1-10-2020

Automatic rectification of warped Bangla document images

Arpan Garai
Indian Institute of Engineering Science and Technology, Shibpur

Samit Biswas
Indian Institute of Engineering Science and Technology, Shibpur

Sekhar Mandal
Indian Institute of Engineering Science and Technology, Shibpur

Bidyut B. Chaudhuri
Techno India Group

IET Image Processing

Research Article

Automatic rectification of warped *Bangla* document images

ISSN 1751-9659
Received on 1st July 2019
Revised 16th September 2019
Accepted on 2nd October 2019
E-First on 21st November 2019
doi: 10.1049/iet-ipr.2019.0831
www.ietdl.org

Arpan Garai[1] ✉, *Samit Biswas*[1], *Sekhar Mandal*[1], *Bidyut B. Chaudhuri*[2,3]

[1]*Department of Computer Science and Technology, Indian Institute of Engineering Sciences and Technology, Shibpur, Howrah, West Bengal 711103, India*

[2]*ProVC, Techno India University, Salt Lake, Sector 5, Kolkata, India*

[3]*Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India*

✉ *E-mail: arpangarai@gmail.com*

Abstract: In this study, a robust algorithm for dewarping of camera-captured document images, mainly in *Bangla* script, is proposed. The algorithm can handle various types of warped document images and they are generated due to different types of document surfaces (convex, concave or multi-folded). The proposed algorithm is independent of font type, font size, font style and camera view angle. After initial preprocessing, the method first demarcates the text lines present in the document image. Then, the headline (*shirorekha*) position of each text line is estimated. Based on the headline position and shape, each text line is dewarped. If the document is highly warped, distorted text (e.g. thinner and shorter characters) is generated after dewarping. Special care has been taken to minimise this distortion based on most undistorted character information. Exhaustive testing shows the robustness and shape improvement of the proposed algorithm. Finally, for shape quality evaluation, some new measures are defined.

1 Introduction

In the present era, smart mobile phones and similar devices, digital cameras are frequently used to capture images of paper documents. As a result, various types of distorted images are often generated. Warping is one of the important distortions in such images. The performance of traditional OCR and other document processing tasks became far from satisfactory if we use warped document images as input. To improve the output quality we have to restore document images. The existing dewarping methods work well on printed *Alphabetic* scripts like *English* but cannot generate such good results for *Alpha-syllabary* scripts like *Bangla*. This drawback has motivated us to propose a new dewarping technique for *Bangla* script.

The alpha-syllabary scripts (like *Bangla/Devanagari*) have some unique shape properties. For example, the characters in these scripts are aligned at the upper region and connected with each other through a top headline, often called 'matra' or 'sirorekha'. On the contrary, the characters in Alphabetic scripts like *English* are aligned at the baseline and are rarely connected within a word.

Document images captured through mobile camera may contain different types of warpings. This is due to the camera angles or curvature of the document surface. A set of typical warped images captured by smartphone are shown in Figs. 1 and 2. In India, a huge number of posters, advertisements or notices are often pasted on round poles and other cylindrically shaped surfaces, which produce warping in captured images. Those documents may be pasted at a certain angle, i.e. skewed. It is difficult to handle errors that consist of skew as well as warping. In such a case, complex perspective distortions may be caused over and above the distortion due to warping.

In this paper, we propose a robust dewarping technique for document images with *Bangla* (sometimes mixed with *English*) script, which may suffer from skewing. The method is mainly based on estimating the top line or '*matra*' of text line written in *Bangla* script. This '*matra*' is used to the text line alignment in the document image. According to Liu *et al.* [1] polynomial fitting is not very suitable for drawing baseline from warped image (for English script). It is difficult to estimate the degree of the polynomial needed for a particular text line. The warping depends on the shape of the surface which may be wavy in nature. Such wavy surface leads to multiple folded warped images (see Fig. 2). So, a polynomial with too lower degree estimates inaccurate curve, while a polynomial with too higher degree leads to an overfitted curve. However, because of 'matra', such problem does not arise for *Bangla*. Here, we have automatically selected some sample points from the 'matra' position of the text line and drawn a curve using cubic spline interpolation. After dewarping, we perform two affine transformations namely, scaling and shearing to make the image distortion-free.

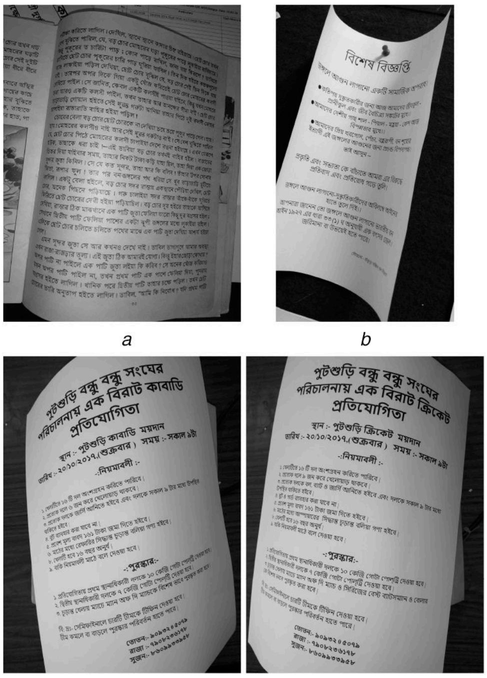
There are some datasets of the warped image available in *English* script [2–4]. However, hardly any dataset with ground truth is available for *Bangla* warped document images. So, the proposed method is tested using the *Warped Document Image Dataset* (*WDID*) [Samples available at https://sites.google.com/site/arpangarai/research-area/sample-images-of-wdid] developed only by Garai *et al.* [5]. The dataset contains images with the above-mentioned varieties of *Bangla* document images.

The rest of this paper is organised as follows. In Section 2, the works related to dewarping are briefly presented. Section 3 describes our proposed dewarping technique. Experimental results and evaluation of the proposed method are presented in Section 4. Finally, in Section 5, some concluding remarks are given.

2 Related work

Several types of dewarping techniques were proposed in the past. Some of them use external 'hardware' for the purpose, e.g. methods are given by Brown and Seales [6], Yamashita *et al.* [7] etc. Most other dewarped only using some software. Thus, the methods are often categorised as 'hard' and 'soft' methods, respectively [8].

The 'soft' methods generally consist of two sub-steps, namely implicit or explicit warp estimation and distortion correction. As per the warp estimation is concerned, the existing related methods can be classified into two categories: two-dimensional (2D) and 3D [1]. In the case of 3D methods, the shape of a page is estimated and transformed into a 3D shape using some geometric models. Some of these are the cylindrical surface model by Cao *et al.* [9], shape-from-shading model based on the luminance of the page surface by



a b c d

Fig. 1 *Different types of single folded warped document images captured from various sources*

(a) Skewed as well as a warped image from a book, ***(b)*** Paper hanging from a notice board, ***(c)*** Document posted on a lamp post (pasted slanted at left), ***(d)*** Document posted on a lamp post (pasted slanted at right)

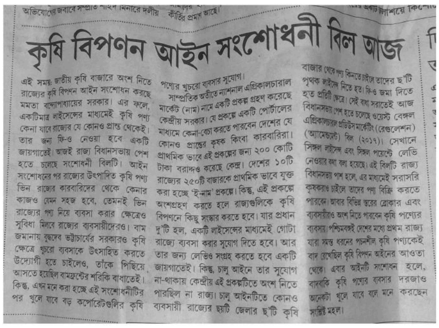


Fig. 2 *Multi-folded newspaper image*

Zhang and Tan [10], general cylindrical surface (GCS) model by Meng *et al.* [11] etc. Liang *et al.* [12] proposed a geometric framework based on texture flow to approximate the 3D document shape and rectify it. Meng *et al.* [11] use the 'line convergent symmetry', a property of the lines in a warped document. They have also introduced the 'para-perspective projection' for approximating the non-linear perspective projection. You *et al.* [13] proposed a method in which geometric property of the surface was estimated in the data and smoothness in the image. The dewarping was done using the geometric properties named 'developable surface'. A few methods proposed for a more specific domain like book page image dewarping system. He *et al.* [14] modelled the approach based on detecting the boundary of the book and approximating the document's 3D surface.

