



Adaptive dewarping of severely warped camera-captured document images based on document map generation

C. H. Nachappa¹ · N. Shobha Rani¹ · Peeta Basa Pati² · M. Gokulnath³

Received: 30 July 2022 / Revised: 13 December 2022 / Accepted: 22 December 2022 / Published online: 9 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Automated dewarping of camera-captured handwritten documents is a challenging research problem in Computer Vision and Pattern Recognition. Most available systems assume the shape of the camera-captured image boundaries to be anywhere between trapezoidal and octahedral, with linear distortion in areas between the boundaries for dewarping. The majority of the state-of-the-art applications successfully dewarp the simple-to-medium range geometrical distortions with partial selection of control points by a user. The proposed work implements a fully automated technique for control point detection from simple-to-complex geometrical distortions in camera-captured document images. The input image is subject to preprocessing, corner point detection, document map generation, and rendering of the de-warped document image. The proposed algorithm has been tested on five different camera-captured document datasets (one internal and four external publicly available) consisting of 958 images. Both quantitative and qualitative evaluations have been performed to test the efficacy of the proposed system. On the quantitative front, an Intersection Over Union (IoU) score of 0.92, 0.88, and 0.80 for document map generation for low-, medium-, and high-complexity datasets, respectively. Additionally, accuracies of the recognized texts, obtained from a market leading OCR engine, are utilized for quantitative comparative analysis on document images before and after the proposed enhancement. Finally, the qualitative analysis visually establishes the system's reliability by demonstrating improved readability even for severely distorted image samples.

Keywords Dewarping · Document images · Smart innovations · Computer vision model · Control point generation

1 Introduction

Affordable smartphones are ubiquitously available with high-end cameras and intelligent camera applications. The breakout of COVID-19 forced the world to think innovatively to communicate and perform day-to-day activities. These have led to an increase in the usage of smartphones to capture and share handwritten document images for many purposes. Unconstrained capture of these images comes with many challenges and poses severe issues for downstream processing to humans and digital agents. These challenges are posed by improper lighting, issues with focus while imaging and pages' non-planar placement on non-planar surfaces.

Additionally, noise is induced owing to paper quality, environmental conditions, and from poor quality sensors. Warping, however, is one of the most common issues faced during capturing photographs of handwritten document images. Moreover, dealing with the warps that occur in document images containing see-through effects for two-sided handwritten documents is particularly challenging and is

C. H. Nachappa, Peeta Basa Pati, and M. Gokulnath authors contributed equally to this work.

✉ N. Shobha Rani
n_shobharani@my.amrita.edu

C. H. Nachappa
nachappach@gmail.com

Peeta Basa Pati
bp_peeta@blr.amrita.edu

M. Gokulnath
gokulnath@my.amrita.edu

¹ Department of Computer Science, Amrita School of Computing, Mysuru, Amrita Vishwa Vidyapeetham, Mysuru, India

² Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, Bengaluru, India

³ Department of Sciences, Amrita School of Physical Sciences, Mysuru, Amrita Vishwa Vidyapeetham, Mysuru, India

highly complex. Current available technologies can estimate control points semi-automatedly for the linear distortions in warped camera-captured documents. In this work, we propose a fully automatic control point detection method for the warped camera-captured document images with various types of geometrical distortions.

Warping may occur with various complexity levels ranging from low to high. Figure 1 depicts sample images with low-, medium-, and high-complexity warps in (a), (b), and (c), respectively. Without an effective automated correction to such distortions, the images are either recaptured, or manual effort is employed to make corrections. Failure to correct these distortions leads to document images which are illegible and non-digitizable. This adversely affects the utility of such document images, thereby rendering them useless. In this work, we aim to provide a solution to perform image dewarping automatically for low-/medium-/high-complex warped documents, so as to improve their readability.

Image warping is a common problem in capturing images of paper documents with handheld mobile devices [1]. Improper orientation of paper documents, wrong angle of capturing photos, wrong settings of the camera in terms of distance, brightness, and contrast, and improper illumination are the significant factors that often lead to warped photos [2]. Additional challenges, such as blur, uneven shading, shadows, and curved text lines, complicate the legibility and readability of documents. All these distortions also adversely impact the optical character recognition (OCR) performance [3, 4]. Enhancements of these distorted images with dewarping and other techniques are highly likely to improve the legibility, readability, and OCR accuracy obtained during the recognition of document images.

The approaches to dewarping take one of the following paths: (i) classical image processing with spatial and spectral techniques, (ii) machine learning-based techniques, and (iii) deep learning-based techniques. Most of these works consider typewritten/printed documents for the design of their algorithms as well as their performance evaluation. Works on the dewarping of handwritten documents are limited and mainly deal with deep learning techniques. Additionally, the warp handled in reported works is restricted to the low or medium complexity categories as per our complexity definitions in Sect. 3.

In this paper, we present our work on two novelties: (i) a generic complexity definition for camera-captured document images and (ii) an automated algorithm that detects unlimited control points to define the boundary of the page in the image leading to effective dewarping of the image. We prove, both quantitatively and qualitatively, that our algorithm generates superior results to existing algorithms as well as works equally effectively for both handwritten and type written images. This is demonstrated with results obtained on both internal and published datasets. The rest of the paper

is organized as follows: Sect. 2 discusses the prior works from a survey of the literature, a description of the proposed methodology is presented in Sect. 4, analysis of experimental results in Sect. 5, and finally, the conclusion in Sect. 6.

2 Prior work

Dewarping is an essential preprocessing step to enhance images for text recognition in printed or handwritten documents. Many researchers have worked on the correction of warps for printed documents. Initial works on the dewarping of document images are based on regression-based curled text line correction [5] and 3D shape reconstruction with specialized hardware [6]. Though the primary focus of dewarping documents is to align the document layout, recognition of text and graphics [7] in documents caught much of the attention.

Some of the early works employed methods such as model-based dewarping [8–13], integrals of cubic polynomial curves [14], text line-based optimization [15], and content-independent dewarping [16]. Subsequently, with availability of deep learning technologies [17–19], many researchers have adapted them for the image dewarping task. The former methods are reliable and deliver good performance for a specific type of document with limited degree of warp while requiring a small database of images. Additionally, the classical model-based or parameter-based holistic approaches deliver reasonable to superior output for a specific category of documents. Deep learning methods, on the other hand, require huge collections of ground truth images and metadata for learning the dewarping process. The performance of deep learning methods, however, is well suited to a wide range of document images.

2.1 Image analysis techniques for printed document analysis

Plenty of research work has been reported based on image analysis methods. These methods are simple and sensitive to thresholding parameters. However, their outcomes mostly are better than model-based approaches. Wu and Agam [7] proposed a two-pass image warping algorithm that uses the texture of a document image to dewarp the printed text documents. The method depends on user input, is language independent, and handles documents with multiple fonts, math notations, and graphics. Their results demonstrate the algorithm's robustness toward images with low-complexity warps. Ulges et al. [20] reported on a method for removal of perspective and page curl distortions in warped printed documents. The method's processing outcome depends on user input for warp corrections. The algorithm validation is conducted with images with low complexity of distortion

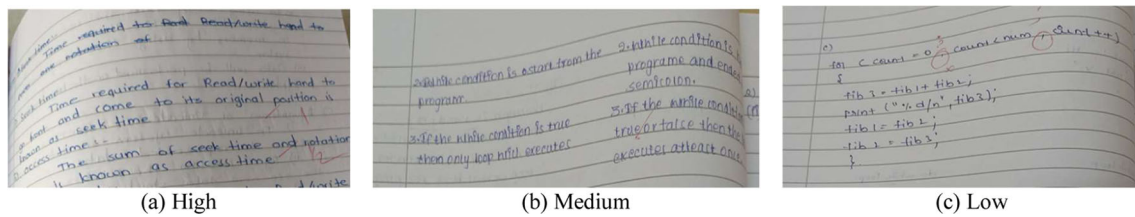


Fig. 1 Samples of warped documents of varying complexity

and text line curls. Ezaki et al. [21] suggested a technique that estimates the warp line-wise by fitting splines to the text lines. Experimental results from this technique demonstrate its effectiveness in printed document images with local irregularities such as formulae, short text lines, and figures. In a subsequent work on dewarping of the document image by [22], text dewarping is carried out by estimating text lines using RANSAC approximation. An attempt is made to correct the nonlinear perspective distortion and page curls in text lines. Datasets employed for experiments include a printed document of pages, and the textbooks are captured using a handheld camera. In a different work, an algorithm is proposed by [23] to compose a geometrically dewarped and visually enhanced image from two views of document images taken by a digital camera at different angles. Document dewarped surface is estimated based on correspondence points between two images by estimating surface and camera matrices via structure reconstruction, 3-D projection analysis, and random sample consensus-based curve fitting with the cylindrical surface model. Image mosaicking is applied to enhance the image by stitching and blending them. Ghods et al. [24] proposed a technique for document dewarping using the Microsoft Kinect camera sensor. The proposed method is based on a 3D model of the warped document obtained using a Kinetic sensor and an external high-resolution camera. A voxelized grid approach is applied with a triangulation mesh generated using the greedy fast triangulation method. Experiments prove that method can correct distortions of different kinds irrespective of the type of content used for dewarping. In another work, dewarping uses text line-based optimization techniques [15]. The method has thrown out lots of limitations of static dewarping methods. A discrete representation of text lines and blocks is used to design a cost function to remove perspective distortion and page curls. Experiments show that the method works well for standard printed document datasets with various layouts and curved surfaces. In the work reported by [25], a multi-stage dewarping algorithm is designed to perform coarse and fine adjustments in warped documents that works based on the regions of interest. Assessing the quality of the dewarped images is carried out without the need for ground truth images based on a page-level decision-making process with a significant time penalty. The results are analyzed

thoroughly with the CBDAR2007/IUPR2011 dataset, which includes different document layouts and types of warps, and the results are reasonably good. Simon and Tabbone [26] proposed a generic approach based on vanishing points to reconstruct the 3D shape of document pages. Experimental results show that the proposed method is robust to various text and non-text block distortions.

