INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Rectification of Curved Scene Text Based on B-Spline Curve Fitting

Kiptanui Linus[1], **C J Prabhakar**[1], **S R Shrinivasa**[1]*

1 Department of Computer Science, Kuvempu University, Shankaraghatta, 577451, Karnataka, India

OPEN ACCESS

Received: 21-09-2023

Accepted: 22-11-2023

Published: 12-08-2024

Citation: Linus K, Prabhakar CJ, Shrinivasa SR (2024) Rectification of Curved Scene Text Based on B-Spline Curve Fitting. Indian Journal of Science and Technology 17(32): 3305-3317. https://doi.org/10.17485/IJST/v17i32.2402

* **Corresponding author**.

srinivasa1991sri@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: In this study, we proposed suitable technique for rectification of curved scene text which is followed by recognition of rectified text in order to improve the accuracy of the existing techniques. **Methods:** In order to rectify curved text, initially, we perform curved text detection using Look More Than Twice (LOMT) model which detects and locates curved text. The detected text area is binarized through adaptive binarizaton technique. Then, we rectify the detected curved text through B-spline based curve fitting which align the curved text into straight line. The rectified text is feed to our recognition module where, we segment the rectified text using Low Variation Extremal Regions (ER) technique and extract the Pyramid Histogram of Oriented Gradients (PHOG) features from the segmented texts. Finally, we perform recognition of text using Tesseract OCR. The rectification and recognition performance of the proposed method on CUTE80 dataset was evaluated using Mean Square Error (MSE) metric and word level recognition accuracy respectively. We compared the recognition results of the proposed method with state-of-the-art methods using challenging benchmarks such as ICDAR2013, Street View Text (SVT), IIIT5k-Words (IIITK), ICDAR2015, SVT-Perspective (SVTP) and CUTE80 dataset based on word level recognition accuracy. **Findings:** From the experimental results, it is observed that the proposed rectification method removes highest error (MSE is minimum) of 99.34% for perspective text with large angle. Also, the proposed recognition module achieves highest recognition accuracy of 93.80% compared to the state-of-the-art methods on the selected six benchmarks. **Novelty:** We proposed a technique for rectifying curved scene text using B-spline curve-fitting technique followed by recognition using handcrafted features. The novelty of the proposed method is that we employ B-spline curve fitting in order to compute local transformation of individual character followed by rectification of each individual character through the computation of normal vectors. The recognition of rectified text is done through the handcrafted features.

Keywords: Rectification; Curved Scene Text; Text Detection; Curve Fitting; Recognition; BSpline

1 Introduction

In this new era of Internet, digital mobile devices are equipped with advanced camera with additional software applications helps mankind to utilize the captured image/video for real time applications. Compared to traditional Optical Character Recognition (OCR) which is efficient and limited performance only with single font style, simple background, and uniform illumination, text recognition in natural scene images is more challenging. Text recognition in natural scenes has a wide range of applications in image understanding, product search, real-time translation, and other scenarios. The majority of scene text embedded in billboards, T-shirts, signboards and building walls of realworld environment is regular in nature. Recognition of regular text which is horizontal and frontal is achieved with high accuracy even though with many challenges posed by this problem. In some situations, text occurred in the natural scene may have irregular shape and some scene text is perspective text, which is caused by side-view camera angles. Some scene text has curved shapes due to placing of characters along curves rather than straight lines. We call such text as irregular text, in contrast to regular text. The performance of the existing scene text recognition methods are very poor on curved texts or irregular texts. This is due to the fact that the existing methods can only deal with horizontal or near to horizontal texts /characters. Curved scene text recognition is uncommon and challenging task because of its complex nature in terms of orientations of the characters in the text and complexity in the background. Therefore, it is very essential to develop a technique that is capable to recognize curved texts available in scene images in order to utilize its rich semantic information.

Huge amount of literature are available for recognition of curved scene text in images. However, the issue of curved scene text recognition is relatively less addressed and therefore, it is necessary to look into this issue to come up with improved method with high accuracy. The deep learning methods have significantly improved the performance of the curved scene text recognition. The existing deep learning based methods attempted to fill the research gap in terms of efficiency for real-world applications. Recently, researchers have proposed end-to-end hybrid methods which are deep learning in nature in order to improve the performance of arbitrarily-shaped scene text recognition. Given an input arbitrary text image with complex background, an end-to-end system aims to directly convert all text regions into string sequences through sequence of steps such as text detection, text recognition, and post-processing. In an end-to-end system, text detection branch and recognition branch can share information and can be jointly optimized to improve overall performance.

The existing end-to-end methods for curved text recognition are classified into two categories such as rectification-based methods and 2D-decoder-based methods. The rectification-based approaches first transform the curved text into horizontal text, and then recognize the transformed text. The rectification process before recognition can improve text recognition accuracy by rectifying text in irregular shapes into regular forms. From the literature review, we found that rectification of curved text can be carried out either using conventional methods or using deep learning methods. The most popular deep learning based text rectifiers are combination of Thin plate Spline (TPS) and spatial transformer networks (STN). In these methods, an input curved scene text image is spatially transformed into a rectified image that contains aligned or straightened text, which is fed into recognition module. Shi B et al[1] proposed rectification method using TPS transformation with only word-level annotations. The method uses regression to regress control points of image borders towards text borders. Then, TPS transformation parameters are calculated based on the movement of these points and the image is rectified using these TPS parameters. The method achieves promising results but it is less effective and not achieved satisfactory rectification result for more challenging text. This is due to the fact that it computes the TPS transformation parameters based on weakly supervised manner and also due to regression of text borders, the text is not well localized with very little text information. Therefore, it leads to character deformation. The major drawback of this method is that it requires an end-to-end training of the model which is expensive and time consuming.