In the case of 2D methods, the restoration is guided by the text line flow and slantness of character in the text part of the document image. Ezaki *et al.* [15] have estimated the warp from the text lines using an elastic curve model. The method can handle local irregularities like formulae, short text lines and figures. A grid-based method was proposed by Lu and Tan [16] to dewarp the image. Here, the text direction was estimated using a vertical stroke boundary. The authors estimated x-line and baseline to get the horizontal and vertical curvatures by tracing the points. Ulges *et al.* [17] proposed a method for the rectification of both perspective and page curl distortion. The method estimated the geometric shape of the warped document using the local distances of adjacent baselines. Gatos *et al.* [18] segmented both words and lines for dewarping. Here, word-wise lower and upper baselines were constructed and slant of each word was calculated and corrected using its neighbourhood relationship. A line segmentation method to dewarp the document image by using coupled-snakes was proposed by Bukhari *et al.* [19]. Stamatopoulos *et al.* [20] rectified a warped document image by coarse rectification using page borders, followed by word baseline. In the technique proposed by Fu *et al.* [8] the boundary of the warped region is estimated. That boundary is assumed to be a rectangle in the un-distorted image of that document. The coordinates of each pixel of the document image are transformed according to the correspondence of that boundary and its approximated rectangular shape. Kim *et al.* [21] have used discrete representation of text line instead of using the baseline calculated from the text line. They have applied their approach in each of the text blocks present in the document. Liu *et al.* [1] proposed a technique where warping was estimated based on the shape of the baselines and slant of the characters. Ultimately, the authors used 'thin-plate spline' to rectify the warping. Kil *et al.* [22] have proposed an approach, in which the text lines and text line segments are used to measure the warping. The text lines are generally aligned in a non-warped document. Using this property, they have mapped the problem into a cost function. Using an iterative approach the document is dewarped. Recently, Yang *et al.* [23] have proposed a mesh-based technique to dewarp mainly the historical documents.

The common idea in most of these methods is that the authors had calculated the baseline and used it to compute the amount and direction of the warping. The characters in Alpha-syllabary scripts are top-aligned; as a result, the baseline does not reflect the exact text flow. Nevertheless, the characters are not separated like the characters in the alphabetic script. So, the conventional baseline drawing technique, proposed in the methods mentioned earlier, will not work here with good accuracy. Garai *et al.* [5] have proposed an alternative approach to dewarp document images containing Indic scripts like *Bangla* and showed some preliminary results. Morphological operations were used to find the headline. This paper is a highly extended version of that work.

3 Proposed approach

The proposed method consists of the following major steps. At first, the image is preprocessed by binarisation and removing border noise. Next, the text lines are segmented. Then, the headlines are estimated in the segmented text lines and dewarped using those headlines. In the case of the skewed document, the shearing distortion is still present after the horizontal alignment. So, a shearing-based correction is performed along with this alignment to correct the skewness of the document. Moreover, due to the variation of distances from the camera and camera angle, some of the text in the document becomes shorter and thinner. So, two post-processing steps are performed. They are scaling and shearing based corrections, respectively. Here, scaling-based correction is performed using scaling factors estimated for each pixel. The scaling factors are estimated from slopes and curvatures of the 'matra' and the local shearing factors for each pixel position. Finally, the shearing-based correction step is performed. The entire process is schematically presented in Fig. 3. Each step is discussed briefly in the following.

3.1 Preprocessing

As a preliminary step, we binarise and de-noise the document image. Several methods are proposed for document image binarisation in this literature like those in [24–26] etc. Among them, the method of [26] is robust and suitable enough to binarise the images of our dataset. In this method, a contrast map from a degraded document image is constructed using an adaptive way. From this contrast map, the text stroke edges are identified. Next, the text portion is segmented by local thresholds. The threshold in a local window is estimated from the intensities of text stroke edge pixels detected previously. Here we binarise the camera captured images using the method presented in [26].

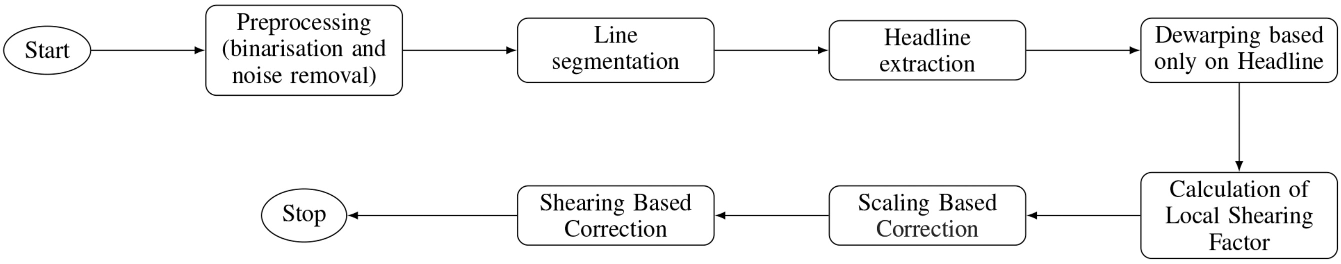


Fig. 3 *Flowchart of the proposed approach*

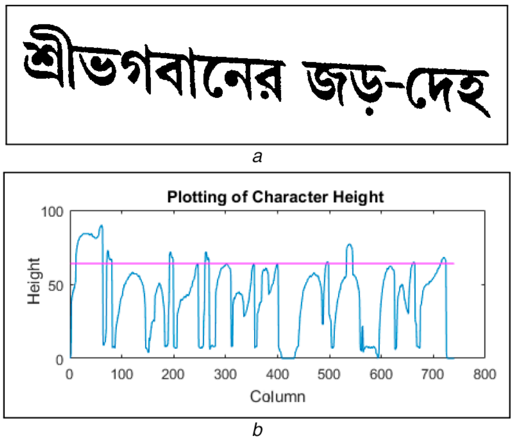


Fig. 4 *Calculation of height of a text line*
(a) Sample line, ***(b)*** Plotted height in each column; horizontal straight line (pink colour) represents the median of 'local maxima's of each 'mountain'

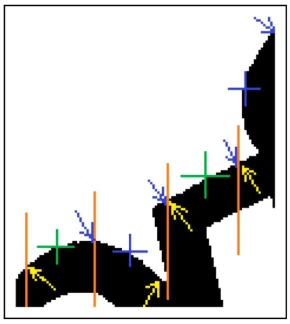


Fig. 5 *Choosing of the sample points: ' + ' are candidate sample points; vertical lines (orange), blue arrow and yellow arrow represent neighbourhood boundary, point with maximum and minimum y-values inside that neighbourhood region, respectively; points do not satisfy (1), marked with blue ' + ' are discarded; points satisfy (1) marked with green ' + ' are accepted*

In [27], the noise present in a document image is classified into six categories. They are ruled line noises [28], marginal noises [29], clutter noises [30], stroke-like pattern noise [31], salt and pepper noises [32] and background noises [33]. Here we have used images from our WDID. Most of the images there are captured in the presence of sufficient amount of light. The quality of the pages is also good, but there is the presence of marginal noises. So, we have removed the marginal noises of the binary images and forwarded the images in the next step. There are several methods to remove the border noises presented in [29, 34–38]. Among these methods the approach presented in [38] is designed to remove the border noise present in a camera captured document image. So, we adapt the way presented in [38]. In this method, the text and non-text parts are separated. Next, the text lines are segmented using a ridge-based method. After that, the page frame is detected from segmented text lines and previously segmented text and non-text parts.

3.2 Text-line segmentation

There are some text-line segmentation algorithms already proposed in the recent literature [39, 40–42]. However, for most of the cases the algorithms are such that it can segment the text lines in a document image having a flat surface. These works do not perform well for camera captured warped images. So, to segment the text lines we used methods proposed in [1], which can segment text lines in a document image having curved surface. In this method, the minimum bounding rectangle (MBR) of connected components is calculated. Next, the MBRs of the components are analysed and a minimum spanning tree (MST) is formed. Based on this MST, the text lines are segmented.

3.3 Headline extraction

As mentioned before in *Bangla* script, generally characters within a word are connected through the horizontal line, called a headline. Here this headline is extracted in each extracted text line. This headline is further used for dewarping.