Wagdy et al. [27] exploited the distorted and undistorted connected components of a document image to handle multi-warping. Interactive input is taken on pairs of control points for a connected component. A mapping function is generated to map the component between the input image and its corresponding ground truth. A technique by [28] based on checkerboard pattern and geometric transformation is evaluated on CBDAR 2007 printed datasets with promising results. Xie et al. [29] proposed a geometric transformation-based dewarping method that exploits the control points to estimate the warp. These control points are generated with the help of a semantic encoder employed on the image content. An article by [30] summarizes the multiple works reported in the document image dewarping contest held by 2nd CBDAR in 2007. The contest utilized a dataset of 102 documents captured with a handheld camera and dealt with various document analysis tasks such as text line, text zone, and comparison of recognition against ASCII text ground truths. Stamatopoulos et al. [31] presented a comprehensive performance evaluation method to demonstrate the performance of dewarping techniques in a concise quantitative manner. The evaluation measure considers the deviation of the dewarped text lines from a straight horizontal reference and expressed by the integral over the text line curves. A point-to-point matching procedure that corresponds between the manually marked warped document image and the dewarping counterpart is also devised. However, the selection of correspondence points is still made interactively. Garai et al. [32] proposed an automatic method for dewarping of comic document images which rectifies the warping in both horizontal and vertical directions. The folds present in the image are considered as a crucial challenge which are dealt with the help of morphological operations and cubic smoothing splines. Though the work provides good outcomes toward the comic documents, the experimentations are not conducted on

images that consist of works in multiple directions or hand-written images with multiple challenges.

2.2 Image analysis-based techniques—handwritten document analysis

The complexity of analyzing (layout analysis, segmentation, and recognition) handwritten documents is much high compared to printed documents. Though dewarping is an enhancement method, the outcome of dewarping shows its impact on handwriting recognition outcomes. Very few works are reported on dewarping documents based on image analysis techniques for handwritten document images.

Bolelli [33] proposed a dewarping approach with the help of a transformation model which maps the projection of a curved surface to a 2D rectangular area. The method exploits the possibilities of applying dewarping to document images to handwritten and typewritten text documents, and projection of the curved surface is extracted. Though the method produces promising outcomes, an assumption of precompiled surface projections is employed to estimate the structure. Experiments are demonstrated using limited severity surface projections, and perspective distortions do not exist in most documents. In work by [34], a document image workflow system is proposed for scaling the handwritten student assessments of universities. Few image processing algorithms are used to improve the quality, readability, and scaling with simple transformations. Around 101 samples of assessments of very few distortions and skew are considered for experimentation. In work by [35], dewarping and deskewing are performed based on various color band details of CMY using a threshold. Removal of geometrical and perspective distortion using centroid and mid-point of bounding box height and connected components is accomplished. Experimentations are conducted on Kannada handwritten documents with limited severity of distortions and text curls. Garai et al. [36] proposed a technique to correct the warps in Bangla handwritten documents. A dataset is contributed by [37] covering a wide variety of historical Arabic documents containing clean information and primary and homogeneous-page layouts. The tests are executed on printed and handwritten documents from some imperative libraries. Ground truth is created using the Aletheia tool.

2.3 Deep learning techniques—printed document analysis

With cost-effective high-end computing power availability, deep learning architectures and techniques are employed for solving various kinds of problems. Researchers have employed these algorithms to solve the warping problem in document images.

Ramanna et al. [17] devised a dewarping method by employing conditional generative adversarial networks (CGAN). The network is trained on the UW3 dataset (with low severity of page curls and distortions) for an image-to-image translation. The trained network helps to remove text curls and geometric distortion of current and historical documents with images and graphics. Xie et al. [19] devised a fully convolutional network (FCN) to rectify distorted document images and remove other objects available in the background by estimating pixel-wise displacements. Synthesized distorted documents are utilized for training the FCN for displacement estimation. The local smooth constraint (LSC) technique is used in training regularization for control of smoothness in the displacement of pixels. Experiments prove that the method is capable of removing various geometric distortions effectively on printed document pages. The work by [38] introduces document image transformer (DocTr) framework to address the issue of geometry and illumination distortion of the warped document images. DocTr model consists of a geometric unwarping transformer and an illumination correction transformer. Bandyopadhyay et al. [39] proposed a supervised gated and bifurcated stacked U-Net module to predict a dewarping grid. The network is trained on synthetically warped document images while its performance is evaluated on the DocUNet dataset. Bandyopadhyay et al. [39] performed dewarping using a gated and bifurcated stacked network that employs residual paths to enhance the flow of information within the network. In this later work, the weighted loss metric is used to enable greater focus on boundaries which indicates the dewarp quality in the results obtained toward the DocUNet dataset.

2.4 Deep learning techniques—handwritten document analysis

Deep learning techniques are also employed to handle the warp issue in handwritten document images. An overview of a few works in this field is presented here.

Ma et al. [40] developed and trained U-Net with intermediate supervision to map a distorted image to its rectified version. With reduced ground truth deformation details through an interactive process, a synthetic dataset of 100,000 images is created by warping non-distorted document images for this training. Both qualitative and quantitative evaluation is performed on the datasets with moderate-to-severe perspective distortions. Dutta et al. [41] proposed a method to segment text lines from documents captured with a flatbed scanner and camera. The dataset contains images that are heavily warped and contains both printed and handwritten text. Semantic segmentation is performed using a multi-scale convolutional neural network. The performance is evaluated with various publicly available datasets, including ICDAR, Alireza, IUPR, CBDAR, Tobacco-800, IAM, and a synthetic

dataset. An encoder–decoder architecture is proposed by [42] based on vision transformers to enhance both machine-printed and handwritten document images. Experiments are conducted on several DIBCO benchmarks to validate the model. However, this reported work focuses on enhancing the input image rather than warping corrections.

Based on the above survey, we observe that most of the research works deal with image analysis techniques toward printed datasets with only very few attempts based on deep learning techniques. The rectification of text block curls due to document warping is seldom addressed for printed or handwritten documents. To the best of our knowledge, the research challenges of dewarping handwritten document images with varying levels of complexity in terms of perspective distortion and efforts toward improvising recognition performance are not addressed in the literature. The various benchmark datasets used for experimentations are mostly with limited severity in terms of perspective distortions and text block curls. Therefore, we have created a dataset that involves complex warp distortions. We also propose an algorithm that handles the warp complexities equally effectively for this dataset and the earlier datasets. From the review of dewarping techniques toward warped camera-captured document images, it is noticed that,

- Control point detection is carried out by applying various geometrical transformations that depend on the choice of arbitrary boundary points on the document boundaries.
- More than four control points are chosen in most cases to perform dewarping.
- The datasets employed are from collections of CBDAR competitions in past works.
- A few works consider smartphone images with simple-to-medium-level warps.
- Works addressing the challenges of recognition regarding post-dewarping tasks seldom exist.

3 Dataset image categorization

For experimentation and evaluation purposes, we have categorized the dataset images into three classes, namely—low (L), medium (M), and high (H), based on the complexities involved in the image. The two primary factors used for categorization are the number of available image corner points and the nature of warps present in the image.

The low-complexity input image class primarily deals with no to low warps. These images are mostly trapezoidal with reduced or no localized curvatures (due to the flat surface of the photographed paper). Four control points connected by straight lines can easily describe these images. Skew and orientation may also be present in the images, as graphical

content outside the page image boundary may be found. The following parameters can be considered to categorize images into low complexity.

- Image can be described using four control points.
- Four straight lines connect the four control points forming a rectangle or trapezoid.
- The skew present for various text lines is the same or is in close range.
- Text lines become horizontal after skew/orientation correction—no curvature is present in text lines.
- The contrast between the page background and outside page areas is high.
- Low illumination variation in the page image.

The medium complexity class of input images deals with low warps. These images require more than four control points to describe their boundary. Each boundary line may contain curvatures representing up to three control points (instead of two) connected by straight lines. Thus, the boundary of the page from the medium complex image class may have up to six control points. Small quantities of localized curvatures may be present. Skew and orientation may also be present in the images, as graphical content outside the page image boundary may be found. The following parameters can be considered to categorize images into medium complexity.

- An image consists of visually perceivable control points in the range of $4 \leq C < 6$, where C denotes the control points.
- Absence of straight boundary lines that can directly connect any two given control points.
- Presence of slight warps across the page surface.
- Presence of slight curvature in text lines.