Wang W et al[2] have proposed TPS based representation for arbitrary shape text using reverse process of rectification. The TPS parameters are computed using control points of original shape of text region, which helps to transform source shape into the target rectangular shape. In the reverse process of rectification, the coordinates of the target rectangle are transformed to the coordinate on the source arbitrary shape. The predicted TPS parameters are used to rectify arbitrary oriented text to a near-horizontal one which helps the subsequent recognition. The major drawback of this method is that, the rectification is not accurate and at the same time, the recognition fails when the curve angles are too large. The method suffers from high training time and cannot rectify small targets. Zheng T et al[3] have proposed TPS based rectification method for oriented text. The proposed method is called as TPS++, which is an extension of TPS transformation that consists of the attention mechanism for text rectification. TPS++ estimates the parameters through foreground control point regression and content-based attention score estimation. TPS++ provides rectification process which is content-aware rectifier corrects a natural text that helps and make easier for subsequent recognition process. The drawback of the method is that it fails to get a satisfactory rectification. The main reason for this is that it uses regressed text borders to estimate the TPS transformation parameters and employs weakly supervised learning and not well localized.

Wang J et al[4] have proposed the detection and recognition method for curved text. The method estimates TPS transform fiducial points using CNN based positioning network. The original curved text is fed to the positioning network where reference point is marked and TPS transformation parameters helps to obtain the corrected text image. If the obtained result is not satisfactory, repeat the above process for number of iterations. In each iteration, the previously computed TPS transformation is added to the original input image and the corrected image is used to predict control points. The horizontal corrected image is feed to the sequence recognition network. The major drawback of this system is that it requires huge amounts of scene text images for training and consumes much more computation resources, which limits its application in many situations. Heng H et al[5] have proposed scene text recognition module, particularly, for curved scene text embedded in low quality natural scene images. The proposed approach consists of a rectification module, convolutional neural network (CNN) module, a semantic context module (SCM), a global context module (GCM), and a light weight transformer decoder. The rectification module consists of three parts. A localization network is used to detect the control points. The grid generator calculates the coordinate position of points. The sample generator utilizes computed point positions to generate the rectified image. After rectification, they utilized CNN based module in order to extract 2D features from the rectified image. Further, they also proposed to adopt the SCM to extract semantic information and generated semantic context features. The major drawback of this method is that, the rectification is not accurate for blurred characters and at the same time, due to transformer module, the model is too complex and leads to slower in performance.

Zhong D et al[6] have proposed a novel method called as Text Proposal with Location-Awareness-Attention Network (TPLAANet) for arbitrarily oriented/shaped text detection and recognition. The text detection module locates text instances using central mask prediction, regressing rectangular bounding box for detecting the text region, and positions of arbitrarily oriented text instances are detected using mask branch. The recognition module uses character information using a Location-Awareness-Attention Network (LAAN). Xiong L et al[7] have proposed a scene text detection and recognition model based on iterative correction. The model consists of a rectification network which employs Bezier curve to estimate the bounding rectangle of the scene text region and uses TPS transform to iteratively rectify orientation of the text into horizontal text. The drawback of the method is that the estimation of parameters of TPS depends on the quality of regressed text borders. The method ignores the text content which gives not promising rectified results for severely distorted text.

You Y et al[8] have proposed curved text detection model based on Deformable Detection Transformer (DETR) called BSNet. They used pre-trained ConvNeXt model for extraction of visual feature maps at different scales. Then, feature maps are fed into the Transformer encoder to extract the features of the text regions. Text regions are further used to predict the B-Spline control point coordinates of the text contours using Transformer decoder. Finally, the boundary coordinates of the text regions are reconstructed using predicted B-Spline control points and obtains the final detection result. The proposed model employs sequence of steps such as pre-trained ConvNeXt for initial feature maps, utilization of Transformer encoder and decoder for B-Spline control points regression and reconstruction of text regions. Hence, the model requires huge number of images for training and need of fine tuning of excessive number of parameters which leads to high complex computation. Further, the Transformer encoder accepts visual feature maps of the input image which leads to loss of major visual information such as larger orientation of text. Hence, the model is not robust to major irregularities but yields highest accuracy for slight irregularities.

Xu N et al[9] have proposed a B-spline based text spotting method for curved scene text spotting. They represent curved text layout using a B-spline method with few parameters. In order to detect the scene text location, they employed special network called as Feature Pyramid Networks (FPN) to extract three levels feature maps from a single images and each level having its own module to detect text region. Each module is small network predict the text region and supervised method is used to merge the predicted text regions from three level networks. Then, they apply Spatial Transformer Network (STN) to rectify the text region. Finally, they used CNN-LSTM framework for recognition of scene text. Chen L et al[10] have proposed end-to-end text spotting network called as SPRNet for arbitrary scene text. They used parametric B-spline based representation model to represent the global shape of the text. The text is detected by regressing its shape parameters. Based on the text's shape parameters, they rectify an irregular text using adaptive projection transformations. The rectified text is fed into the subsequent light – weight text recognition network.