The estimation of the headline is roughly done in three steps. They are (a) generation of candidate sample points, (b) collection of essential sample points and (c) estimation of headline by fitting a smooth curve through corrected sample points. Next, we briefly discuss each of these steps. It is found that the amount of distortion in a warped image varies with the size of the text present in the image. Hence, the thresholds used for the generation of the sample points are estimated based on the height of the text line (η). The values of the other set of thresholds are set empirically. So, the median height of a text line (η) in the warped document is calculated as follows.

Each column is scanned of the text line under consideration. Let $P_t(i)$ and $P_b(i)$ be the position of top-most and bottom-most of text pixels for the ith column, respectively, where $i = 1, 2, 3, \ldots, M$ and M is the number of columns of the considered text line. The difference of $P_t(i)$ and $P_b(i)$ ($H_s(i)$) are obtained. We plot H_s against the horizontal span of the text line and an example of such plot is shown in Fig. 4*b*. Next, from the plotting of H_s we determine the local maxima. Then the median of these maxima is considered as the median height of the text line, η.

(a) Generation of candidate sample points: We have found that for thinner components (i.e. components having one/two characters) in a text line often does not contain 'matra'. So, these components are separated using a threshold. After studying a large number of components in different types of images (more than 150 number of document images) we have selected the threshold as η. The connected components that are present within a text line are divided into two groups. The components having width $< \eta$ belong to one group (A) and the rest of the components fall in the other group (B). The top-most point of a component belonging to group A is considered as a sample point. For each component in group B, at first the upper profile is obtained. We get candidate sample points P_i (x_i, y_i) by scanning vertically at an interval $\sqrt{\eta}$ from the upper profile. This period between two sample points is set by doing experiments on different variations of warping in different types of text lines (more than 4500 number of text lines). Such points are marked by ' + ' in Fig. 5. Now, we consider all pixels within the horizontal distance $\sqrt{\eta}/2$ (i.e. pixels within $x_i \pm \sqrt{\eta}/2$) of P_i and their y-coordinates are noted. This region for each candidate sample point is shown using vertical straight lines in Fig. 5. Let $y_{\max}$ and $y_{\min}$ be the maximum and minimum y-coordinate values,

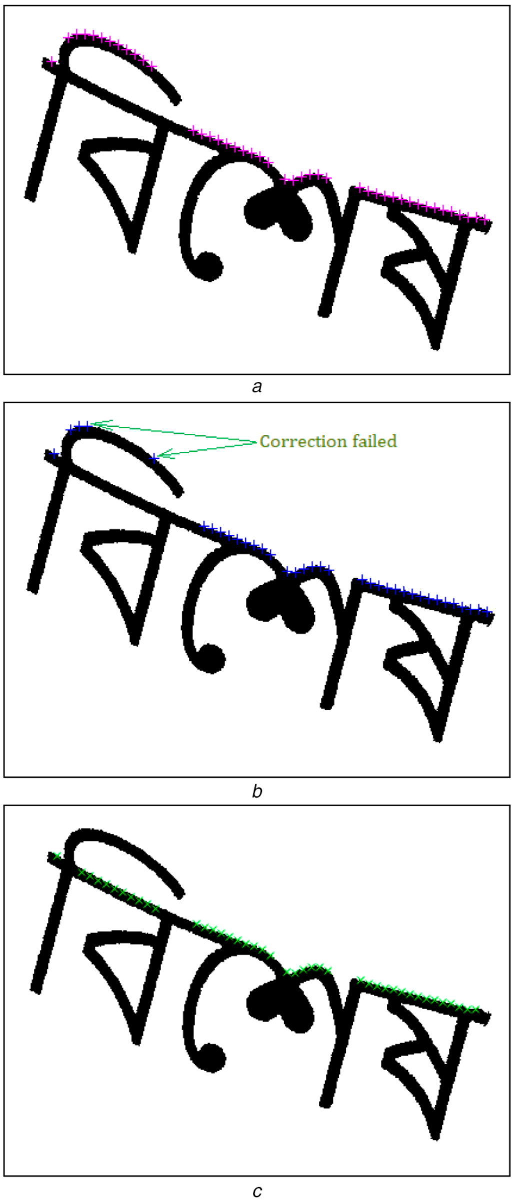


Fig. 6 *Sample point correction*
(a) Pink '+': initial sample points, ***(b)*** Blue '+': Corrected sample points (scanning left to right), ***(c)*** Green 'x': Corrected sample points (scanning right to left)

respectively. The point P_i is kept if the following equation is satisfied, otherwise it is discarded

$$\tan^{-1}\left|\frac{y_{\max} - y_{\min}}{\sqrt{\eta}}\right| < \frac{\pi}{6} \tag{1}$$

Here, these thresholds of the slopes are set by experiment over more than 4500 text lines. Fig. 5 visually illustrates the process.

(b) Collection of essential sample points: We assumed that the headline of a word locally follows (2), which is the equation of a curved line

$$y = ax^2 + bx + c \tag{2}$$

All the candidate sample points for a particular text line may not lie on this curve. Hence, the position of a few candidate points need to be corrected and some of the candidate points may be removed. For this, we calculate the slope (D_i') and the change of slope (D''_i) at each candidate point, where $i = 1, 2, \ldots, n$ and n is the number of candidate sample points. In the digital domain, (D_i') and (D''_i) are reduced to

$$D_i' = \frac{\mathrm{d}y_i}{\mathrm{d}x_i} = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}$$

$$D''_i = \frac{\mathrm{d}^2 y_i}{\mathrm{d}x_i^2} = \frac{D'_i - D'_{i-1}}{x_i - x_{i-1}}$$

To correct the positions of the candidate sample points, we scan the text line from both ends and the correction starts from the second sample point from its ends. Two consecutive sample points are considered (say ith and ($i-1$)th points). Let d_x be the horizontal distance between the ith and ($i-1$)th points and D_i' be the slope at the ith point. If $d_x \le \eta/3$ and $D_i' \le \tan(\pi/4)$, then the ith sample point is considered as essential sample point and a vertical scan starts from the ith sample point. We have done experiments on the different types of text lines having different font type, size and style of more than 150 number of images to set these thresholds. When the scan line encounters a white to black transition, the scanning is stopped and corresponding text pixel is taken as a candidate sample point. This sample point is considered as essential sample point if (1) is satisfied and the slope at this point is less than equal to $\tan(\pi/4)$.

Now, if $d_x > \eta/3$ and $D''_i \le T$, where T is a threshold value defined below, then the point is considered an essential sample point. If $d_x > \eta/3$ and $D''_i > T$, then a vertical scan starts from the ith sample point. When the scan line encounters a white to black transition, the scanning will stop and corresponding text pixel is taken as a candidate sample point. This candidate sample point is considered as an essential sample point if (1) is satisfied and $D''_i \le T$. Here, T is defined empirically as follows:

$$T = 1.5 \times \left| \frac{\dfrac{y_{i-2} - y_{i-1}}{x_{i-2} - x_{i-1}} - \dfrac{y_{i-3} - y_{i-2}}{x_{i-3} - x_{i-2}}}{|x_{i-3} - x_{i-1}|} \right|$$

Figs. 6*b* and *c* show the essential sample points when scanning is done from left-to-right and right-to-left, respectively. It is evident from these figures that the position of a few candidate sample points are not correct. Hence, extra effort is needed to correct their position.

Let S_l and S_r be the set of essential sample points when scanning is done from left-to-right and right-to-left, respectively. These two sets may not be the same, which is evident from Figs. 6*b* and *c*.

The sample points belonging to the S_l and S_r can be divided into three categories, as follows:

(i) Type-1: The set of points that belong to $S_\mathrm{l} \cap S_\mathrm{r}$.
(ii) Type-2: The element of this set is a pair of points whose x-coordinates are the same, but y-coordinates are different.
(iii) Type-3: The rest of the points belong to Type-3 category.

Fig. 7*a* shows the three types of sample points.

