General Document Image Correction Method
Based on Text Line in Natural Scene

Yuze Xiang, Jiani Gao, Haijiao Shan and Xiaokun He

ABSTRACT

This paper presents a general document image correction method based on the text line. The proposed method roughly screens the text area and implements rotation correction based on the TextSnake text detection model. Then by the relative positions of characters, the principle of text direction consistency and the two-way matching strategy, the precise positioning is realized. At last, the information of text ridges is obtained to acquire the nonlinear mapping relationship of the whole graph, Therefore achieving correction reconstruction of the general document image. The experiment results show that the correction method can overall solve image quality problems such as tilt distortion and improve the recognition rate of tilt distortion image.

KEYWORDS
Component, Formatting, Style, Styling, Insert (key words).

INTRODUCTION

Optical character recognition (OCR) of document materials has always been an important issue in image processing and computer vision. With the increasing popularity of mobile camera devices, more and more mobile camera devices have replaced traditional scanning devices. Digital devices have the advantages of portability, fast response, and non-contact. However, there are geometric distortions (affine changes, perspective transformations) and geometric distortions caused by curved surfaces in the images captured by the camera. In order to reduce geometric distortion, many methods have been proposed in the industry. At present, there are mainly three solutions to this problem: (1) scanning the 3D structure of the document by laser scanner[1-2], which can accurately estimate the structural shape of the text. But the method is constrained by hardware devices and is currently used less. (2) Document geometric correction is realized by the texture features in the image[7], which requires that the document content in the image be relatively neat. In the case of a large number of non-text object interference factors in the complex scene, the texture features are unstable, and the method is greatly limited. (3) Based on the line correction image[6,11], the image correction model parameters are calculated by the feature attributes of the text line. However, this method relies heavily on the accuracy of text line detection and has poor robustness.

Considering the above limitations of the current methods, and the method
based on text lines has been the most widely concerned because of its good performance. Based on TextSnake text localization, this paper expands the image detail feature processing and designs a line detection algorithm with high robustness to the warped text lines, which realizes local correction and overall reconstruction of linear and non-linear geometric distortion of the captured image. This method has better stability and robustness than traditional methods.

**SYSTEM DESIGN**

Due to the interference of perspective distortion, geometric distortion and non-text objects (such as messy background, etc.), text detection is prone to the phenomenon of multiple text lines sticking together, leading to the wrong detection of text lines. In this paper, the pre-trained TextSnake model is firstly used to detect the target text area[8], and the model outputs the target area to detect the energy diagram. The OTSU method is used to binarize the energy diagram, therefore to eliminate the interference from non-text areas. The following four steps have been used in addition to the above methods to complete the correction of distorted documents: (1) binarization; (2) global rotation correction; (3) text line detection based on character concatenation; (4) fitting the baseline of text lines to achieve single-line text distortion correction; (5) layout restoration completes the overall reconstruction. As shown in the flow chart Figure 1.

**BINARY IMAGE**

Image binarization is an important step in image processing, especially in text line detection. The processing effect directly influences the accuracy of subsequent steps. At present, there are many methods for image binarization, including the traditional OTSU method, the Sauvola method, and the Niblack method. By comparing the effects of each binarization method, this paper adopts adaptive binarization, which can effectively suppress the effect of shadow on the binarization effect. The basic formula is as follows.

$$ T(x,y) = m(x,y) + k \ast s(x,y) $$

The threshold value of the image $I(x,y)$ at the point $(x,y)$ is determined by the local mean $m(x,y)$ and local standard deviation $s(x,y)$. $k$ is adjustment coefficient of the expression, which is usually set as -0.2.
ROTATION CORRECTION

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

The bounding rectangle is an important geometric feature of the image. Geometric features are of great importance in image processing and analysis. This paper uses the minimum bounding rectangle of the text line to calculate the rotation correction angle of the text. From the above, here this paper obtain the distributed energy map of the text line in the entire document by the pre-trained TextSnake model. The energy map is binarized by a fixed threshold binarization method to obtain a mask map of the text region in the original image. The connected domain is detected in and the minimum circumscribed rectangle of each connected domain is calculated. Let each connected domain have a width
$W$, a height $H$, and connected domain with $H/W < 4$ and $H < T_H$ is eliminated, where $T_H$ is the fixed text box height threshold. In this step, the misdetected non-text area that may exist is removed. Calculate the average value $\theta$ of the minimum circumscribed rectangle rotation angle of all connected domains as the image rotation correction angle, and the rotation transformation is as shown in formula (2)

$$
\begin{bmatrix}
    x \\
    y
\end{bmatrix} =
\begin{bmatrix}
    \cos \theta & -\sin \theta \\
    \sin \theta & \cos \theta
\end{bmatrix}
\begin{bmatrix}
    x_0 \\
    y_0
\end{bmatrix}
$$

(2)

Where $(x_0, y_0)$ is the coordinate of original image $I(x, y)$, $(x, y)$ is the pixel point coordinates on the rotated image $I_d(x, y)$.