1.1 Research Gap

The review of the recent existing papers developed for arbitrary oriented text indicates that some methods meant only for detection followed by rectification and other methods meant only for recognition task, because, some of the real time applications only need either detection or recognition. There is a need of both tasks simultaneously for some applications such as self-driving vehicles/visual navigation where precise and accurate results are expected. The researchers have developed methods[1–4] that perform detection/rectification and recognition simultaneously for sensitive real time applications. In these methods, the success of recognition module depend on the detection/rectification module, hence, these methods yield inconsistent results for arbitrary oriented text. In order to provide consistent results, the researchers have developed hybrid methods that perform both tasks simultaneously such as detection/rectification and recognition for arbitrary oriented text [5–7]. However, these methods which consider both text detection and recognition do not detect the text and recognize the characters correctly for arbitrarily oriented/shaped text. The above discussion concludes that detection/rectification and recognition of arbitrarily shaped text are challenging. The challenges are irregular space between characters, oriented shape of every character in the curved text line. Therefore, there is a requirement of an effective method which should detect/rectify and recognize irregular spaced characters, and oriented characters simultaneously instead of only text detection and only recognition. It indicates that there is a need of effective and reliable model to address challenges of arbitrarily shaped text recognition. We claim that there is a research gap between the existing methods and challenges which are mentioned above. In order to overcome the above mentioned challenges, we developed a method for detecting/rectifying and recognizing oriented scene text in natural scene images based on the following observations in the existing methods.

All the existing end-to-end hybrid methods [1–4] are deep learning approaches yields inconsistent and inaccurate results for irregular spaced and oriented characters because during rectification process, they yield severe distortion or even important visual information loss which results in inconsistent recognition. Particularly, end-to-end hybrid systems i.e. Thin Plate Spline (TPS) - spatial transform network (STN) models [1–4] may results in unsatisfactory rectified text for irregular spaced and oriented/distorted text due to its non-linearity and the calculation of TPS transformation parameters purely depends on the quality of regressed text borders. The poor performance of TPS-SPN methods during rectification process leads to inconsistent results during recognition because there is an inter-dependency of rectification and recognition module. The methods [7–10] are also deep learning based end-to-end hybrid system that utilizes a parametric B-spline curve model to detect the global shape of irregular spaced and oriented text through weakly supervised learning based regressing of shape parameters. Then, they rectify oriented text using transformations such as SPN or adaptive projection transformations. Finally, the rectified text is fed into text recognition network which is also based on deep learning. The end-to-end hybrid methods such as TPS-SPN methods [1–4] and B-spline curve based methods [7–10] which are reviewed above yields unsatisfactory rectified text for irregular spaced and oriented/distorted text which results in inconsistent recognition. Also, these methods are based on weakly supervised learning and often require large amounts of labeled data for training. Further, the curved scene text image datasets are so limited in size resulting in poor performance on unseen datasets and collecting/ annotating large datasets is time-consuming and expensive.

In order to overcome the above mentioned difficulties of the end-to-end hybrid methods, we propose a rectification model for curved scene text based on B-spline curve fitting method and proposed to use handcrafted features for recognition of rectified scene text. The proposed method consists of rectification module which is different from the rectification module of the existing end-to-end hybrid methods [7–10] where they use weakly supervised regression process of shape parameters. Our rectification module does not require weakly supervised learning based regression process for rectifying the entire scene text area. Instead of computing the transformation of global shape of entire scene text, we compute local transformation of individual character in the distorted text. We consider individual character and employed B-spline curve fitting in order to compute simpler local transformations. Then, we rectify the distortion of each individual character through the computation of normal vectors on the curve and we rotate each character individually based on the resulting angle. The rectified scene text is fed into the text recognition model where handcrafted features are utilized unlike deep learning network adopted by end-to-end hybrid existing methods. The handcrafted features reduce lot of hurdles posed by deep learning network such as lack of labeled data and computation time leads to good recognition with high accuracy.

1.2 Contributions

The major contribution of this work is developing a system that is able to rectify and recognize curved texts in natural scene images. In summary, the main contributions are listed as follows:

1. We propose a text shape rectification model based on B-spline curve fitting that can rectify arbitrary shaped scene text for robust recognition.
2. The proposed method don't require manually designed networks such as TPS-SPN and supervised learning based regressing is not required for regressing the boundary coordinates of text regions.
3. We propose to adopt hand crafted features for recognition which are more precise and robust and reduces extra computation complexity compared to deep features.
4. The proposed rectification method effectively improves the performance of the text recognition model compared to state-of-the-art methods which is investigated using benchmark datasets in the experiments.

2 Methodology

We proposed a technique to rectify and recognize curved texts in scene images. The Figure 1 shows flow diagram of the proposed method. We perform curved text detection on the input scene image using text detection method [11]. After detection of curved texts, the image is binarized through a local thresholding technique [12]. To rectify the curved text, we adopted B-spline curve fitting technique [13] which provides local approximation of the contours. Finally, we compute the normal vectors on the curve and each character is rotated individually such that the normal vector is vertically aligned. Finally, we perform recognition of scene text using sequential steps such as character segmentation, features extraction from characters and then recognition using Tesseract OCR. The rectified text is feed to our recognition module where we segment the rectified text using Low Variation Extremal Regions(ER) technique and extracted the Pyramid Histogram of Oriented Gradients (PHOG) features from the segmented texts. Finally, we perform recognition of text using Tesseract OCR.