The high-complexity class of input images deals with greater warps. The boundaries of these page images may require multiple control points, with each boundary line containing multiple curvatures of both concave and convex types. The nature and quantity of localized curvatures in different page regions may vary. The surface of the paper would contain multiple curvatures, which reflect on the boundary lines as well as curvatures in the text lines of the image. Skew and orientation may also be present in the images, as graphical content outside the page image boundary may be found. The following parameters can be considered to categorize images into complex complexity.

- An image consists of visually perceivable control points which exceed the range of $4 \leq C < 6$, where C denotes the set of control points.
- Absence of straight boundary lines that can directly connect any two control points.

- Significant warps across page surfaces severely affect textual readability and recognition.
- Presence of significant curvature in text lines.
- The contrast between the page background and outside page areas is low.

4 Proposed methodology

The primary focus of the proposed method is to enhance the state-of-the-art functionality of image dewarping systems to deal with low-/medium-/high-complexity warps. In this work, an efficient methodology for warped handwritten documents is proposed, which reduces distortion in text blocks enabling text legibility. In the proposed work, image dewarping is carried out in multiple phases, as shown in Fig. 2.

Initially, the input image is subject to preprocessing protocol for pre-enhancements, followed by the corner point detection for document region detection and boundary extraction. The document map is generated based on the control point detection outcome. Finally, image rendering is done using an isometric mesh-based technique based on a document map to obtain a dewarped image.

4.1 Preprocessing

Given a noisy image W^0 (superscript denotes the original image) W with dimensions of an image being $M \times N$ representing M rows and N columns. Pre-enhancement process of W^0 precedes with conversion to RGB color space of noisy image W^0 . For an RGB image, W^0 composed of W^r , W^g , and W^b denoting the red, green, and blue channels, a green channel image variant W^g is retrieved via function F_1 (1).

$$F_1 : W^0 \rightarrow W^g \quad (1)$$

W^g is a green channel with intensity values I_{\min} to I_{\max} as minimum to maximum intensity level. For a green channel image W^g , that is obtained from a noisy image W^0 with Gaussian noise [43] component n , noise-corrected image W can be estimated by (2).

$$W \leftarrow W^g - n \quad (2)$$

The proposed method retrieves noise-free image W by applying a linear Weiner filter [44] function F_2 for distortion correction (3). To suppress the noise component in independent identically distributed regions in an image, function F_2 preserves the gradient details and enhances visual characteristics that help improve text legibility.

$$F_2 : W^g \rightarrow W \quad (3)$$

From (3), function F_2 effectively removes the noise component n through the smoothing of distinct pixels in W^g . Though denoising is effective with function F_2 , the implication of (2) results in a smoothing effect producing a denoised blurry image W . Therefore, a deblurring function F_3 is applied on W to reduce the smoothing effect as in (4).

$$F_3 : W \rightarrow I_0 \quad (4)$$

Given a deblurred image I_0 , a blind deconvolution function [45] F_4 reduces the blurring effect with the help of point spread function f of strength k with 15 as its size. A larger kernel size would result in loss of information, whereas minimal values would show no blur reduction. Deblurring function F_3 reduces the blurring effect by a factor of b from W , as indicated in (5). Additionally, in the subsequent step, morphological bridging operation [46] via function F_4 to bridge the gaps in edges in W is applied as in (6).

$$I_0 \leftarrow W - b \quad (5)$$

$$F_4 : I_0 \rightarrow I_1 \quad (6)$$

Image I_1 indicates the morphologically processed image, which would positively impact the recognition performance of handwritten OCR. Given a morphologically processed image I_1 , which is subject to adaptive thresholding [47] function F_5 that categorizes the pixels into binary values 0 and 1 through a local sliding window approach producing a binary image I as given by (7).

$$F_5 : I_1 \rightarrow I \quad (7)$$

The image I is further proceeded to corner point detection in the subsequent stage. Figures 3 and 4 show the results of preprocessing and its visual implications for the enhancement of the original image.

4.2 Control point detection

The detection of control points plays a crucial role in the dewarping of documents. Hough transformation in combination with mean-based control point selection is employed in the proposed method. The proposed system can handle images with more control points by approximating the varying range to four boundary control points. Given a noise-corrected image I , detection of boundary control points C_1 , C_2 , C_3 , and C_4 is carried out using linear Hough transform H via the Hough lines technique. H performs the line detection by varying the line approximately lying at an angle of 90° from one to another. Each line is aligned in perpendicular directions as detected via Hough transform H .

Fig. 2 Workflow of image dewarping algorithm

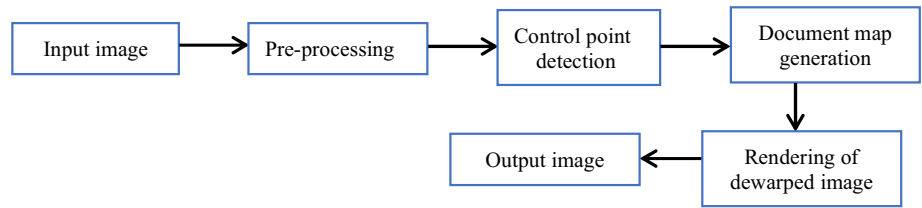


Fig. 3 Workflow of the preprocessing stage

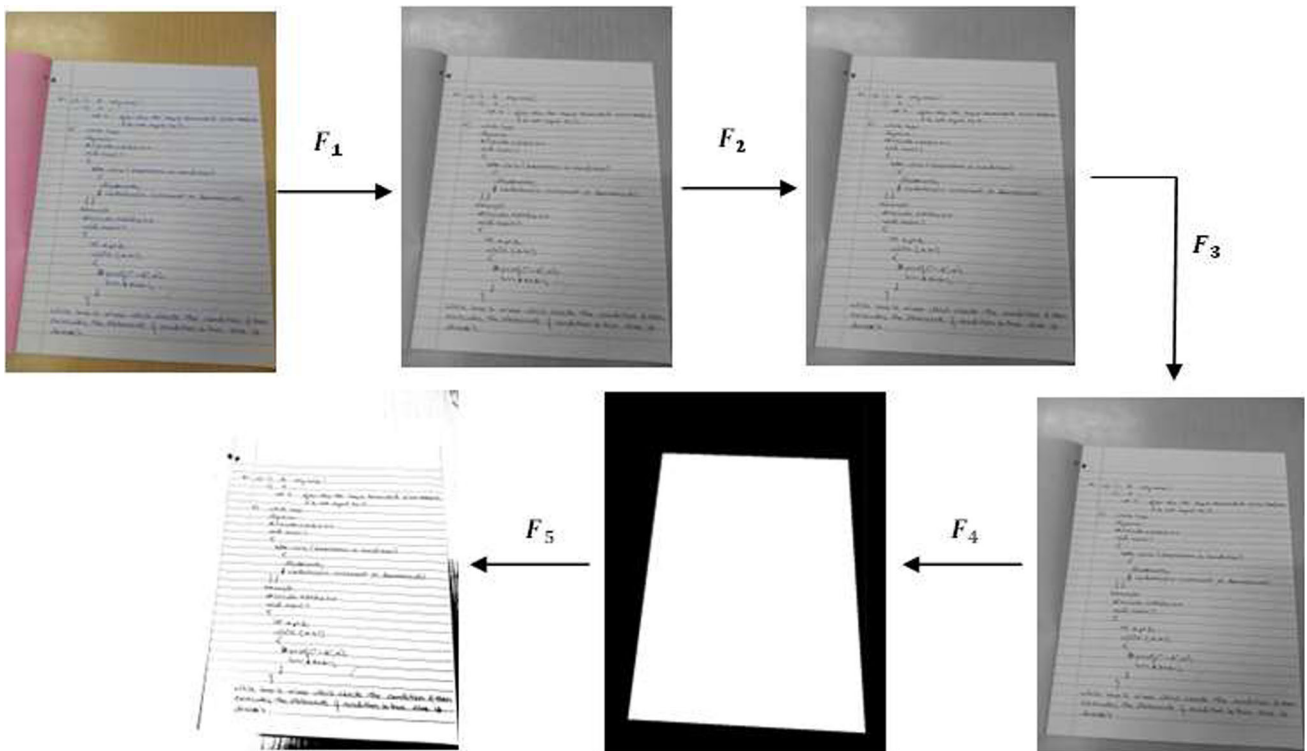
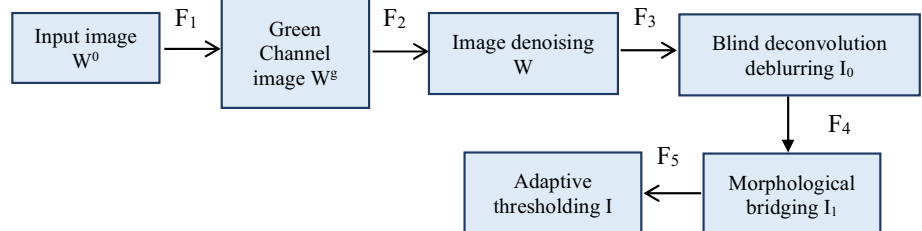