Figure 3. Image rotation correction result.

**TEXT LINE DETECTION**

The lines of text in the captured picture is not completely horizontal, so the algorithm of text line based on horizontal projection is not applicable. This paper designs a method based on the character connection to implement text line detection. The method is in the form of a character block, which is a small-range character area connected by several characters, compared to the traditional one, making the angle information of the text introduced and the text further improved by the principle of distance closest principle and angle consistency. The accuracy of the block connection and the introduction of a two-way matching strategy further improve the robustness of the algorithm. The specific algorithm steps are as follows:

1) Let the binarized image after the text target area selection and rotation correction used for morphological expansion. The connected domain is detected for the expanded image. Calculate the circumflex moment and the minimum circumflex moment of each connected domain. The test results are shown in Figure 4.

2) For any text blocks $C_i$, global search text blocks $C_j, (j \neq i)$, and the conditions meet:

$$
|C_i^\theta - C_j^\theta| \leq T_\theta
$$

(3)

$C_j$ is called the candidate neighboring text block of $C_i$ (where $T_\theta$ is the set angle threshold, which represents the maximum angular difference allowed
between the two text blocks). Iterate through all the text blocks and search for all candidate neighboring characters. Then find the nearest text block from the selected text block in the candidate neighboring character set on the left and right sides, and empty if the text block is not satisfied. The distance between text blocks is defined as the Euclidean distance between the center points;

Figure 4. Text block detection result.

(3) Apply two-way matching of the connection to the text block, and determine the mutual connection relationship between the two text blocks only when the two text blocks are mutually adjacent to each other;

(4) Repeat steps 2 and 3 until the nearest neighbor characters of all text blocks are found;

(5) Combine all adjacent character sets with overlapping parts, form a character set of single text lines, and construct a text line rectangle according to the circumscribed rectangle between the characters. The text line detection result is shown in figure. 5.

Figure 5. Text line test result.

TEXT LINE DISTORTION CORRECTION

In the text line correction, this paper uses the least-squares method to obtain the text ridge and uses the moving least-squares method to fit the text deformation and achieve the distortion text correction. The specific algorithm is as follows:

(1) Let the original image after the rotation correction represented. Extract the single text line graph respectively based on the text line detection result.

(2) Perform character connection and text area extraction through morphological expansion operation, and connect into a complete text line outline to represent the overall shape direction of the text line; on the outline of the text line, find the coordinate point position of the center point of the contour Data point set S. At this time, the upper left corner of the image is taken as the
coordinate origin, and the direction in which the image width is located in the x-direction, and the direction in which the height is located in the y-direction. At this time, the coordinates of each pixel in the set S are as shown in the formula (4);

\[ S_i = \left[ x_i, \frac{y_{i1}+y_{i2}}{2} \right], i \in [0, w) \] (4)

(3) Image Deformation Line Matrix M (Equation 4), the corresponding position \( f(v) \) in the deformed image can be obtained by the deformation matrix M, the coordinates of the pixel point \( v \) on the original image.

\[ f(v) = (x - p')M + q' \] (5)

where \( x \) is the coordinates of the pixel points of the original image, \( p' \) and \( q' \) are the weighted average positions of the coordinates of the respective control vertices before and after the deformation, and the weighted average of the coordinates of the preset control vertices before the \( p' \) deformation, \( q' \) is the weighted average of the coordinates of the preset control vertices after deformation, and exists in pairs; \( (x - p') \) is the distance from each pixel of the original image to the weighted average position, and the deformation matrix formula is

\[ M = \left( \sum_j P_j^T w_j P_j \right)^{-1} \sum_j P_j^T Q_j \] (6)

\[ P_i = p_i - p', Q_i = q_i - q \] (7)

where \( P \) is a matrix of coordinates of each control vertex before deformation, and \( Q \) is a matrix formed by the coordinates of each transformed vertex after deformation, which is a reciprocal of the distance from the pixel point to the control vertex on the original image;

![Image](image1)

(a) Before correction

![Image](image2)

(b) After correction

Figure 6. Single text line distortion correction result.

![Image](image3)

(a) Pre-correction image (b) Corrected image

Figure 7. Image correction result.
In this paper, two sets of experiments are conducted on the test set. The test set is a collection of conventional typesetting mathematics and English test papers and a picture taken by a general student practice in a natural scene. It is divided into normality, perspective distortion, distortion, blur, and other different types of pictures, a total of 200.