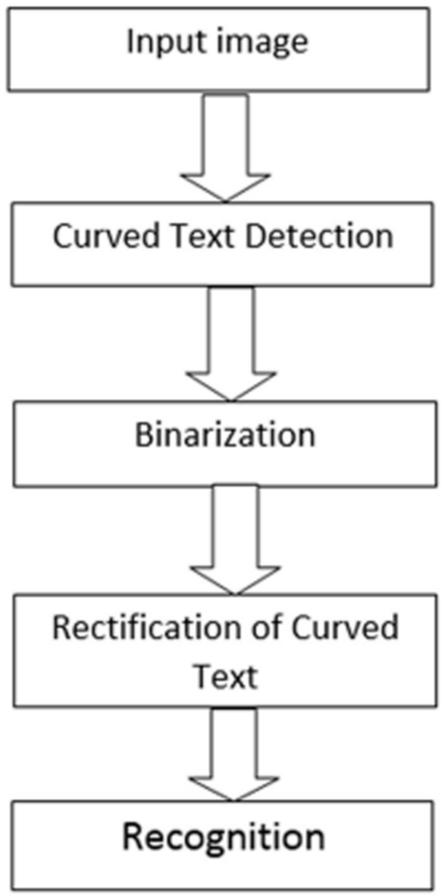


Fig 1. The flow diagram of the proposed method

The proposed method was evaluated using benchmark datasets through various experiments in order to prove efficacy of the method. The experiments were conducted in two phases. In the first phase, only the proposed method is evaluated using CUTE80 dataset for both rectification and recognition module. In the second phase, we compared the results of the proposed method with state-of-the-art methods using challenging benchmarks such as ICDAR2013, Street View Text (SVT), IIIT5k-Words (IIITK), ICDAR2015, SVT-Perspective (SVTP) and CUTE80 dataset based on word level recognition accuracy. While evaluating the proposed method for rectification, we used Mean Square Error (MSE) metric and word level recognition accuracy is used to evaluate the text recognition module. MSE metric helps to measure the amount of error removed by curve fitting on the input and the output images. We calculate MSE using the following equations:

$$y = a + bx \tag{1}$$

The Least Squares use the above equation to approximate the set of data $(x_1,\ y_1),\ (x_2,\ y_2),\ \ldots,\ (x_n,\ y_n)$, where n $\geq$ 2. It is always true that, the best fitting curve (x) should have the least square error.

$$\sum_{i=1}^{n}\left(y_i - f\left(x_i\right)\right]^2 = \sum_{i=1}^{n}\left(y_i - \left(a + bx_i\right)\right]^2 = min \tag{2}$$

The obtained recognition results are evaluated using word level recognition accuracy is defined as follows:

$$accuracy = \frac{Number\ of\ \ correctly\ recognized\ \ words}{Total\ Number\ \ of\ \ Word\ \ images} \tag{3}$$

2.1 Curved Text Detection

In the proposed method, the first step is detection of curved text in order to detect and locate curved text area which results in boundary of text region. We adopted the method proposed by Fan et al [11] which is an instance aware curve text detector called as Look More Than Twice (LOMT). It consists of two modules such as curved text approximation and component merging modules. Figure 2 shows curved text detection results on sample images of CUTE80 dataset.

Fig 2. Curved Text detection results on sample images of CUTE80 dataset

2.2 Binarization

The detection of curved text yields boundary of text region which is need to be binarized which facilitates further process. In order to binarize the image, we adopted the binarization technique [12], which is an improved type of local binarization. The adopted binarization method performs very well on scene text images. It has three major sections: variance calculation, boundary linking and adaptive thresholding. In variance calculation, the boundary pixel of a text region always has higher gray values than its background pixels even in presence of noise. Therefore, detecting text boundary location using n x n window based on variance is more robust to noise compared to any other techniques. The variance calculation produces a variance map, which is a rough sketch of the text boundaries region. Due to low variance in some regions, a discontinuity arises in the variance map. To overcome this challenge, a canny edge detector is used on the gray image to generate all the important edges. Edges connected to the boundary lines are kept. This task recovers all the boundary lines including those that might have been missed by the variance calculations. Lastly, adaptive thresholding is performed on the image based on the variance map that was generated on the previous steps. The Figure 3 depicts the binarization results on sample images of CUTE80 dataset.

2.3 Rectification of Curved text using B-Spline

We applied B-spline curve fitting technique [13] on binarized image which approximates the region of the curved text. We need to get the pixels in a certain order that follows a path coinciding with the curve. This is achieved by applying thinning to convert the region to one pixel thick line. It works very well to produce the desired results. B-spline curve fitting is described as follows:

Given n+1 control points $P_0,\ P_{1,}\ldots, P_n$ and a knot vector $u = u_{0,}, u_1, \ldots, u_m$, the B-spline curve of degree p defined by these control points and knot vector u is

$$C\ (u) = \sum_{i=0}^{n} N_{i,p}\,(u)\,P_i, \tag{4}$$

Fig 3. The sample results of binarization on scene text images of CUTE80 dataset. The first column shows the original scene text images and second column shows the binarized curved text.

where $N_{i,p}$(u) are B-spline basis functions of degree p. The degree of B-spline basis function is defined by p. the ith B-spline basis function of degree p, is written as $N_{i,p}$(u), is defined recursively as follows:

$$N_{i,0}(\mathrm{u}) = \begin{cases} 1 & \text{if } u_i \leq u < u_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$N_{i,p}(u) = \frac{u-u_i}{u_{i+p}-u_i} N_{i,p-1}(u) + \frac{u_{i+p+1}-u}{u_{i+p+1}-u_{i+1}} N_{i+1,p-1}(u) \tag{6}$$