Fig. 4 Results of applying denoising functions $F_1, F_2, F_3, F_4,$ and F_5

Let L be the set of all boundary lines. Let C be the set of all points of intersection of line L . Using the following procedure, let us identify four subsets of C , say $C_1, C_2, C_3,$ and C_4 . First, let us randomly choose four points from C , say $c_1, c_2, c_3,$ and c_4 . Next, calculate the Euclidean distance between all the points in C with $c_1, c_2, c_3,$ and c_4 . Add each point p to C_i , for $1 \leq i \leq 4$, whenever p is the closest point to c_i when compared with other c_j 's, where $j \neq i$. Later,

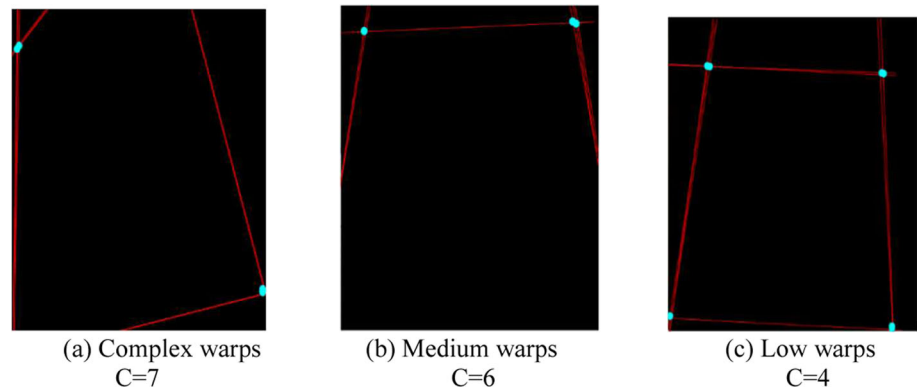
update the value of c_i for $1 \leq i \leq 4$, as given in (8).

$$c_i = (\overline{p_x}, \overline{p_y}) \tag{8}$$

where $\overline{p_x} = \frac{1}{|C_i|} \sum_{(p_x, p_y) \in C_i} p_x$ and $\overline{p_y} = \frac{1}{|C_i|} \sum_{(p_x, p_y) \in C_i} p_y$

Let us remove all the C_i points for $1 \leq i \leq 4$. Then, re-calculate the Euclidean distance between all the points

Fig. 5 Boundary control point detection



in C with the updated c'_j s, for $1 \leq i \leq 4$, and repeat the process. That is, add each point p to C_i , for $1 \leq i \leq 4$, whenever p is the closest point to the updated c_i when compared with other updated c'_j s, where $j \neq i$. This process should be repeated until the points in the C_i , for $1 \leq i \leq 4$, remain unchanged. After this process, the final points c_1 , c_2 , c_3 , and c_4 are considered for boundary detection in the subsequent stage. Figure 5 shows the results of applying boundary control point detection on denoised images toward severe/moderate/slightly warped documents.

The corner points do not have much importance in the proposed algorithm, it is the control points which help trace the boundary of the document. A boundary may be severely curved due to high warp. In such a situation, along with corner points there will be many more control points to define the boundary. The border of the page gets defined with small line segments joining two consecutive control points. Thus, no text in the page goes out of the boundary lines of the page. However, in a situation where the text goes out of the visual areas of the camera, the algorithm will not be able to reconstruct the missing areas.

4.3 Document map generation

Control points c_1 , c_2 , c_3 , and c_4 are quadrilateral control points given as input to the boundary detection process. These control points constitute the area of the document image, and the lines are composed from c_1 to c_2 for the top horizontal document boundary, c_1 to c_3 for the left vertical document boundary, and followed by c_2 to c_4 for the right vertical document boundary, and finally c_3 to c_4 to obtain the bottom horizontal boundary. The boundary control points are used as a reference to remove the background details, i.e., the image region which comprises the document area. If (x_i, y_j) represents a pixel lying outside the limits of control points c_1 , c_2 , c_3 , and c_4 , then (x_i, y_j) will be assigned as 0; otherwise, 1. This process of evaluating the pixel positions will be continued from $i = 1, 2, 3 \dots M$, and $j = 1, 2, 3 \dots N$ leading

to the formation of a document map M which serves as a reference image for the dewarping process.

Document map M is subjected to rotation correction filter R , and the referenced region R in the boundary detected image compensates for the loss of details of curled text blocks. Applying the rotation correction filter R will suffice the loss of detail in curled text blocks. Therefore, this step is essential in the dewarping process, especially for documents with severe/moderate warps with curled text blocks. Figure 6 shows the outcome of document map generation for the image in Fig. 5a–c.

4.4 Image rendering

The resampling of pixels with the help of a 2D mesh grid involves a spatial transformation of interior points bounded by control points c_1 , c_2 , c_3 , and c_4 . This image dewarping involves mapping the four corners of the non-rectangular document region into a rectangular region. To manipulate the non-rectangular shape of the document region from the document map M , the boundary control points c_1 , c_2 , c_3 , and c_4 are stretched with the help of a 2D mesh grid resulting in a rectangular shape. Mesh grid acts as a 2D resampling grid to rectify the distortions such as curls in text and curvy document surface and edges detected by control points.

Let the document region in document map M be denoted as D and the non-document region as D' . Initially, a 2D mesh grid is created on the document region D and which is then mapped with the help of a mapping function g by the warps present in an image I . The mapping function g resamples the pixels in D of map M to correct the curls in text blocks. Mapping function g replicates the transformation onto the mesh in M onto the preprocessed image I concerning the region D in document map M . Perspective transformation p is employed by mesh grid function g to perform the rendering of a transformed image I into a flattened rectangular form resulting in dewarped image G . Figure 7 demonstrates the workflow of image dewarping process.

Fig. 6 Document map generation for Fig. 5a–c

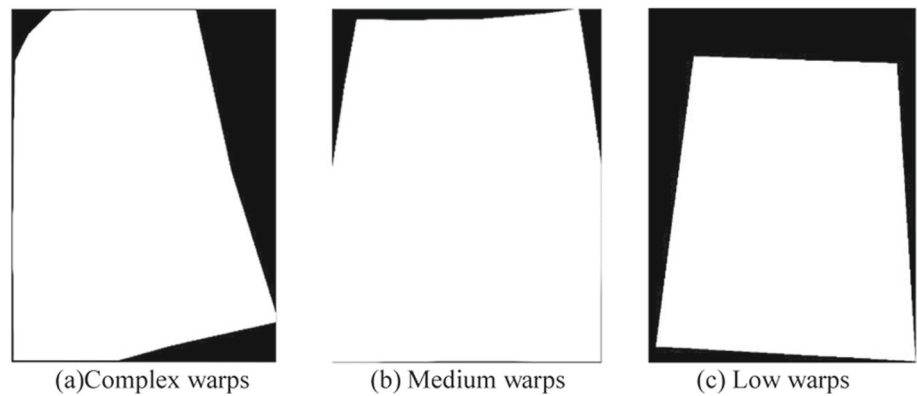


Fig. 7 Image dewarping workflow

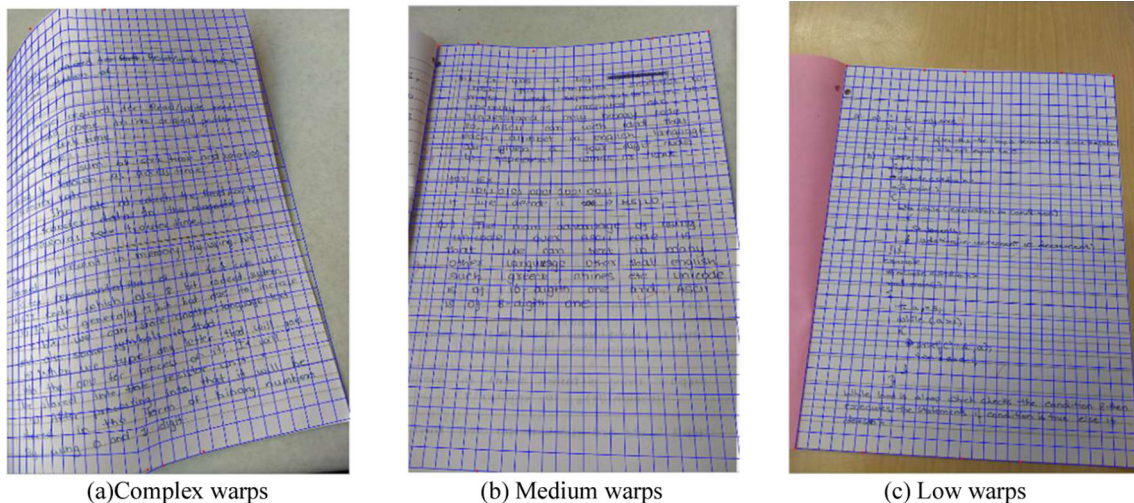
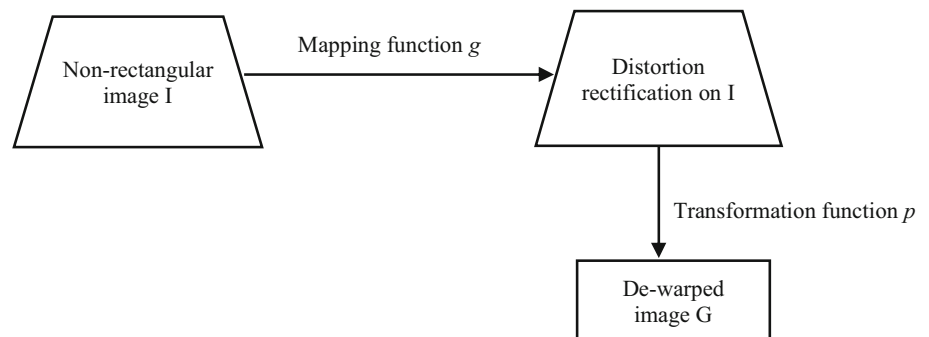


Fig. 8 2D grid mesh mapping

A 2D isometric mesh is created on the surface of the document image using a generally cylindrical surface to map the structure of the warps present in the image. The horizontal lines are estimated based on vanishing points and document boundaries, whereas vertical lines are estimated based on aspect ratio [48]. The warps are corrected using the 2D mesh as a reference, and transformations are applied to the mesh reference to render it into a rectangular shape.

