binarization method. (c) Shape reconstruction by our proposed method. The second row is obtained using Canny edge detection on characters in the first row, respectively.

Fig. 7: Another example of our proposed method and existing ones. Characters in the first row from left to right are the original input, Otsu’s binarization, our proposed method, respectively. The characters in the second row is the corresponding Canny edge of the first row.

along the boundary and make the character smoother. Canny detector is then employed to get the shape of the character.

III. EXPERIMENTS

The evaluation of our proposed methods is performed on 1025 character images from ICDAR 2003 character testing dataset [14]. Those scene characters are chosen because they vary greatly in terms of fonts, size, color and so on. Meanwhile, many of them suffer from uneven illumination and perspective distortion. Even worse, both foreground and background text regions are very complex due to different colors and textures, especially the background. Some examples are shown in Fig. 1. In a recent work [15], the authors annotated pixel-level binarization ground truth for ICDAR 2003 dataset for evaluation. We compare our methods with previous works based on the ground truth. The evaluations are on two aspects, shape reconstruction error and character recognition accuracy.

A. Qualitative Measure for Shape Reconstruction

We expect that our shape reconstruction results can well recover the shape of the input characters in terms of edge map. So we perform Canny edge detection on the binary character image we obtained and the ground truth in [15]. At the same time, Canny and Sobel edge detections are conducted on the original input character for comparison.

To evaluate our shape reconstruction results, we use relative error to measure the quality. We consider two features in the measurement: (1) number of pixels in the edges. (2) area of convex hull involving all the edges. These two features even though cannot give a very accurate measurement, they can effectively reflect the reconstruction error based on edge map. The following formula is used to measure the relative error:

$$E = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} |Gr_{ij} - Test_{ij}| \quad (1)$$

where $n$ is the total number of images, $m$ is the number of features used, i.e., number of pixels in the edges and area of convex hull involving all the edges. $Gr_{ij}$ represents the ground truth feature and $Test_{ij}$ represents the testing feature to compare. Table I shows the relative errors of existing methods and our proposed method.

From the table we can see that among the three existing methods, Canny detector produces the highest error since it gives many noisy edges while the Iterative-Midpoint-Method (IMM) [4] achieves the lowest error since it removes some false edges and fills small gaps between disconnected edges. However, even IMM incurs more than twice the error of our proposed method which eliminates almost all the false edges in the background while maintaining the text boundary quite well.

B. Quantitative Measure for Shape Reconstruction

Character recognition accuracy is an effective measurement of character shape reconstruction. We compare our method with Otsu’s [5] and Niblack’s [7] binarization methods which are widely used in document image analysis area. The recently proposed IMM method [4] also involves character shape reconstruction and performs quite well on scene characters.

Tesseract OCR [2] is employed for shape reconstruction evaluation and the results are shown in Table II. The result in the second column of Table II is the accuracy of Tesseract OCR on raw color images and the third column are results after applying the corresponding methods in the first column. As shown in the second column, if we just apply OCR directly to the characters without any preprocessing, only 20.78% recognition accuracy is achievable. We next apply OCR to the pixel-level binarization ground truth [15], and obtain 65.17% recognition accuracy. This is the best achievable as this is an idealized situation when the OCR is given the ground truth. We will now test our method together with the other methods mentioned above to see how close we can achieve with respect to the best result of 65.17%.

As shown in Table II, our method achieves the closest result of 62.15%. Among the binarization methods, Niblack’s method gives the lowest accuracy since the local window size is fixed in the method while the scene character size varies significantly (character height ranges from 21 to 731 pixels and width ranges from 11 to 444 pixels).

In Table III, we demonstrate some samples using Otsu’s and Niblack’s binarization methods in comparison with our method.
TABLE II: Recognition accuracy on ICDAR 2003 test dataset (1025 characters)

<table>
<thead>
<tr>
<th>Method</th>
<th>Before</th>
<th>After</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground Truth (idealized)</td>
<td>20.78%</td>
<td>65.17%</td>
<td>44.39%</td>
</tr>
<tr>
<td>Proposed</td>
<td>41.37%</td>
<td>62.15%</td>
<td>20.78%</td>
</tr>
<tr>
<td>IMM</td>
<td>45.37%</td>
<td>48.20%</td>
<td>2.83%</td>
</tr>
<tr>
<td>Otsu</td>
<td>24.59%</td>
<td>27.42%</td>
<td>2.83%</td>
</tr>
<tr>
<td>Niblack</td>
<td>25.27%</td>
<td>24.59%</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

TABLE III: Sample shape reconstruction results of ICDAR 2003

<table>
<thead>
<tr>
<th>Input</th>
<th>Otsu</th>
<th>Niblack</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>'g'</td>
<td>null</td>
<td>null</td>
<td>'g'</td>
</tr>
<tr>
<td>'k'</td>
<td>null</td>
<td>null</td>
<td>'k'</td>
</tr>
<tr>
<td>'N'</td>
<td>null</td>
<td>null</td>
<td>'O'</td>
</tr>
<tr>
<td>'y'</td>
<td>'N'</td>
<td>'R'</td>
<td>'N'</td>
</tr>
<tr>
<td>'R'</td>
<td>null</td>
<td>null</td>
<td>'O'</td>
</tr>
<tr>
<td>null</td>
<td>null</td>
<td>null</td>
<td>'S'</td>
</tr>
<tr>
<td>'D'</td>
<td>null</td>
<td>null</td>
<td>'D'</td>
</tr>
<tr>
<td>'E'</td>
<td>'E'</td>
<td>'K'</td>
<td>'F'</td>
</tr>
</tbody>
</table>

proposed method to give a direct view of the improvements. Those listed characters are exposed to different ill conditions such as complex foreground and background, fancy fonts and varying illumination. The results are evaluated via Tesseract OCR and shown below each character. As we can see, our method can recognize most of the characters while the other methods fail completely.

IV. CONCLUSION AND FUTURE WORK

In this paper we propose a character shape reconstruction method based on medial axis. We extend the dominant text pixel selection method in [12] to principal diagonal and secondary diagonal directions to obtain more dominant text pixels. In addition, a new Reverse Gradient Orientation method is proposed to get additional dominant text pixels. Then medial axis is extracted after applying ring radius transform on candidate text pixels. A novel method is presented to reconstruct the character by setting all the pixels within the radius indicated by the medial axis values as text. In this way, we can directly get a solid binary character. Experiments show that our proposed method greatly outperforms existing binarization methods in terms of shape reconstruction error and character recognition rate on ICDAR 2003 scene character dataset.

Currently parts of the boundary of the reconstructed binary character are not very smooth and the noise in the medial axis may affect the final reconstruction result. Therefore, our future work will deal with such problems to obtain a much smoother and more accurate binary character.

ACKNOWLEDGEMENT

This research is supported in part by A*STAR grant 092 101 0051.

REFERENCES