Let S be a set of points that are used to find the headline of a particular text line. The Type-1 points are surely the members of the set S. Next, we consider Type-3 points. Let P_3 be a point of Type-3 and P_p and P_s be the nearest predecessor and successor of P_3. Note that P_p and P_s belong to Type-1 set. We obtain the slope (D'_{3p}) between P_3 and P_p as well as slope (D'_{3s}) between P_3 and P_s. If $|D'_{3p} - D'_{3s}| < \tan(\pi/6)$, the P_3 is included in the set S.

For Type-2 points, each pair of points is considered such that the coordinates of the pair of points are (x, y_1) and (x, y_2). Consider Fig. 7*b*, the sample points at A and B are the Type-2 points and N_l and N_r are the left and right nearest neighbours of A and B, respectively. N_l and N_r belong to the set S. The Euclidean distances between A and N_l, and between A and N_r are obtained. The sum of these distances is calculated and it is given by d_A. The same has been done for point B. Suppose the summation of the distances is d_B. If $d_B < d_A$, then B is included in the set S and A is discarded.

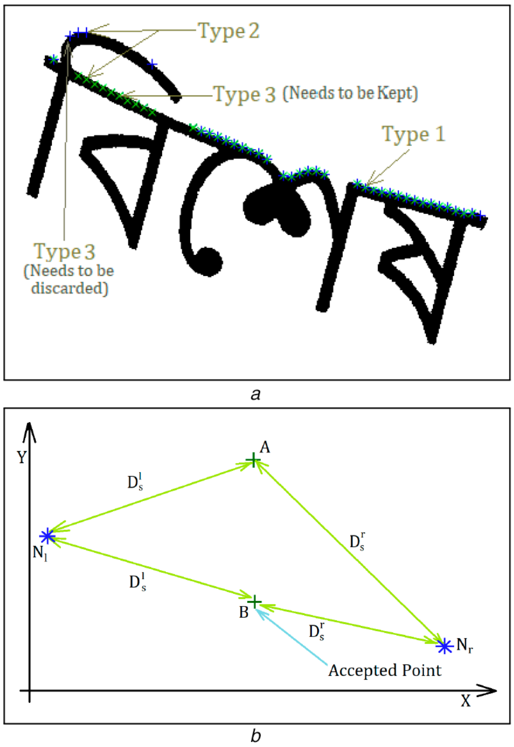


Fig. 7 *Different types of points*
(a) Type 1, 2, 3 points, ***(b)*** Choosing the correct point from each pair of point in Type 2

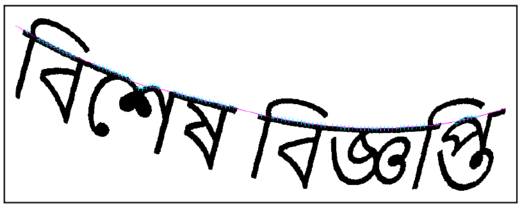


Fig. 8 *Finally considered sample points along with the calculated 'headline'*

Otherwise, A is the essential sample point. The final result for essential sample point collection is shown in Fig. 8.

(c) Smooth curve fitting: The set of points S is used to estimate the headline for a particular text line. We use a cubic curve fitting [43] algorithm on the set S which gives us the headline for a particular text line. Fig. 8 shows the result of the estimated headline.

3.4 Dewarping based on the headline

Now, we have to draw a horizontal line for each text line in the document image and we know that the equation of a horizontal line is $[Y_i = C_i + mX \mid i = 1, 2, \ldots, N; C_i = \text{constants}; m = 0]$. Here, N is the number of text lines present in a document. We have to calculate the value of C_i to draw the ith horizontal line. Consider the first horizontal line (AB see Fig. 9) corresponding the first line of the document. The vertical distances $[d_{1j} \mid j = 1, 2, 3, \ldots, M_1]$ between each point of the estimated headline and the first row of the image are obtained. Let, M_i be the number of points in the ith line of the document. Then, the value for C_1 (for the line AB) is calculated as $\left[C_1 = (1/M_1)\sum_{j=1}^{M_1} d_{1j}\right]$. For the rest of the text line in the document, the C_i is calculated as follows. For the ith $(i = 2, 3, \ldots, N)$ text line, the vertical distances $[d_{ij} \mid j = 1, 2, 3, \ldots, M_i]$ between each point of the ith and the $(i-1)$th estimated headline are measured. The mean (D_i) of these vertical distances is obtained $\left[D_i = (1/M_i)\sum_{j=1}^{M_i} d_{ij}\right]$ and C_i is calculated as $C_i = C_{i-1} + D_i$. The above-mentioned process is

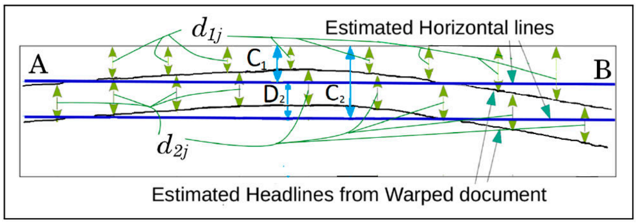


Fig. 9 *Estimated horizontal line (blue colour) from the headline (black colour)*

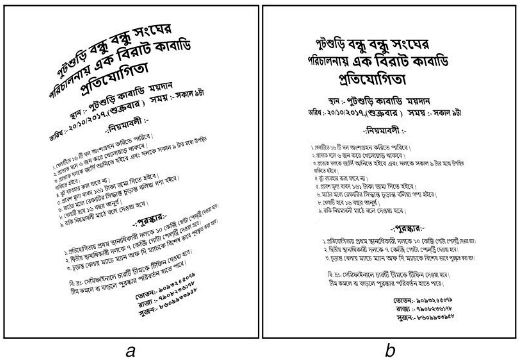


Fig. 10 *Results before scaling and shearing*
(a) Preprocessed image, ***(b)*** Dewarped image without scaling and shearing

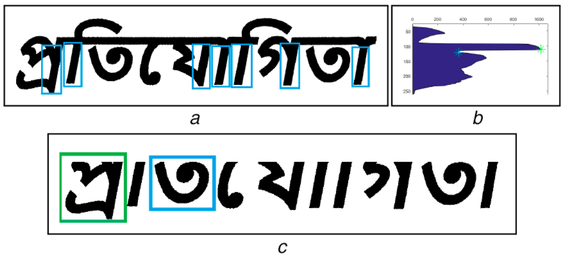


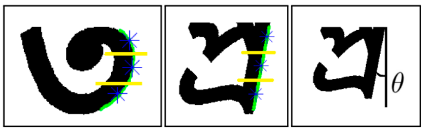
Fig. 11 *Automatic removal of 'matra'*
(a) Sample line (ROI for calculating shearing factor are marked by the blue box), ***(b)*** Horizontal projection profile image of Fig. 11*a*, ***(c)*** Separated components after erasing matra (headline)

depicted in Fig. 9. Each pixel in a particular text line is translated in such a way that its headline coincides with estimated horizontal line of that particular text line. In other words, the text pixels of each column of each text line are translated towards the y-direction. Let the equation of the ith headline is $Y_i = \rho_i(X_i)$. The vertical distances $[t_i(X_i) = C_i - \rho_i(X_i)$ for the ith text line] from calculated horizontal line to the estimated headline in each column are measured. These distances (t_i) are used as translation factors during the translation process. A preprocessed warped image and its dewarped version are shown in Figs. 10*a* and *b*, respectively.

3.5 Correction of dewarped image

The results of dewarping process may still have two types of errors; firstly, the size of the characters may not be the same within a text line, and secondly, the dewarped character image looks sheared. Hence, post-processing is required to minimise these errors. This kind of error occurs due to the presence of skew and/or perspective projection in the original document image. In the next subsections, we discuss the correction of these errors, where the scaling correction is followed by the shearing correction. However, at first, local shearing factors are calculated (discussed in Section 3.5.1) since it is needed for both scaling and shearing based corrections.