The applied transformations are then mapped into the actual image. Finally, the image is rendered into a rectangular form using perspective transformations. Figure 8 presents results of 2D grid mesh mapping on the document area.

Figure 9 illustrates the workflow from Sect. 4.1 preprocessing to Sect. 4.4 image rendering with each stage of document transformation with the help of a test image.

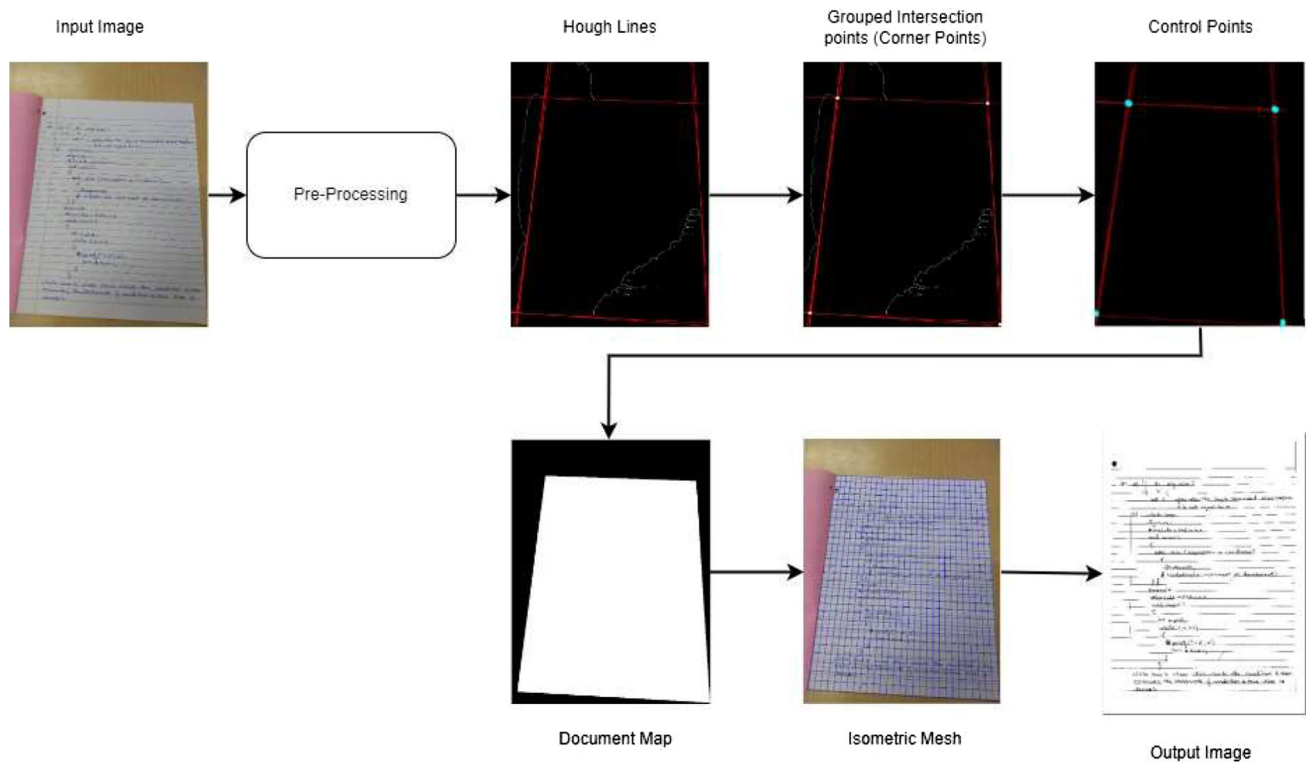


Fig. 9 Illustration of steps in the transformation of warped to dewarped image using the proposed method

5 Results and discussion

5.1 Dataset creation and acquisition

We have created a robust dataset by capturing real-time images of handwritten answer scripts of students. These answer scripts are in the form of booklets which helps us address the warping problem commonly seen in images of thick bound book pages or rolled papers. Moreover, to increase the complexities involved in the dataset, we have also altered the camera perspectives resulting in complex and geometric distortions. The dataset comprises complex images with distortion, low contrasts, shadows, warps, and complex backgrounds, which can be encountered in a real-time day-to-day scenario. For experimentation and evaluation purposes, we have categorized the datasets into low, medium, and high based on the complexities involved in the image. The datasets are made publicly available via the GitHub repository URL provided [49].

The metadata of the dataset is generated with the help of the CVAT (Computer Vision Annotation Tool) annotation tool, a free and open-source tool used for labeling data in Computer Vision and Image Processing-based algorithms. The images were firstly annotated with the PgBoundary label, which represents the document image's boundary, and also with PgLine labels, which represent the curve lines present in an image containing warps. Four equidistant curve lines

(PgLine) placed in four sections of the image were used to roughly map the shape of the warp present in the image. Finally, the generated annotations were used for the ground truth creation. Along with creating our dataset, we also collected dataset images from two different sources belonging to different institutions. However, these images do not comprise warps but have complexities in terms of contrasts and distortions. These datasets helped us understand the influence of warps present in an image based on the optical character recognition rate. The dataset details are presented in Table 1.

The system was also evaluated on two standard datasets—the CBDAR 2007 dewarping contest dataset [30] and the DDFCN120 dataset [19]. The CBDAR 2007 dewarping contest dataset has around 102 binarized images, and the DDFCN120 dataset has 120 images. The nature of both the datasets is printed and camera captured with low-, medium-, and high-complexity warps.

5.2 Qualitative analysis

Various aspects of qualitative analysis were performed to ascertain the enhancement accomplished with the proposed method. This analysis involved inspection of the visual quality and legibility of textual content in the images, the corrections to the skew, the orientation of the page images, and the curvatures of lines of text in the document. Various

Table 1 Characteristics of the collected datasets

Source	Dataset name	Image count	Complexity type	Ground truth count	Image resolution (approx.)	Additional remarks/nature of the dataset images
Camera-captured handwritten answer scripts	AVV-CD234	234	Low Medium High	234	3024 × 4032 pixels	Shadows, complex background, perspective, and geometric distortions
Camera-captured handwritten notes	CCHD124	124	Low Medium	15	1140 × 1667 pixels	Low contrasts
Camera captured	RTHD600	600	Low Medium	15	1185 × 1584 pixels	Slight blurs and uneven illumination
[30]	CBDAR 2007	102	Low Medium High	9	600 × 800 pixels	Camera-captured printed documents
[19]	DDFCN120	120	Low Medium High	15	960 × 1024 pixels	Camera-captured printed documents

sample images were taken from different complexity categories. They were tested for the effectiveness of two reported techniques, OCV DocScan and GBS U-Net, reported in the literature alongside the proposed technique. OCV DocScan [50] is designed on an architecture that emphasizes preprocessing rather than a robust dewarping architectural core. GBS U-Net [39] uses a U-Net based approach with a dewarping mechanism. Figure 10 provides the sample inputs and the enhanced outputs obtained from the three techniques for comparative visualization and analysis. The following subsections detail our observations on the outputs of these images.

5.2.1 Low complexity

The low-complexity class of input images mainly deals with no to low warps (refer to Sect. 3). These images are primarily trapezoidal with reduced or no localized curvatures (due to the flat surface of the photographed paper). Four control points connected by straight lines can easily describe these images. Skew and orientation may also be present in the images and graphical content outside the page image boundary may be found.

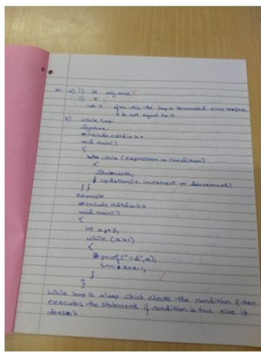
When such images are enhanced, the proposed algorithm can detect the paper's boundaries along with the skew and orientation that may be present. The boundary detection can employ the isometric mesh technique and dewarp the trapezoidal image to near perfection, as shown in Fig. 10d. Localized adaptive thresholding can binarize the image correctly, retaining good visual quality and taking care of any illumination variations in the input image. Our denoising algorithm can remove the noise present to generate output

with more excellent legibility. Additionally, the output generated is bound to the regions of textual content, and all the outside regions are removed. Close examination of many outputs from these image classes informs us that none of the output images had the textual regions being cut out during such segmentation. A comparison of the outputs of our algorithm with the other two reported methods reveals that our algorithm is superior to the others. It may be noted from Fig. 10b that while good visual output is generated from OCV DocScan, the output page image also contains portions of non-textual regions. GBS U-Net has created more warp distortions to its output than was initially present. This may be because GBS U-Net is designed to deal with cylindrical distortions and assumes its presence in all input images. Since the input image had no such distortion, the application of the algorithm affects adversely.