If we have $n+1$ data points $D_{0,}D_1,\ldots,D_n$ and want to find a B-spline curve that can follow the shape of the data polygon containing data points, then two more inputs are still required. The first input is the number of control points and the second input is the degree(p), where $n > h \geq p \geq 1$. With these two inputs, a set of parameters and a knot vector can be determined. Let the parameters be $t_{o,}t_1,\ldots,t_n$. The number of data points is equal to the number of parameters. Then, the approximation B-spline of degree p is given by:

$$C(u) = \sum_{i=0}^{h} N_{i,p}(u) P_i \tag{7}$$

Where $P_{0,}P_{1,}\ldots, P_h$ are the $h+1$ unknown control points. After passing the first and the last data points to curve, $D_0 = C(0) = P_0$ and $D_n = C(1) = P_{h,}$ there are only $h-1$ unknown control points. Considering this, the curve equation becomes as follows:

$$C(u) = N_{0,p}(u) D_0 + \left(\sum_i^{h-1} N_{i,p}(u) P_i\right) + N_{h,p}(u) D_n \tag{8}$$

To find out control points $P_1,\ldots,P_{h-1}$ such that the function $f(P_1,\ldots,P_{h-1})$ is minimized. The approximation is done using least square method. The sum of all the squared distances is

$$f(P_1,\ldots,P_{h-1}) = \sum_{k=1}^{n-1} |D_k - C(t_k)|^2 \tag{9}$$

$$\sum_{k=1}^{n-1} N_{g,p}(t_k) \sum_{i=1}^{h-1} N_{i,p}(t_k) P_i = \sum_{k=1}^{n-1} N_{g,p}(t_k) Q_k \tag{10}$$

Since we have $h=1$ variables, g runs from 1 to $h-1$ and there are $h-1$ such equations:

$$P = \begin{bmatrix} P_1 \\ P_2 \\ \cdot \\ \cdot \\ \cdot \\ P_{h-1} \end{bmatrix} \tag{11}$$

$$Q = \begin{bmatrix} \sum_{k=1}^{n-1} N_{1,p}(t_k) Q_k \\ \sum_{k=1}^{n-1} N_{2,p}(t_k) Q_k \\ \cdot \\ \cdot \\ \sum_{k=1}^{n-1} N_{h-1,p}(t_k) Q_k \end{bmatrix} \tag{12}$$

$$N = \begin{bmatrix} N_{1,p}(t_1) & N_{2,p}(t_1)\ldots & N_{h-1,p}(t_1) \\ N_{1,p}(t_2) & N_{2,p}(t_2)\ldots & N_{h-1,p}(t_2) \\ & \cdot & \\ & \cdot & \\ N_{1,p}(t_{n-1}) & N_{2,p}(t_{n-1})\ldots. & N_{h-1,p}(t_{n-1}) \end{bmatrix} \tag{13}$$

The system of linear equation can be re-written as:

$$(N^T N) P = Q \tag{14}$$

$$P = (N^T N)^{-1} Q \tag{15}$$

Since N and Q are known, solving this system of linear equations for P gives the desired control points.

Computing Normal vectors

The i^{th} point of a normal vector on the curve is computed as follows:

$$n_i^k = \begin{bmatrix} \cos\left(\frac{\pi}{2}\right) - \sin\left(\frac{\pi}{2}\right) \\ \sin\left(\frac{\pi}{2}\right) \cos\left(\frac{\pi}{2}\right) \end{bmatrix} \times \frac{1}{2} \left(\frac{(p_i^k - p_{i-1}^k)}{\left\| p_i^k - p_{i-1}^k \right\|_2} + \frac{(p_{i+1}^k - p_i^k)}{\left\| p_{i+1}^k - p_{i-1}^k \right\|_2} \right) \tag{16}$$

where i represents the position on the curve where the normal vector is being computed.

Horizontal alignment of text

The normal vectors gives the rough estimate of character orientation. To rectify the text, we rotate individual characters until the normal vectors are aligned vertically. The angle of rotation is computed as follows

$$\theta_i^k = (90^\circ - \angle n_i^k) \quad i = 1, 2, \cdots, N^k \tag{17}$$

where, θ^k is the type of orientation of text string. The value of θ may be either positive or negative depending on the direction of rotation. Figure 4 demonstrates rectification results of the proposed method on curved scene text images. First column shows the original curved scene text images from CUTE80 dataset. Second column shows the binarized images and the third column shows the rectified scene text using B-spline curve fitting.

2.4 Recognition using Handcrafted Features

After performing the rectification of the curved text, we perform character segmentation where, we use Low Variation Extremal Regions (ER)[14] to segment the character candidates. The Low Variation Extremal Regions gives a good distinction in pixel values, the pixel values may be higher or lower compared to the boundary pixels. This variation in pixel values gives the capability to split the local paths more accurately. We extract the features from the segmented scene text using Pyramid Histogram of Oriented Gradients (PHOG)[15]. The PHOG extracts low-level features from the intended characters and passes these extracted features to the Support Vector Machine (SVM)[16] which classifies the characters as either a character or non-character based on the features. Lastly, we use the Tesseract Optical Character Recognition system for recognizing scene text.

3 Results and Discussion

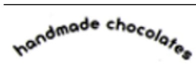
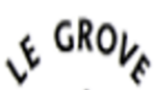
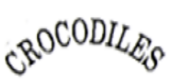
We implemented the proposed method using PyTorch framework and on a NVIDIA Tesla V100 GPU. We used the text centerline by a cubic B-spline (order n = 4) with 5 control points. We used Mean Square Error (MSE) metric while evaluating the proposed method for rectification and word level recognition accuracy is used to evaluate the text recognition module.

Fig 4. The rectification results of the proposed method; first column shows the original curved scene text images from CUTE 80 dataset. Second column shows the binarized images and the third column shows the rectified text using B-spline curve fitting.