3.5.1 Calculation of local shearing factor: In *Bangla,* if we remove the matra of a word, then the characters within a word get separated. To do this, the horizontal projection profile (Fig. 11*b*) of each text line of the dewarped image is obtained. Analysing the



a b c

Fig. 12 *Calculation of the shearing factor*
(a) Component having a non-linear right profile (marked by blue box Fig. 11*c*), ***(b)*** Component having a linear right profile (marked by green box Fig. 11*c*), ***(c)*** Shearing factor $\tan\theta$

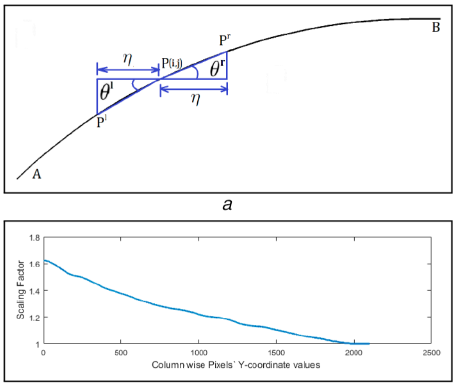


Fig. 13 *Estimation of scaling factors*
(a) Calculation of scaling factor, ***(b)*** Calculated scaling factor (column-wise) of the text line having 'matra' *AB* in Fig. 13*a*

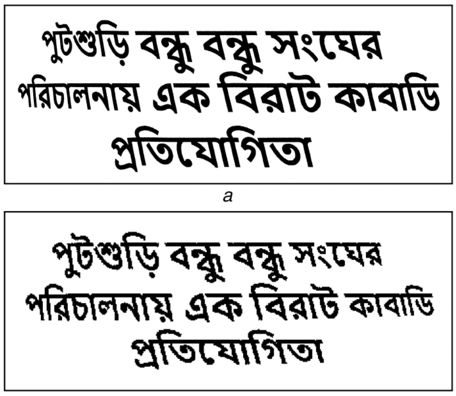
projection profile, the position of matra for each text word is estimated and removed. The result of character segmentation is shown in Fig. 11*c*.

The right projection profile of each sufficiently large character is obtained and the profile is divided into nearly equal parts. The middle point of each part is considered and curvature of the profile is estimated. If the curvature of the profile is close to zero (<0.001), then the angle (θ) between the profile and imaginary vertical line passing through the topmost point of the profile is calculated. We call $\tan\theta$ (see Fig. 12*c*) as a local shearing factor $\left(S_i^f\right)$. The local shearing factor for each character having linear right profile within a line is estimated. The mean of these slopes $\left(S^{mf}\right)$ for the entire text line is obtained.

3.5.2 Scaling-based correction: The restoration of size for each component in the dewarped document image is done through scaling. The scaling factor is computed with the help of slope and curvature. To illustrate the process of estimation of scaling factors, consider Fig. 13*a*. Here, *AB* is an estimated headline for a particular text line present in the warped document. Let $P(i, j)$ be a point on *AB*; P^l and P^r are its left and right neighbouring points at a distance η, respectively, η is the median height of the text line computed in Section 3.3. Three slopes considered here are m^l, m^r at a point P and the mean slope, S^{mf} for the entire document. The mean slope is the overall skew of the document.

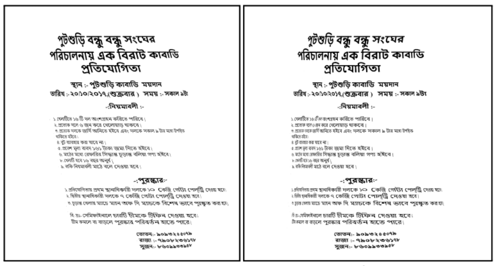
At first, the slopes (m^l and m^r) at P are determined with respect to P^l and P^r. These slopes are calculated using θ_l and θ_r, as shown in Fig. 13*a*. The ratio (R_c) between the curvature (C_k^r) of each kth text line and the minimum of those curvatures is ($C_{k\min}^r$). For multiple folded document images, we have considered the average curvature of all the folds, i.e. $R_c = C_k^r / C_{k\min}^r$. The mean slope, S^{mf} for the entire text line is already computed in the previous subsection. Now, the scaling factor (S_j^c), for all the pixels at column j for a particular text line is estimated as follows:

$$S_j^c = \frac{e^{\left(S_j^f - S_{\min}^f\right)} - 1}{R_c} + 1 \qquad (3)$$



a b

Fig. 14 *Results after scaling*
(a) Enlarged crop of the top portion of Fig. 10*b*, ***(b)*** After scaling-based correction of Fig. 14*a*



a b

Fig. 15 *Final Result*
(a) Result of scaling based correction for the image shown in Fig. 10*b*, ***(b)*** Result of shearing based correction after scaling

where S_j^f is obtained using the following equation:

$$S_j^f = \sqrt{1 + |\tan S_A|}$$

$$S_A = \left[\tan^{-1}\left(\frac{|m^r| + |m^l|}{2}\right) - \tan^{-1}\left(S^{mf}\right)\right]$$

Here, $S_{\min}^f$ is the minimum of S_j^f. During scaling, the effect of skew needs to be corrected. This correction is done by subtracting the overall skew $\left(\tan^{-1}S^{mf}\right)$ from the warping angle, $\tan^{-1}((|m^r| + |m^l|)/2)$. The calculated scaling factors of the line *AB* in Fig. 13*a* are shown in Fig. 13*b*. S^{mf} is the mean slope calculated in the previous section. The modified coordinates of the point P are obtained as $i' = (i - C) \times S_j^c$ and $j' = j \times S_j^c$. Here, C is the intercept value of the headline for a particular text line in the dewarped document. This process is done for each column of a particular text line. Due to scaling, the image size is increased and nearest-neighbour interpolation is used to interpolate the missing pixel values. The inter-line gap is incremented by the $[(1 - \max\,(S_j^c)) \times \eta]$ unit. An output of the scaling operation is shown in Fig. 14. The position of the local slants is adjusted according to the modification done in this step on the document image.

3.5.3 Shearing-based correction: Shearing correction is done using the local shearing factor, which is estimated in Section 3.5.1. There are some components for which the local shearing factor is not estimated due their shape. For such components, shearing factor is approximated considering their neighbouring components. The result of shearing correction is shown in Fig. 15*b*.

4 Experimental results

We have implemented our algorithm in C in a PC (Intel(R) Core(TM) i7-6700 3.4 GHz CPU, running Ubuntu 16.04). The

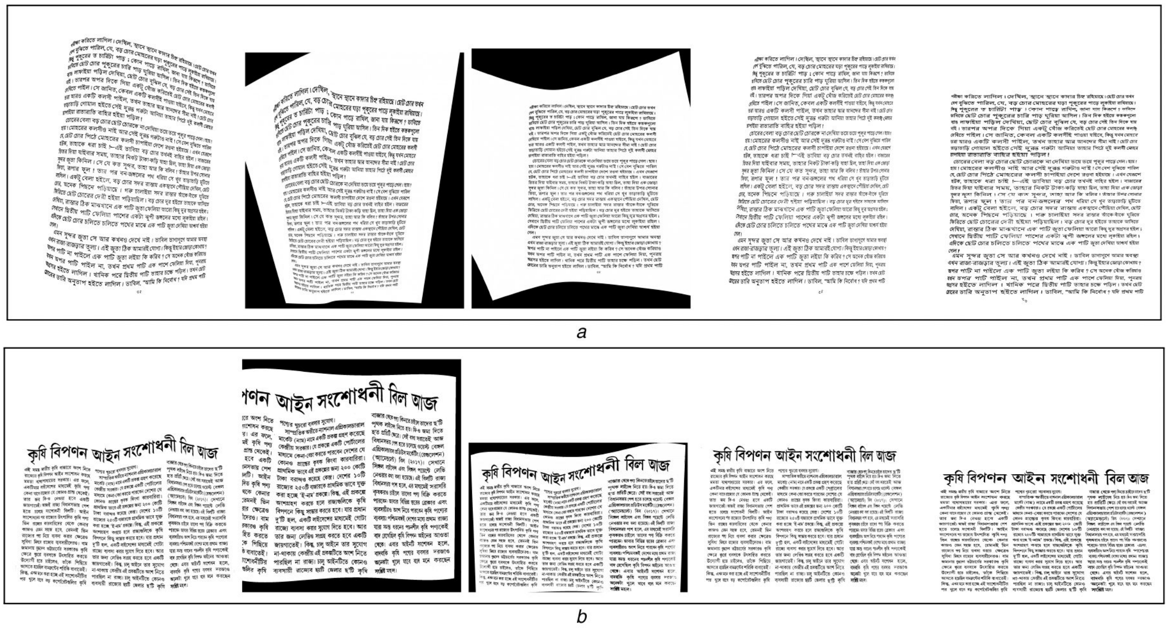


Fig. 16 *Preprocessed input image (1st column) and corresponding results using Kim et al. [21] (2nd column), Kil et al. [22] (3rd column), Garai et al. [5] (4th column) and proposed method (5th column), respectively*
(a) Book page image, ***(b)*** Multiple folded newspaper image

proposed method has been tested over various warped document images taken from different sources, scanned at 300 DPI, using HP Scanjet 5590. Since there is no benchmark database available for Bangla script, we have created our own *Bangla* warped document image dataset.