In summary, noise reduction and deblurring processes improved the image significantly, while boundary detection helped remove the background regions. Thus, we may infer that images belonging to a low-complexity class are almost equally enhanced by OCV DocScan and the proposed method. At the same time, GBS U-Net induces more distortion in the output initially present.

5.2.2 Medium complexity

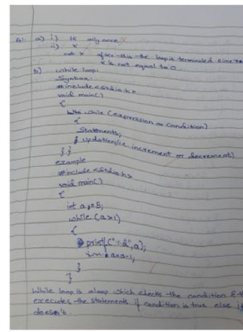
The medium complexity class of input images deals with medium warps (refer to Sect. 3). These images require more than four control points to describe their boundary. Each boundary line may contain curvatures representing up to three control points (instead of two) connected by straight



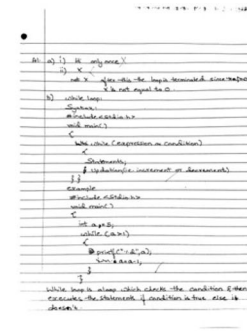
(a) Sample-1



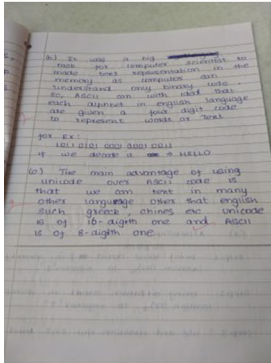
(b) OCV DocScan - Sample 1



(c) GBS U-Net - Sample 1



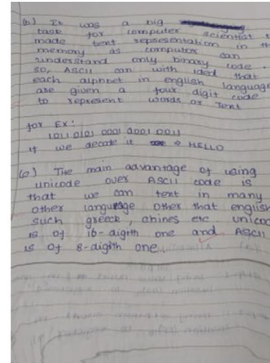
(d) Proposed Method - Sample 1



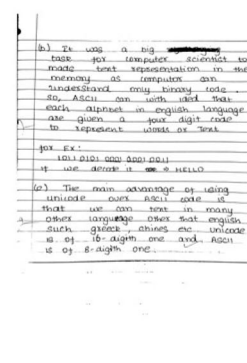
(e) Sample-2



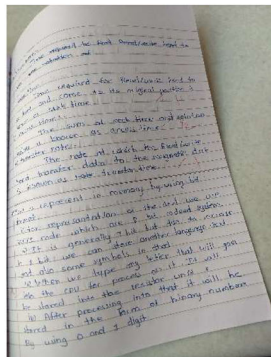
(f) OCV DocScan - Sample 2



(g) GBS U-Net - Sample 2



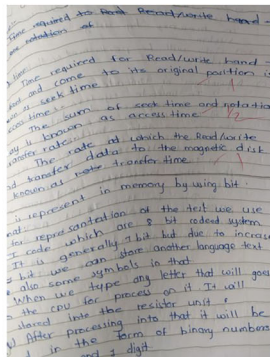
(h) Proposed Method - Sample 2



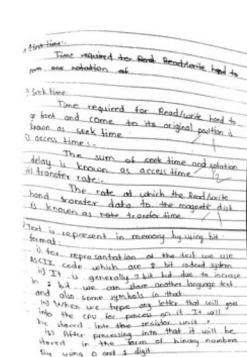
(i) Sample-3



(j) OCV DocScan - Sample 3



(k) GBSU-Net- Sample 3



(l) Proposed Method -Sample3

Fig. 10 Comparative visual analysis of the outputs generated for sample input images of low- (row 1), medium- (row 2), and high-complexity class of images (row 3). The input images are presented in column 1;

outputs generated by OCV DocScan & GBS U-Net are presented in columns 2 and 3, respectively, while the outputs of the proposed algorithm are presented in column 4

lines. Thus, the boundary of the page from the medium complex image class may have up to eight control points. Small quantities of localized curvatures may be present. Skew and orientation may also be present in the images, as graphical content outside the page image boundary may be found. Figure 10e represents a sample input image subjected to enhancement by the three algorithms. It may be noted in this sample image that the text lines present at the top of the page carry mild curvature and are not straight.

The output of the proposed algorithm may be observed in Fig. 10h. It may be noted that the output is corrected for skew, orientation, and non-page regions. The isometric mesh-based dewarping technique straightens the curvatures present in the text lines on top. Noise reduction and deblurring processes have improved the image and the document boundary was detected in congruence with the expectation. It efficiently removes the background regions that are not part of the region of interest without losing the textual regions in the image. The output generated by OCV DocScan (present

as Fig. 10f) is very distorted. Not only are portions of the page image cropped out, but high curvature is induced into the output page. While the algorithm has done an excellent job in binarization, the warp distortion induced is more significant than the corrections brought about by the algorithm.

GBS U-Net (output in Fig. 10g) has performed well due to mild cylindrical distortion. It may be observed that while the top text lines are better straightened, the lower ones, void of any curvature in the original image, have some curvature induced into them. The algorithm has detected the page boundaries well and removed the non-page areas from the image.

The proposed technique best enhances images in the medium complexity class. The visual quality of outputs from all three tested algorithms testify to this claim. This impression is further reinforced with better recognition accuracy observed in Table 4.

5.2.3 High complexity

The high-complexity class of input images deals with greater warps (refer to Sect. 3). The boundaries of these page images may require multiple control points, with each boundary line containing multiple curvatures of both concave and convex types. The nature and quantity of localized curvatures in different page regions may vary. The paper's surface would contain multiple curvatures, which reflect on the boundary lines and curvatures in the text lines of the image. Skew and orientation may also be present in the images, as graphical content outside the page image boundary may be found. Figure 10i represents a sample input image subjected to enhancement by the three algorithms.

The proposed algorithm has the best visual quality of the output images (Fig. 10l) from all three algorithms. The boundaries of the page are detected well, with the fixation of the localized curvatures reduced to the minimum. While the output image still contains curvatures, the legibility of the page has improved significantly. A closer inspection of the top left corner of both input and output images certifies this claim. Noise removal and adaptive thresholding have also contributed to enhancing image quality. This is further substantiated by the character recognition accuracy reported in the quantitative analysis section. OCV DocScan has enhanced the image with fairly well boundary detection and binarization. However, the dewarping achieved has failed to handle the complex page curvatures. GBS U-Net, in comparison, has failed on multiple counts—boundary detection, dewarping, and even skew correction.

5.2.4 Output visualization

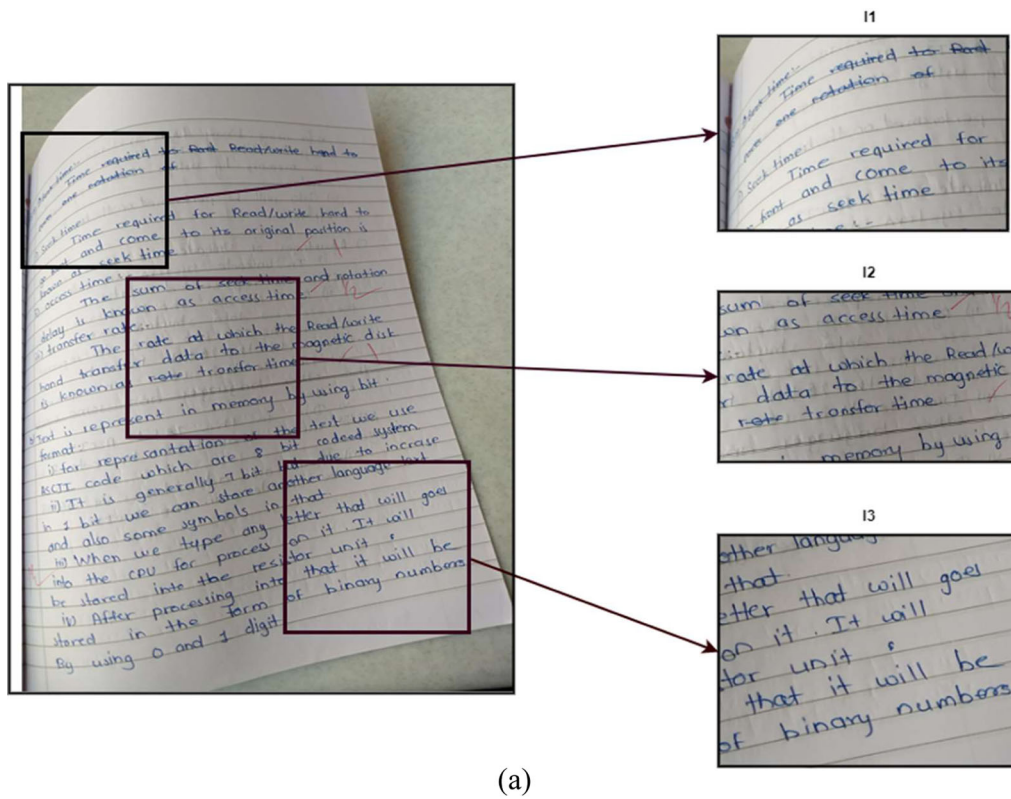
In summary, while our proposed algorithm falls short of generating the perfect output, it has generated a far better result