3.1 Performance Evaluation of Our Approach

The CUTE80 dataset is used to evaluate the performance of the proposed method with respect to rectification and recognition. CUTE80 dataset consists of 80 curved scene text images with complex background, perspective distortion and poor resolution effects. We evaluated proposed rectification method through Mean Square Error (MSE). MSE metric helps to measure the amount of error removed by curve fitting on the input and the output images.

Table 1. The evaluation of rectification results of the proposed method using Mean Square Error (MSE) metric on CUTE80 dataset. The first column shows binarized text, second column shows MSE value before applying rectification, third column shows rectified text, fourth column shows MSE value after rectification and last column shows error removed

Binarized Text	MSE	Rectified Text	MSE	Error Removed(%)
	2,705,283162		77,912.550	97.12
	2,834,756.321		4,3655.227	98.46
	2,637,838.936		49,063.8228	98.14
	2,414,569.314		15,936.157	99.34

From Table 1, it is observed that the proposed method removes highest error (MSE is minimum) from the curved text (97.12% of error is removed from the curved scene text shown on the first row, 98.46 % for curved scene text shown on the first row, 98.14% for third row text, 99.34% for text shown on the last row). It is also clear from the Table 1 that the rectified texts (column three) completely aligned without loss of original information. Also, the proposed method is robust to perspective and rectifies texts with large angle. After rectifying the scene text, we performed recognition of the rectified images. The obtained results are evaluated using word recognition level accuracy. The correctly recognized word image is the one that has the same word string as the one in the word image. The characters in the image can range from upper case, lower case, special characters and digits ranging from '0-9'. Table 2 shows sample of recognition results of the proposed method after the rectification of the curved text. From the Table 2, it is observed that rectifying curved scene texts before recognition yields accurate results. The average recognition accuracy of the proposed method is 93.8% for curved scene text images of CUTE80 dataset.

Table 2. The recognition results of the proposed method on CUTE80 benchmark. First column shows the original images. The second column displays result of the binarization. The third column shows rectification results of our method. The fourth and fifth columns are the ground truth and the recognition results of the proposed method respectively

Input Image	Binarized Text	Rectified Text	Ground Truth	Recognition Result
	CROCODILES SWIMMING	CROCODILES SWIMMING	CROCODILES SWIMMING	CROCODILES SWIM-MING
	LE GROVE	LE GROVE	LE GROVE	LE GROVE
	WELCOME TO HISTORIC	WELCOME TO HISTORIC	WELCOME TO HISTORIC	WELCOME TO HISTORIC
	handmade chocolates	handmade chocolates	handmade chocolates	handmade chocolates

In order to prove the efficacy of the proposed rectification method, we performed the curved scene text recognition without rectification. The Table 3 shows recognition results on curved scene text images without rectification. The average recognition accuracy before rectification is 36.8% whereas the average recognition accuracy after rectification is 93.8%. This is due to the fact that the proposed method is failed to recognize some samples due to the orientation of the characters in the curved texts and they are not horizontally aligned.

Table 3. The recognition results of the proposed method without rectification of the curved text

Input image(original)	Binarized Image	Ground Truth	Recognition
	ALLAHABAD UNIVERSITY	ALLAHABAD UNIVERSITY	BAD UNIVP
	CROCODILES SWIMMING	CROCODILES SWIMMING	oolBDo
	AVILON	AVILON	Prilov

3.2 Comparison with State-of-the-Art Methods

In order to show the effectiveness of the proposed method, we compared the performance of the proposed method with state-of-the-art methods using six challenging benchmarks: ICDAR2013, Street View Text (SVT), IIIT5k-Words (IIITK), ICDAR2015, SVT-Perspective (SVTP) and CUTE80. These benchmarks are widely used in assessing Scene Text Recognition models and these benchmarks contain large number of scene text samples with irregularities. ICDAR 2013 dataset consists of 462 scene text images. The SVT dataset contains 350 images with blur, low resolution and irregular text. IIIT5k-Words (IIITK) consists of 5000 images. ICDAR 2015 dataset consists of 1500 images with text regions are multi-oriented. The CUTE80 dataset contains

80 high-resolution images with 288 cropped text instances. The images in CUTE80 have a complex background, perspective distortion, and poor resolution. We evaluate the performance of proposed method with state-of-the art methods which are popular rectifiers and designed to handle curved shape text through rectification prior to recognition. The state-of-the-art methods considered for comparison are

1. Shi et al[1] proposed an End-to-end Attentional Scene Text Recognition (ASTER) model that performs rectification and recognition of scene texts. It utilizes a flexible TPS transformation that rectifies various text irregularities, this model utilizes attentional sequence-to-sequence model for recognition.
2. Zheng T et al[3] proposed TPS++, which consists of the attention mechanism for text rectification. TPS++ estimates the parameters through foreground control point regression and content-based attention score estimation. TPS++ provides rectification process which is content-aware rectifier corrects a natural text that helps and make easier for subsequent recognition process.
3. Luo et al[17] have proposed a multi-object rectified attention network (MORAN) that is used to recognize general scene text objects, this model has two networks the multi-object rectification network and an attention-based sequence networks.
4. Zhang et al[18] have proposed a new learnable geometric-unrelated module, the Structure-Preserving Inner Offset Network (SPIN), which is based on the color manipulation of an input within the SPIN module. The SPIN consists of two components: the Structure Preserving Network (SPN) and the Auxiliary Inner-offset Network (AIN). The SPN helps alleviating the irregularities due to inconsistent intensities and AIN is an accessory network transforms the irregularities caused by consistent intensities.
5. Chen L et al[10] have proposed end-to-end text spotting network called as SPRNet for arbitrary scene text. They used parametric B-spline based rectification model for rectifying the curved text. The text is detected by regressing its shape parameters. Based on the text's shape parameters, they rectify an irregular text using adaptive projection transformations. The rectified text is fed into the subsequent light – weight text recognition network.