4.1 Dataset used

To the best of our knowledge, two warped document image datasets are publicly available, namely *DFKI-1* [2], *IUPR dataset* [3] and Ke Ma dataset [4]. All these datasets consist of warped document images containing the English script. We have created a WDID for our purpose. The dataset contains 221 various types of warped *bangla* document images; like document posted on a lamp post, document hanging on a notice board, multiple folded big document pages like newspapers etc. These varieties are not present in the previously stated datasets. In WDID, we have captured the images using digital cameras and mobile phone cameras. The images are taken from various distances. The distance from the surface and the lens of the camera varies from 10 to 80 cm. The lighting conditions are also different. Some pictures are taken at day time and some pictures are taken at night where the document is illuminated by neon/LED light. The text content of the pages has difference in fonts types, size and style. To get that, we have chosen books/magazines/pages printed at different times. The time of printing the documents ranges from 1985 to date. To create multiple folded documents, we have considered sufficiently big size of the document and folded it multiple times. Then unfolded it to capture the image. Some of the images are warped and some of those also have a fair amount of skew associated with it. The maximum amount of skew angle is considered in the dataset is $\pm 12^{\circ}$. In our data set, we have considered documents with a high degree of curl. The maximum amount of curl angle [44] considered is $\pm 45^{\circ}$. Most of the images have resolution more than 3000×4000.

4.2 Performance analysis and comparison

We have evaluated the performance of the proposed method on the WDID dataset, which contains different types of warped document images. The performance of the proposed method is also compared with existing state-of-art techniques. The comparison is made in two ways, (i) qualitative and (ii) quantitative. These are detailed in the subsequent sections.

4.2.1 Visual comparison: The performance of the proposed approach is visually compared with three existing related techniques. We refer to these here as Kim-method [21], Kil-method [22] and Garai-method [5]. For visual comparison, four sets of images are presented in Figs. 16 and 17, respectively.

It is evident from Figs. 16 and 17 that Kim-method and Kil-method do not produce an ideal result for Bangla scripts, though the authors claimed that their methods are script independent. Black patches are present at the border of the output images of these two methods. Garai-method is developed for Bangla script and its result is better than Kim-method and Kil-method. In the case of concave and convex surfaces (Fig. 17), Garai-method does not produce good results, whereas our method produces fairly good results. For all the input images, which are presented in Figs. 16 and 17, the proposed method visually outperforms the others.

4.2.2 Quantitative evaluation: To evaluate the performance of the proposed method, we have an indirect way of measuring the performance of the method using *Google Doc OCR*. The OCR accuracy is estimated as presented in [5]. The performance of the proposed method is presented in Table 1, along with performances of the other existing method. It is evident from Table 1 that the performance of the proposed method is better than the existing methods.

OCR for historical documents or handwritten documents may not always perform well. Moreover, it does not provide a quantitative measure of the visual correctness of the dewarped image. Stamatopoulos *et al.* [45] have already proposed an approach to evaluate the dewarping methods. However, it does not check whether the slant error is corrected or not. Therefore, we evaluate the performance of the proposed method in terms of restoration accuracy for the italic and bold type characters (α_i and β_i), which are used in [5]. The α_i and β_i are defined as $\alpha_i = (\theta_d/\theta_g) \times 100\%$ and $\beta_i = (B_d/B_g) \times 100\%$, respectively. Here, θ_d and θ_g are slant angles w.r.t vertical axis of the same italic characters in dewarped images and its equivalent ground truth images, respectively. Here, B_d and B_g are stroke widths of the same characters in dewarped images and its equivalent ground truth images, respectively. Table 2 shows the restoration accuracy for italic and bold characters of the different methods. Here, also the performance of the proposed method is superior. α_i fails to give an accuracy of non-italic fonts as θ_g for normal fonts often goes to 0. Hence, we introduce a new parameter to evaluate the local *slant-ness* in a document defined as *mean local slant error percentage*

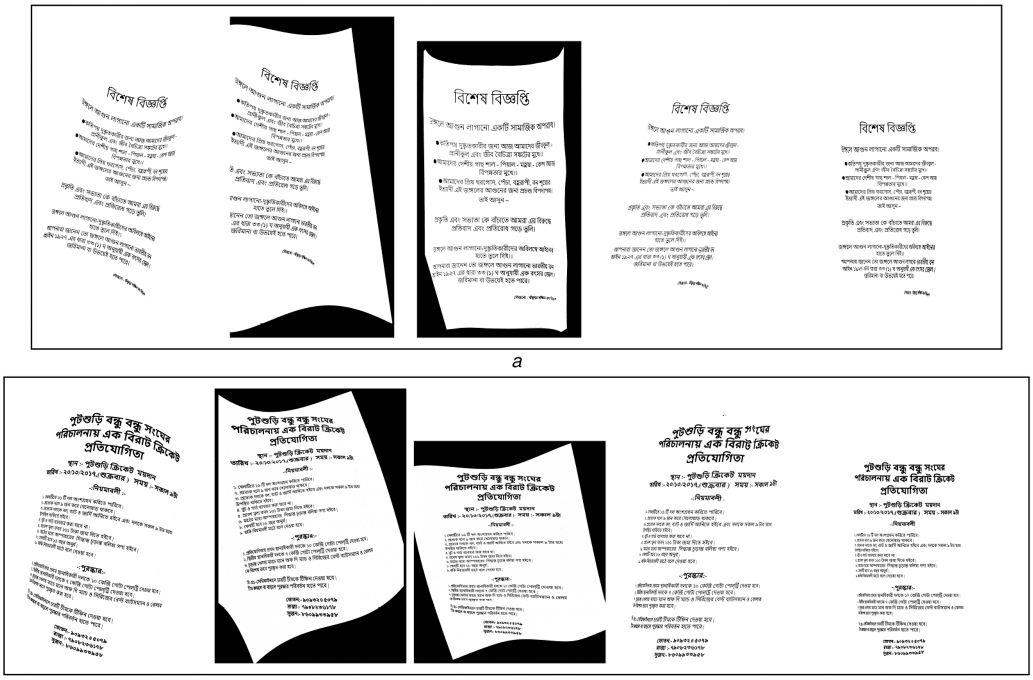


Fig. 17 *Preprocessed input image (1st column) and corresponding results using Kim et al. [21] (2nd column), Kil et al. [22] (3rd column), Garai et al. [5] (4th column) and proposed method (5th column), respectively; images captured form*
(a) Paper hanging on a notice board, ***(b)*** Paper pasted on a Lamppost

Table 1 Performance analysis in terms of OCR

Born Digital Database	Curve type	Language	OCR rate				
			Before dewarp	Kim *et al.* [21]	Kil *et al.* [22]	Garai *et al.* [5]	Proposed approach
Set 1(78 images)	left	bangla	negligible	51.03	84.5	97.08	98.32
Set 2(60 images)	right	bangla	negligible	55.07	83.5	97.40	99.09
Set 3(81 images)	other	bangla	negligible	50.04	86.1	95.81	99.18
Set 4(20 images)	—	mixed	negligible	54.05	85.2	98.04	98.14

Table 2 Restoration accuracy for *Bold* and *Italic* type characters

Text type	Kim *et al.* [21]	Kil *et al.* [22]	Garai *et al.* [5]	Proposed approach
		Restoration accuracy-bold (β_i)		
bold	88.1	89.3	91.1	91.8
bold + italic	80.5	87.3	92.0	92.5
		Restoration accuracy-italic (α_i)		
italic	61.1	90.3	92.48	92.3
bold + italic	60.5	89.6	91.36	92.5

(*MLSEP*) and it is described in the next paragraph. Nevertheless, we have also proposed an index to judge the warping. In all these cases, lower parameter value suggests better performance of the algorithm.