The first set of experiments was to correct the test set image and calculate the correction accuracy. The results show that more than 86% of the document images are corrected and reconstructed, and 14% of the images do not trigger the correction reconstruction logic due to the small distortion.

The second set of experiments is to use the Baidu open platform OCR recognition model to carry out the recognition rate comparison experiment to verify the impact of the corrected algorithm on the overall OCR recognition rate. Table I and Table II show the comparison results of the recognition rate before and after the correction of the entire document and the correction of the distortion text line respectively. The experimental results surface can effectively improve the accuracy of document recognition.

![Figure 8. Pre-correction image.](image1)

![Figure 9. Corrected image.](image2)

The recognition result of Figure 8 is: Wasemaegevery he ptuland positive aboutfe?
The recognition result of Figure 9 is: Was Meaanage very hepfun and positive about lit?

| TABLE I. CORRECTED THE ENTIRE DOCUMENT RECOGNITION SUCCESS RATE. |
|-----------------|-----------------|-----------------|-----------------|
|                 | Number of words | Number of correct word | Correct recognition rate | Error recognition rate |
| Before correction | 350             | 314              | 89.7%               | 10.3%               |
| After correction | 350             | 322              | 92.0%               | 8.0%                |
|                                | Number of words | Number of correct word | Correct recognition rate | Error recognition rate |
|--------------------------------|-----------------|------------------------|--------------------------|------------------------|
| Before correction              | 231             | 186                    | 80.5%                    | 19.5%                  |
| After correction               | 231             | 202                    | 87.4%                    | 12.6%                  |

CONCLUSION

This paper presents a new general document image correction reconstructed algorithm. Based on TextSnake's text energy information, the algorithm introduces the angle consistency principle and the two-way matching strategy on the text line to improve the robustness of the warped text line detection. The distortion text correction technology based on the text ridge deformation is designed to realize the overall correction and reconstruction of the document image. The experimental data shows that the method has a better correction effect on the text distortion of the document image in the natural scene. At the same time, the corrected OCR recognition effect is significantly improved and is suitable for application in a text recognition scene.

REFERENCES

1. He Y, Pan P, Xie S, et al. A Book Dewarping System by Boundary-Based 3D Surface Reconstruction//Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013.
2. Meng G, Pan C, Xiang S, et al. Metric Rectification of Curved Document Images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4):707-722.
3. Zeng Fan-feng, Guo Zheng-dong, Wang Zhan-dong. Fast correcting method based on connected components for warped Chinese document images[J]. computer engineering and design, 2015(5):1251-1255.
4. H.I. Koo, N.I. Cho, State estimation in a document image and its application in text block identification and text line extraction, in: Proceedings of the 11th European Conference on Computer Vision, 2010, pp. 421–434.
5. Kumar J, Kang L, Doermann D, et al. Segmentation of Handwritten Textlines in Presence of Touching Components[C]//2011 International Conference on Document Analysis and Recognition. IEEE, 2011.
6. Kumar J , Abd-Almageed W, Kang L, et al. [ACM Press the 8th IAPR International Workshop - Boston, Massachusetts (2010.06.09-2010.06.11)] Proceedings of the 8th IAPR International Workshop on Document Analysis Systems - DAS’10 - Handwritten Arabic text line segmentation using affinity propagation[C]// Iapr International Workshop on Document Analysis Systems. ACM, 2010:135-142.
7. Zhang Z, Liang X, Ma Y. Unwrapping low-rank textures on generalized cylindrical surfaces[C]//IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011. IEEE, 2011.
8. Long S, Ruan J, Zhang W, et al. TextSnake: A Flexible Representation for Detecting Text of Arbitrary Shapes[J]. 2018.
9. Wu M, Li R, Fu B, et al. A Model Based Book Dewarping Method to Handle 2D Images Captured by a Digital Camera[C]//International Conference on Document Analysis & Recognition. IEEE, 2007.
10. Wang L, Sun J, Uchida S. Text Line Extraction Based on Integrated K-Shortest Paths Optimization[C]//2018 13th IAPR International Workshop on Document Analysis Systems (DAS). IEEE, 2018.
11. Lbachir I A, Es-Salhi R, Daoudi I, et al. A Survey on Segmentation Techniques of Mammogram Images[M]//Advances in Ubiquitous Networking 2. Springer Singapore, 2017.
12. Schaefer S, Mcphail T, Warren J. Image Deformation Using Moving Least Squares[J]. ACM Transactions on Graphics, 2006, 25(3):533-540.