The recognition results of above mentioned popular rectifiers are used for comparison with our approach. We extracted the results of the state-of-the art methods from the paper[3]. Table 4 shows the comparison of our method with the state-of-the art methods. The comparative analysis demonstrates that our method outperforms the state-of-the-art methods by a large margin for curved text recognition. The performance improvements of the proposed method compared to state-of-the–art methods across six benchmarks for recognition of irregular datasets. The proposed method achieves considerable performance gains of 2.0%, 2.8%, and 8.0% on IIIT5k, ICDAR 2015 and CUTE80 datasets respectively. Also, our approach achieves 92.00% on SVT, 93.00% on ICDAR2013 and 83.00% on SVTP benchmark, which shows outperformance of the proposed method over popular text spotters. Overall, the proposed method yields highest recognition accuracy 93.80% on CUTE80 dataset. Particularly, compared to ASTER[1] which uses TPS-based rectification module and TPS++[3] that uses an attention mechanism for text rectification, our model achieves higher performance on six curved text benchmarks. Further, our model yields superior performance on six benchmarks compared to MORAN[17] and SPIN[18] which are popular rectifiers. We also compared the performance of the proposed method with the method proposed by the Chen L et al[10], where they proposed end-to-end text spotting network with parametric B-spline based rectification model. The performance comparison with the B-spline based rectification method shows that increase in accuracy for text recognition. The improved results indicate that the proposed method has a strong capability to rectify the text with large orientation angle. It also proves the effectiveness of B-spline based rectification prior to recognition. This is majorly attributed by the fact that text rectification increases the accuracy of recognition.

Table 4. Comparison of Recognition accuracy (%) of our approach with popular rectifiers

Method	IIIT5k	SVT	ICDAR2013	ICDAR2015	SVTP	CUTE80
MORAN[17]	93.00	87.50	90.90	74.30	79.10	82.60
ASTER[1]	93.50	88.90	92.30	76.70	80.50	84.70
SPIN[18]	93.70	89.30	91.30	78.20	78.80	83.00
TPS++[3]	94.50	91.50	92.80	78.20	82.80	85.80
Chen L et al[10]	95.00	91.00	92.00	78.50	82.00	86.00
Our Approach	96.50	92.00	93.00	81.00	83.00	93.80

The analysis of experimental results indicate that B-spline based rectification done by our approach is remarkable which is evident through evaluation carried using MSE metric on natural scene curved text images. The recognition module of our approach achieves highest accuracy of 93.80% compared to state-of-the-art methods on six benchmarks. The text in these datasets is curved in nature and contains complex backgrounds. Our method overcomes these challenges through the integration of adaptive binarization to deal with the complexity of the background and B-spline curve fitting is flexible in nature that can handle curved text shapes. The rectification results indicate that the B-spline curve fitting approximates both complex and simple curves with good performance compared to polynomial and Bezier curves. It is observed that polynomial curves are not well performed on approximation of complex curves and on the other hand, the approximation of pixel data on simple curves using polynomial suffers from Runge's phenomenon. The B-spline curve is free from Runge's phenomenon. Therefore, B-spline curve best suits for rectifying curved scene text. B-spline also captures the smoothness of any text layout that might be present in scene images. The proposed rectification model geometrically represents the global shape of every character and its intrinsic smoothness with a parametric B-spline centerline. Compared to boundary-based text representations employed in state-of-the-art text spotting methods, our centerline-based representation is less susceptible on the overall shape and regularity of the text. Moreover, the proposed method yields the complete boundary of the text and there is no need of applying post processing.

4 Conclusion

This study proposes a robust method for rectification of curved scene texts using B-spline curve fitting, which is followed by the recognition of rectified text using handcrafted features. The evaluation of proposed rectification module results using MSE metric obtained on the popular benchmark indicates that our rectification process removes highest error 99.34% for perspective text with irregular spaced and oriented text. Further, the ablation study conducted on the popular benchmark indicates that the recognition accuracy before rectification is 36.80% where as the recognition accuracy after rectification is 93.80%, which clearly shows that the proposed B-spline based rectification module boost the performance of the recognition module through aligning irregular spaced and an oriented text horizontally without loss of information. This is due to the novelty involved in the proposed rectification module compared to the existing rectification methods where we first detect the curved text using Look More Than Twice (LOMT) model, which gives text region boundary. Instead of using whole scene text image which is normally used by existing rectifiers, we use only detected text region boundary for further rectification process. Also, instead of considering the whole text region for computing the transformation which is normally utilized by most of the existing rectifiers, we compute the local transformation of individual character in the detected text region boundary through B-spline curve fitting.