MLSEP: In *Bangla* scripts, some characters have a linear right profile that makes an angle of $\pi/2$ with the 'matra'(for non-skewed type characters Fig. 18). The mean local slant error percentage (γ_l) is estimated as follows:

$$\gamma_l = \frac{\sum_{i=1}^{N} |\theta_l(i)|}{N} \times 100 \tag{4}$$

Here, $\theta_l(i)$ is the angle between the right profile of the ith connected component and the vertical line passing through the rightmost point of that connected component in dewarped image (as shown in Fig. 18). These components are those having a linear right profile that creates an angle of $\pi/2$ with the 'matra' in its equivalent ground truth image. Let N be the total number of such components in a particular dewarped image. Restoration accuracy for the characters with the normal font is measured using *MLSEP*, as shown in Table 3. To estimate the mean local slant error percentage, we have only considered warped images having skew. It is clear from Table 3 that the performance of the proposed method is better than that due to the others.

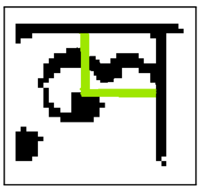


Fig. 18 *Component having a linear right profile parallel to the vertical axis*

Warping correction index: We introduce another measuring index to evaluate the performance of the proposed method. This index is called here as warping correction index. To estimate this index, we consider those points where the curvature of the estimated headline in warped document changes abruptly. For the single folded warped document image, three points are chosen

Table 3 Performance analysis in terms of mean local slant error percentage (γ_l)

Image source	Kim *et al.* [21]	Kil *et al.* [22]	Garai *et al.* [5]	Proposed approach
book	15.85	5.5	15.18	4.2
lamp-post	24.7	7.5	22.57	3.7
newspaper	25	4.8	17	4
noticeboard	19.5	4.2	25.47	3.3

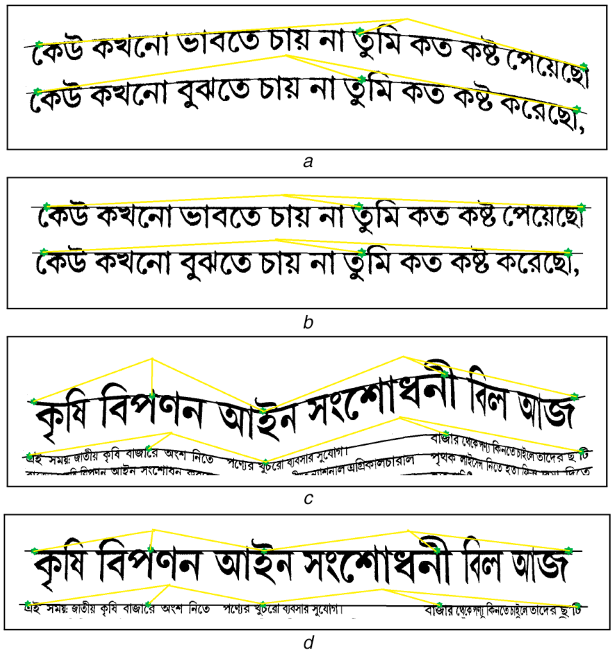


Fig. 19 *Example for estimation of warping correction index*
(a) Single fold warped document, ***(b)*** Output of the proposed method for the input (a), ***(c)*** Example of multi-fold document, ***(d)*** Proposed method for the input (c)

Table 4 Performance analysis using $C_{\rm rms}$

Image Source	$C_{\rm rms}$ ($\times 10^{-5}$)			
	Kim *et al.* [21]	Kil *et al.* [22]	Garai *et al.* [5]	Proposed approach
Book	36.2	15	1.6	0.19
Lamp-post	2.8	3.1	1.75	0.14
Newspaper	15.6	29	0.90	0.47
Noticeboard	70.2	17.1	9.33	1.5

from each line to estimate the curvature of a particular headline in the dewarped image. One point corresponds to the maximum curvature of the corresponding line in the warped image and other two points are the start and endpoints of that line (Figs. 19*a* and *b*). For multiple folded warped document images three points are chosen for each fold (Figs. 19*c* and *d*).

The selected points for a particular text line are used to estimate the curvature of that line. The *root mean square* curvature ($C_{\rm rms}$) is defined in (5), where C_{gd} represents the calculated curvature of the dewarped document. The smaller the value of $C_{\rm rms}$ better the performance of the dewarped method. Here, N is the number of text lines present in the dewarped document

$$C_{\rm rms} = \sqrt{\frac{\sum_{i=1}^{N} C_{gd}^2(i)}{N}} \tag{5}$$

The performance of our method and other methods in terms of warping correction index is presented in Table 4 and it is evident from the table that the performance of the proposed method is better than those of the others.

The morphology-based headline detection algorithm used in [5] may fail to handle complex warping. The proposed method uses a local slope-based technique to extract the headline which is more robust and can handle text line curl angle up to $\pm 45^{\rm o}$. The horizontal alignment is not accurate without proper estimation of the headline. While aligning the components, in our previous approach the translation factor is the same for an entire component, whereas in this method each column of a component has a separate translation factor. So, the alignment of text lines is more accurate in the proposed method. Due to these reasons, the accuracy measures like OCR accuracy, $c_{\rm rms}$ for our previous approach is not as good as the proposed method. Moreover, unlike our method the method proposed in [5] does not handle skew and scaling errors. Hence, the performance using MLSEP (γ_l) of the proposed method is better than [5].

5 Conclusion

In this paper, we proposed a dewarping method for alpha-syllabary scripts *Bangla,* which may be used for *Devanagari, Gurmukhi, Assamese* etc. The proposed method has been tested over various warped document images taken from different sources. The method can handle different kinds of warping like an image of a document having multiple folds, a document hanging in a notice board etc. The errors generated due to the skewness of the document are corrected using horizontal alignment and shearing-based correction. Hence, the method can handle different types of skewed as well as warped document images. The proposed approach works for curled and/or skewed images with the skew angle up to $\pm 45^{\rm o}$. Here, we have also proposed two measuring indices ($v\gamma_l$ and $C_{\rm rms}$) for directly evaluating the efficiency of a dewarping algorithm. The γ_l and $C_{\rm rms}$ are used to measure the skew correctness and the warping correctness of the proposed method, respectively. The performance of the proposed method is encouraging. We also compared the performance of our method with three recent techniques and our method outperforms the others. In future, the method can be extended to solve more complex problems like warped images with large amount of perspective distorted warped document images, or warped images with handwritten text etc.

6 References

[1] Liu, C., Zhang, Y., Wang, B., *et al.*: 'Restoring camera-captured distorted document images', *Int. J. Doc. Anal. Recognit. (IJDAR)*, 2015, **18**, (2), pp. 111–124

[2] Shafait, F.: 'Document image dewarping contest'. 2nd Int. Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 181–188

[3] Bukhari, S.S., Shafait, F., Breuel, T.M.: 'The iupr dataset of camera-captured document images'. 4th Int. Conf. on Camera-based Document Analysis and Recognition (CBDAR 2011), Beijing, China, 2011, pp. 164–171

[4] Ke, M., Zhixin, S., Bai, X., *et al.*: 'Docunet: document image unwarping via a stacked u-net'. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Utah, USA, 2018

[5] Garai, A., Biswas, S., Mandal, S., *et al.*: 'Automatic dewarping of camera captured born-digital Bangla document images'. 2017 Ninth Int. Conf. on Advances in Pattern Recognition (ICAPR), Bangalore, India, 2017, pp. 1–6

[6] Brown, M.S., Seales, W.B.: 'Image restoration of arbitrarily warped documents', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, **26**, (10), pp. 1295–1306

[7] Yamashita, A., Kawarago, A., Kaneko, T., *et al.*: 'Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system'. Proc. of the 17th Int. Conf. on Pattern Recognition, 2004, ICPR 2004, Cambridge, UK, 2004, vol. 1, pp. 482–485

[8] Fu, B., Li, W., Wu, M., *et al.*: 'A document rectification approach dealing with both perspective distortion and warping based on text flow curve fitting', *Int. J. Image. Graph.*, 2012, **12**, (1), p. 1250002