than the compared algorithms. A closer visual analysis of the output generated by the proposed algorithm is performed on the high-complexity images. This provides exciting observations related to the improvements brought about by the algorithm on the legibility of the textual content. Figure 11a presents a sample input image of a highly complex class with three separate regions marked on the image. I1, I2, and I3 are the callouts of these regions for a zoomed-in view. Figure 11b presents the output obtained from the algorithm for the input image in Fig. 11a. Similarly, O1, O2, and O3 are the equivalent zoomed-in localized portions of the output image. A comparative viewing of I2 and I3 vs. O2 and O3, respectively, shows that the legibility of the textual content is similar in both input and output images. It, however, may be stated that the skew and curvature present in these portions of the input image (I2 and I3) have been corrected near ideally in the outputs (O2 and O3). However, the inspection of I1 vs. O1 reveals interesting observations. While it may be noted that the text present in the left portions of the I1 is barely human readable, the same is readable in O1. Thus, it may be concluded that the proposed algorithm can dewarp and enhance the images across complex classes with high efficiency.

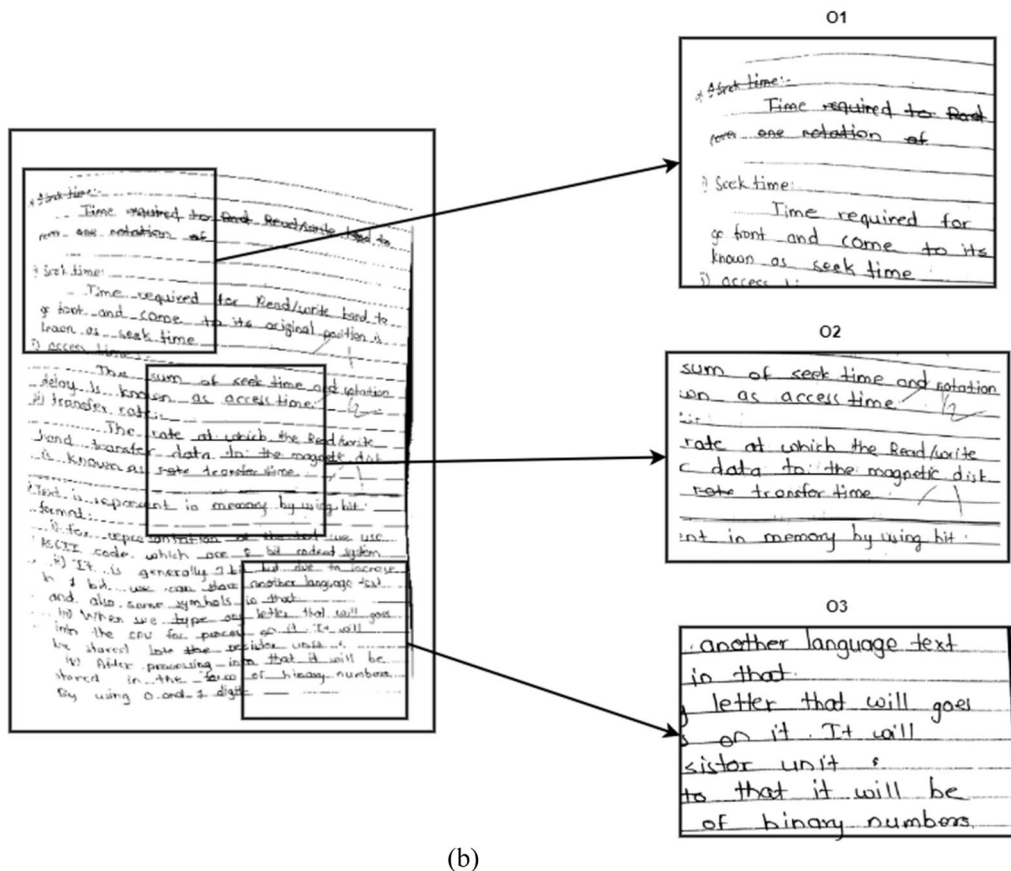
A comparative study is carried out to check the performance of the proposed algorithm on handwritten and typewritten documents. Figure 12 presents the results of such a study. Figure 12a–c represent images of an un-ruled book with handwritten text while the images in Fig. 12g–i present typewritten text documents with plane paper. The outputs for Fig. 12a–c are available as Fig. 12d–f, respectively, while those of Fig. 12g–i are available as Fig. 12j–l, respectively. The input images present in Fig. 12g–i are taken from WDID dataset [51], while the images in Fig. 12a–c are collected from internal sources. A closer look at the visual performance for both sets of input images reveals that the proposed algorithm works equally efficiently on handwritten as well as typewritten text document images. The outcomes are presented in low-, medium-, and high-complexity documents.

5.3 Quantitative analysis

The qualitative analysis performed and reported in Sect. 5.2 reveals that superior enhancement is achieved with the proposed algorithm. This section deals with quantitative analysis to accomplish the same. The quantitative analysis deals with two critical parameters: Intersection Over Union (IoU) and optical character recognition (OCR) output accuracy. The results for these parameters are explained in detail in the following subsections.



(a)



(b)

Fig. 11 Increase in visual quality with no loss of textual content. Sample input image (a) with its corresponding output obtained from the proposed algorithm shown in (b)

Fig. 12 Experiments on un-ruled samples: **a** Low complexity—input sample **b** Medium complexity—input sample **c** High complexity—input sample **g-l** dewarping on WDID datasets of low (**g**), medium (**h**), and high (**i**) complexity documents of Devanagari script

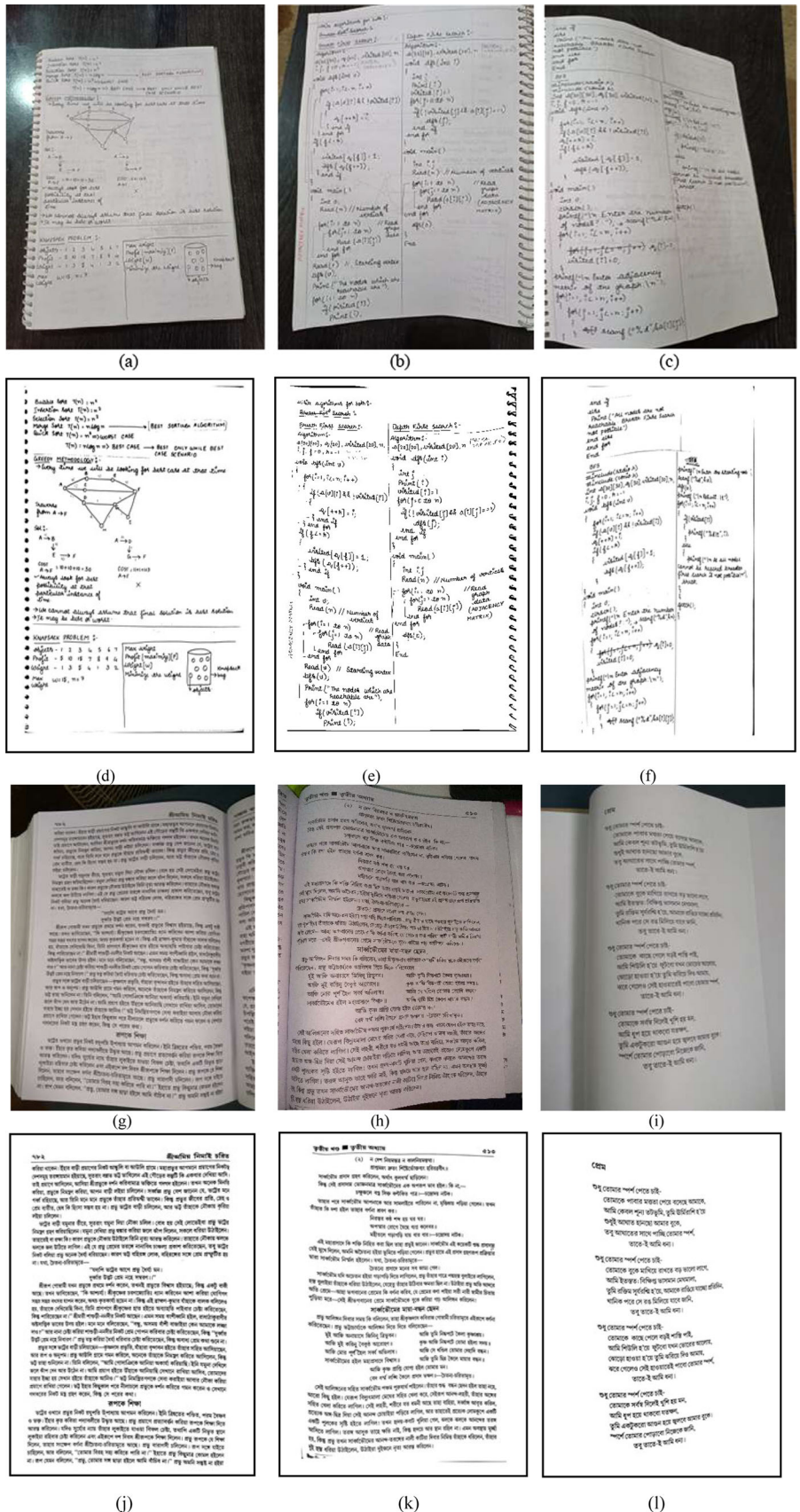


Table 2 IoU values obtained for image enhancement using the proposed method against the two other comparative methods described in Sect. 5.2