Further, the recognition results of the proposed method are compared with the four state-of-the-art methods on six benchmarks which demonstrate that the proposed method achieves highest accuracy of 93.80%. The highest recognition rate of the proposed method compared to the existing methods is due to the novelty involved in the proposed method where we extracted PHOG handcrafted features unlike deep features used by the state-of-the-art methods. Another major reason for achieving the highest recognition accuracy is that an accurate rectification done by our rectification module prior to recognition which improves performance of the recognition module. The advantage of our approach is that unlike the state-of-the-art methods which are end-to-end deep learning systems, our recognition module uses handcrafted features which reduce lot of hurdles posed by deep learning network such as lack of labeled data and computation time leads to good recognition with high accuracy, which helps to utilize the proposed method for real time applications where computation time and accuracy matters a lot. Even though, we achieved the success in recognition of curved text using the proposed method, there are failures for some examples, where our method has an inaccurate performance when dealing with long and dense curved text. Another observation is that the proposed method degrades its performance when dealing with blurred text. In the NLP field, language models attract researchers due to its remarkable performance for various NLP applications, we mitigate the drawbacks of our approach in our future work where we planned to use these language models for multi-modal learning to correct blurred oriented characters in order to improve the performance.

References

1) Shi B, Yang M, Wang X, Lyu P, Yao C, Bai X. Aster: An attentional scene text recognizer with flexible rectification. 2019. Available from: http://dx.doi.org/10.1109/TPAMI.2018.2848939.
2) Wang W, Zhou Y, Lv J, Wu D, Zhao G, Jiang N, et al. Tpsnet: Reverse thinking of thin plate splines for arbitrary shape scene text representation. *Proceedings of the 30th ACM International Conference on Multimedia*. 2022;p. 5014–5025. Available from: https://doi.org/10.1145/3503161.3547882.
3) Zheng T, Chen Z, Bai J, Xie H, Jiang YG. TPS++: Attention-Enhanced Thin-Plate Spline for Scene Text Recognition. *Proceedings of the Thirty Second International Joint Conference on Artificial Intelligence*. 2023;p. 1777–1785. Available from: https://doi.org/10.48550/arXiv.2305.05322.
4) Wang J, Wu W. Research on Natural Scene Text Detection and Recognition. *In Journal of Physics: Conference Series*. 2021;1754(1). Available from: https://doi.org/10.1088/1742-6596/1754/1/01220.
5) Heng H, Li P, Guan T, Yang T. Scene text recognition via context modeling for low-quality image in logistics industry. . Available from: https://doi.org/10.1007/s40747-022-00916-1.

6) Zhong D, Lyu S, Shivakumara P, Pal U, Lu Y. Text proposals with location-awareness-attention network for arbitrarily shaped scene text detection and recognition. *Expert Systems with Applications*. 2022;205:117564–117564. Available from: https://doi.org/10.1016/j.eswa.2022.117564.
7) Xiong L, Gui Z, Ou Y, Xu W. Scene Text Detection and Recognition Based on Iterative Correction. *Proceedings of 5th International Conference on Robot Systems and Applications (ICRSA)*;2022. Available from: https://doi.org/10.1007/s11831-019-09315-1.
8) You Y, Lei Y, Zhang Z, Tong M. Arbitrary-Shaped Text Detection with B-Spline Curve Network. *International Journal of Sensors*. 2023;23(5):2418–2418. Available from: https://doi.org/10.3390/s23052418.
9) Xu N, Lian Z. An End-to-End Basis-Spline Based Text Spotting Network for Scene Text. *Proceedings of IEEE International Conference on Progress in Informatics and Computing*;2020:231–235. Available from: https://doi.org/10.1109/pic50277.2020.9350811.
10) Chen L, Su F, Shi J, Qian Y. Reading Arbitrary-Shaped Scene Text from Images Through Spline Regression and Rectification. *Proceedings of the Asian Conference on Computer Vision*;2022:2629–2645. Available from: https://doi.org/10.1007/978-3-031-26348-4_7.
11) Zhao F, Shao S, Zhang L, Wen Z. A straightforward and efficient instance-aware curved text detector. *International Journal of Sensors*. 2021;21(6):1945–1945. Available from: https://doi.org/10.3390/s21061945.
12) Ghoshal R, Roy A, Banerjee A, Dhara BC, Parui SK. A novel method for binarization of scene text images and its application in text identification. *Pattern Analysis and Applications*. 2019;22:1361–1375. Available from: https://doi.org/10.1007/s10044-018-0687-2.
13) Singh BM, Mittal A, Ghosh D. A novel method for straightening curved text-lines in stylistic documents. *EURASIP Journal on Image and Video Processing*. 2014;(1-8). Available from: http://jivp.eurasipjournals.com/content/2014/1/36.
14) Šarić M. Scene text segmentation using low variation extremal regions and sorting based character grouping. 2017. Available from: http://dx.doi.org/10.1016/j.neucom.2017.05.021.
15) Tan ZR, Tian S, Tan CL. Using pyramid of histogram of oriented gradients on natural scene text recognition. *Proceedings of IEEE International Conference on Image Processing*;2014:2629–2633. Available from: https://doi.org/10.1109/ICIP.2014.7025532.
16) Sarkar A, Chatterjee S, Das W, Datta D. Text classification using support vector machine. *International Journal of Engineering Science Invention*. 2015;4(11):33–37. Available from: http://www.ijesi.org/papers/Vol(4)11/G411033037.pdf.
17) Luo C, Jin L, Sun Z. MORAN: A multi-object rectifed attention network for scene text recognition. *Pattern Recognition*;2019:109–118. Available from: https://doi.org/10.1016/j.patcog.2019.01.020.
18) Zhang C, Xu Y, Cheng Z, Pu S, Niu Y, Wu F, et al. Spin: Structure-preserving inner offset network for scene text recognition. *Proceedings of AAAI Conference on Artificial Intelligence*. 2021;35:3305–3314. Available from: https://doi.org/10.1609/aaai.v35i4.16442.