[9] Cao, H., Ding, X., Liu, C.: 'A cylindrical surface model to rectify the bound document image'. Proc. Ninth IEEE Int. Conf. on Computer Vision, Nice, France, 2003, vol. 1, pp. 228–233

[10] Zhang, L., Tan, C.L.: 'Restoringwarped document images using shape-from-shading and surface interpolation'. 18th Int. Conf. on Pattern Recognition (ICPR'06), Hong Kong, China, 2006, vol. 1, pp. 642–645

[11] Meng, G., Pan, C., Xiang, S., *et al.*: 'Metric rectification of curved document images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (4), pp. 707–722

[12] Liang, J., DeMenthon, D., Doermann, D.: 'Geometric rectification of camera-captured document images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, (4), pp. 591–605

[13] You, S., Matsushita, Y., Sinha, S., *et al.*: 'Multiview rectification of folded documents', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2018, **40**, (99), pp. 505–511

[14] He, Y., Pan, P., Xie, S., *et al.*: 'A book dewarping system by boundary-based 3d surface reconstruction'. 2013 12th Int. Conf. on Document Analysis and Recognition, Washington, D.C., USA, 2013, pp. 403–407

[15] Ezaki, H., Uchida, S., Asano, A., *et al.*: 'Dewarping of document image by global optimization'. Eighth Int. Conf. on Document Analysis and Recognition (ICDAR'05), Seoul, South Korea, 2005, vol. 1, pp. 302–306

[16] Lu, S., Tan, C.L.: 'Document flattening through grid modeling and regularization'. 18th Int. Conf. on Pattern Recognition (ICPR'06), Florida, USA, 2006, vol. 1, pp. 971–974

[17] Ulges, A., Lampert, C.H., Breuel, T.M.: 'Document image dewarping using robust estimation of curled text lines'. Eighth Int. Conf. on Document Analysis and Recognition (ICDAR'05), Seoul, South Korea, 2005, vol. 2, pp. 1001–1005

[18] Gatos, B., Pratikakis, I., Ntirogiannis, K.: 'Segmentation based recovery of arbitrarily warped document images'. Ninth Int. Conf. on Document Analysis and Recognition (ICDAR 2007), Curitiba, Brazil, 2007, vol. 2, pp. 989–993

[19] Bukhari, S.S., Shafait, F., Breuel, T.M.: 'Dewarping of document images using coupled-snakes'. Proc. of Third Int. Workshop on Camera-Based Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 34–41

[20] Stamatopoulos, N., Gatos, B., Pratikakis, I., *et al.*: 'Goal-oriented rectification of camera-based document images', *IEEE Trans. Image Process.*, 2011, **20**, (4), pp. 910–920

[21] Kim, B.S., Koo, H.I., Cho, N.I.: 'Document dewarping via text-line based optimization', *Pattern Recognit.*, 2015, **48**, (11), pp. 3600–3614

[22] Kil, T., Seo, W., Koo, H.I., *et al.*: 'Robust document image dewarping method using text-lines and line segments'. 2017 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, vol. 1, pp. 865–870

[23] Yang, P.: 'Effective geometric restoration of distorted historical document for large-scale digitisation', *IET Image Process.*, 2017, **11**, (12), pp. 841–853

[24] Pratikakis, I., Zagoris, K., Barlas, G., *et al.*: 'Icdar2017 competition on document image binarization (dibco 2017)'. 2017 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, vol. 1, pp. 1395–1403

[25] Meng, G., Yuan, K., Wu, Y., *et al.*: 'Deep networks for degraded document image binarization through pyramid reconstruction'. 2017 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, vol. 1, pp. 727–732

[26] Su, B., Lu, S., Tan, C.L.: 'Robust document image binarization technique for degraded document images', *IEEE Trans. Image Process.*, 2013, **22**, (4), pp. 1408–1417

[27] Farahmand, A., Sarrafzadeh, A., Shanbehzadeh, J.: 'Document image noises and removal methods'. Int. Multi Conf. of Engineers and Computer Scientists (IMECS 2013), Hong Kong, 2013, vol. 1, pp. 436–440

[28] Said, J.N., Cheriet, M., Suen, C.Y.: 'Dynamical morphological processing: a fast method for base line extraction'. Proc. of 13th Int. Conf. on Pattern Recognition, Vienna, Austria, 1996, vol. 2, pp. 8–12

[29] Dey, S., Mitra, B., Mukhopadhyay, J., *et al.*: 'A comparative study of margin noise removal algorithms on marnr: a margin noise dataset of document images'. 2017 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, vol. 4, pp. 35–39

[30] Fan, H., Zhu, L., Tang, Y.: 'Skew detection in document images based on rectangular active contour', *Int. J. Doc. Anal. Recognit. (IJDAR)*, 2010, **13**, (4), pp. 261–269

[31] Agrawal, M., Doermann, D.: 'Stroke-like pattern noise removal in binary document images'. 2011 Int. Conf. on Document Analysis and Recognition, Peking, China, 2011, pp. 17–21

[32] Premchaiswadi, N., Yimgnagm, S., Premchaiswadi, W.: 'A scheme for salt and pepper noise reduction and its application for ocr systems', *W. Trans. Comput*, 2010, **9**, (4), pp. 351–360

[33] Leung, C.C., Chan, K.S., Chan, H.M., *et al.*: 'A new approach for image enhancement applied to low-contrast–low-illumination ic and document images', *Pattern Recognit. Lett.*, 2005, **26**, (6), pp. 769–778

[34] Dutta, A., Garai, A., Biswas, S.: 'Segmentation of meaningful text-regions from camera captured document images'. 2018 Fifth Int. Conf. on Emerging Applications of Information Technology (EAIT), Howrah, India, 2018, pp. 1–4

[35] Shafait, F., Breuel, T.M.: 'A simple and effective approach for border noise removal from document images'. 2009 IEEE 13th Int. Multitopic Conf., Islamabad, Pakistan, 2009, pp. 1–5

[36] Shafait, F., van Beusekom, J., Keysers, D., *et al.*: 'Document cleanup using page frame detection', *Int. J. Doc. Anal. Recognit. (IJDAR)*, 2008, **11**, (2), pp. 81–96

[37] Dey, S., Mukhopadhyay, J., Sural, S., *et al.*: 'Margin noise removal from printed document images'. Proc. of the Workshop on Document Analysis and Recognition. DAR'12, New York, NY, USA, 2012, pp. 86–93

[38] Bukhari, S.S., Shafait, F., Breuel, T.M.: 'Border noise removal of camera-captured document images using page frame detection'. In, Iwamura, M., Shafait, F. (Eds.): '*Camera-based document analysis and recognition*' (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 126–137

[39] Rath, T.M., Manmatha, R.: 'Word image matching using dynamic time warping'. 2003 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Proc., 2003, vol. 2, pp. II–II

[40] Gatos, B., Louloudis, G., Stamatopoulos, N.: 'Segmentation of historical handwritten documents into text zones and text lines'. 2014 14th Int. Conf. on Frontiers in Handwriting Recognition, Crete, Greece, 2014, pp. 464–469

[41] Moysset, B., Louradour, J., Kermorvant, C., *et al.*: 'Learning text-line localization with shared and local regression neural networks'. 2016 15th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR), Shenzhen, China, 2016, pp. 1–6

[42] Vo, Q.N., Kim, S.H., Yang, H.J., *et al.*: 'Text line segmentation using a fully convolutional network in handwritten document images', *IET Image Process.*, 2017, **12**, (3), pp. 438–446

[43] Green, P.J., Silverman, B.W.: '*Nonparametric regression and generalized linear models: a roughness penalty approach*' (Taylor & Francis group: Chapman and Hall/CRC, UK, 1993)

[44] Bukhari, S.S., Shafait, F., Breuel, T.M.: 'Coupled snakelets for curled text-line segmentation from warped document images', *Int. J. Doc. Anal. Recognit. (IJDAR)*, 2013, **16**, (1), pp. 33–53

[45] Stamatopoulos, N.: 'Performance evaluation methodology for document image dewarping techniques', *IET Image Process.*, 2012, **6**, (7), pp. 738–745