Methods	Complexity of warps		
	Low	Medium	High
Andrewdcampbell [50] OCV DocScan—IoU	0.78	0.73	0.51
Bandyopadhyay [39] GBS U-Net—IoU	0.91	0.85	0.74
Proposed method—IoU	0.92	0.88	0.80

AVV-CD234 dataset is used for this study

5.3.1 IoU

Since the proposed dewarping technique is heavily dependent on the boundary of the page image, an evaluation of the accuracy of the boundary detection technique employed is necessary. IoU has been extensively used in the literature to assess various object detection techniques and, therefore, can logically be extended to assess the page boundary detection activity.

Set A consists of all pixels manually labeled to belong to the page in the input image. Similarly, Set B consists of all pixels which belong to the page image as per the boundary detection algorithm. Union of sets A & B contains pixels that belong to the page image either as labeled manually or detected by the algorithm. Intersection or Overlap of A & B consists of all pixels labeled to be page image pixels both by a human agent and the algorithm. IoU, as depicted in Eq. (9), is defined as the ratio of the area of overlap to the area of union.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{A \cap B}{A \cup B} \quad (9)$$

Images from all different complexity classes are considered for the IoU computation. IoU is evaluated based on 234 images of the AVV-CD234 dataset. Table 2 presents the IOU evaluation results on the various complexity classes using the three methods.

It may be observed from Table 2 that images belonging to the low-complexity class produced better IoU scores over the scores for medium- or high-complexity image classes. This phenomenon is observed across different methods employed for enhancement. Additionally, with the increase in complexities, there is a reduction in the IoU score for all the methods employed. The low gap in IoU values for the low-complexity class of images may be observed in column 2 of Table 2. It may also be observed from Fig. 10 and row

2 of Table 2 that the average IoU values obtained for GBS U-Net (0.91) are close to the IoU value attained with the proposed method (0.92). However, it is recalled from observations noted in Sect. 5.2.1 that Method 2, assuming the image carries cylindrical distortion, induces curvatures into the text lines even though it can perform boundary detection reasonably accurately. Since the proposed method can detect boundaries accurately without inducing further distortion, it is apt to infer that it performs superior quality enhancement for the low-complexity class of images.

When the IoU values are observed for medium and high-complexity classes, present in columns 3 and 4, respectively, the proposed method achieves the highest IoU values across the classes. Additionally, the gap of IoU values between the methods keeps expanding as the level of complexity increases (see Fig. 13). In all cases, the IoU values of GBS U-Net are closer to the proposed technique than the same values for OCV DocScan. From these results, it may be inferred that the proposed technique works equally well on images of various complexity levels and outweighs the enhancement by available competitive algorithms.

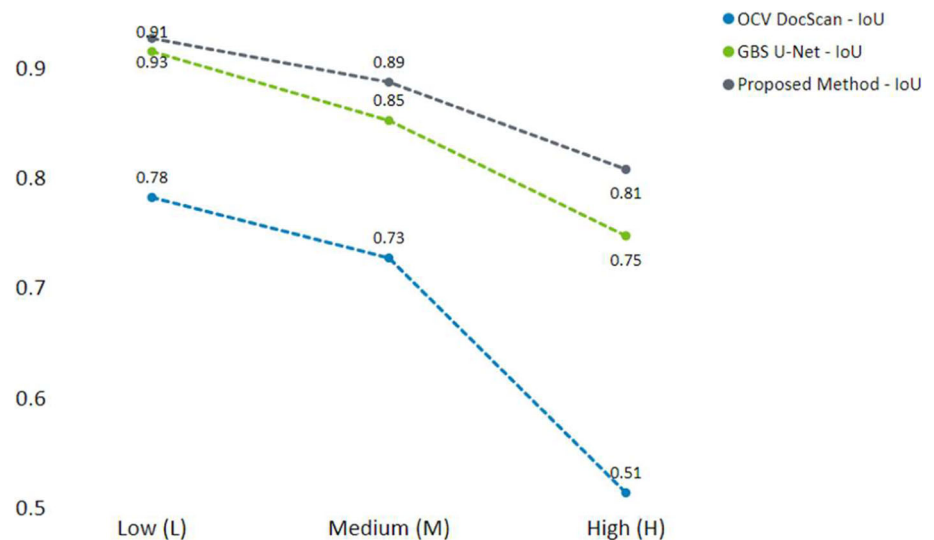
5.3.2 Computational complexity

The computational time taken by the proposed dewarping algorithm is explained in this section. As speed is not the primary research objective of this work, we have considered the time taken by the proposed dewarped algorithm against the state-of-the-art methods. The computational complexity is analyzed by considering the three categories of datasets; low, medium, and high warps. The performance of the proposed method is measured against the OCV DocScan [50] and GBS U-Net [39]. The amount of time taken is quantified in seconds for the test set images considered and presented in Table 3.

5.3.3 OCR Accuracy

The most crucial purpose of our enhancement technique is to feed the output image to an OCR engine for content extraction. This shall help achieve various automation activities downstream. An algorithm that creates excellent visual quality output but fails to generate better accuracy at the OCR stage is not entirely usable for automation needs. Thus, the accuracy of OCR output makes it an apt measure for the enhancements achieved.

Google Vision OCR is used for character recognition of page images. The generated output from the OCR engine consists of text blocks with corresponding bounding box coordinates. The bounding box coordinates help us map the recognized word to its location in the document image for accurate result analysis. The mapping between the text word and its location enables a comparative analysis of the OCR

Fig. 13 IoU of existing methods**Table 3** Computational time obtained for dewarping using the proposed method against the two other comparative methods described in Sect. 5.2

Complexity	Number of images	Andrewdcampbell [50] OCV DocScan (in seconds)	Bandyopadhyay [39] GBS U-Net (in seconds)	Proposed method (in seconds)
Low	15	5.932	12.112	6.948
Medium	15	6.607	13.516	8.824
High	15	7.236	15.248	11.072

AVV-CD234 dataset is used for this study

performance between the original and enhanced images (output from the dewarping process).

OCR is done on a word-by-word basis and compared with the ground truth data with significant emphasis on words located in the curved/warped regions. The bounding box coordinates help analyze the displacement of words affected by warps after dewarping. Recognition of such words present in curved/warped regions directly correlates to the system's efficiency in terms of OCR. Images of various complexities are used before and after they are enhanced with the proposed and other algorithms. The results obtained were compared against the ground truth data collected manually. This helps us determine recognition accuracy for original images and those after enhancement. The results obtained from this analysis are tabulated in Table 4 and presented graphically in Fig. 14.

It may be observed from the data available in Col. 1 of Table 4 that the OCR accuracies of the original images are pretty low and below 55% across different image classes of the dataset. The results from recognizing images from OCV DocScan are presented in Col. 2 of Table 4. A closer examination of the numbers here reveals that while the method brings enhancements that boost OCR accuracy for low complex images, the accuracy obtained for high-complex images

has gone for deterioration. The reason for such behavior of OCV DocScan is quite perplexing and needs further investigation.

When the results from GBS U-Net, present in Col 3 of Table 4, are analyzed, it is observed that the accuracy improvement brought about is maximum for low complex images. This contradicts the observations during the visual inspection presented in the qualitative analysis section (Ref. Sect. 5.2.1). The visual inspection indicated that the text lines are curved in the output image of the algorithm, which suggests a possible deterioration of the OCR accuracy. The only possible explanation that may be provided here is that the technique improves the image in noise cleaning and binarization. At the same time, the OCR engine can handle the induced curvatures as they were uniformly induced. This method has significantly improved OCR accuracy for medium and high-complexity images. The improvements achieved by the algorithm for medium and high-complexity image classes are 17 and 18.2%, respectively. This is a significant improvement brought in by the algorithm for character recognition with image enhancement. However, it may be observed that the improvement brought about is proportionate to the accuracies of the original images.

Table 4 OCR accuracy measure (in %) for images before and after the enhancement process using dataset AVV-CD234

complexity type	Original image OCR	OCV DocScan OCR	GBS U-Net OCR	Proposed method OCR
Low (L)	53.1	56.3	78.9	89.4
Medium (M)	43.0	48.0	60.0	95.0
High (H)	35.1	36.6	53.3	93.3
Avg. Accuracy	43.7	46.9	64.0	92.5

Google Vision OCR is used as the recognition engine for this study

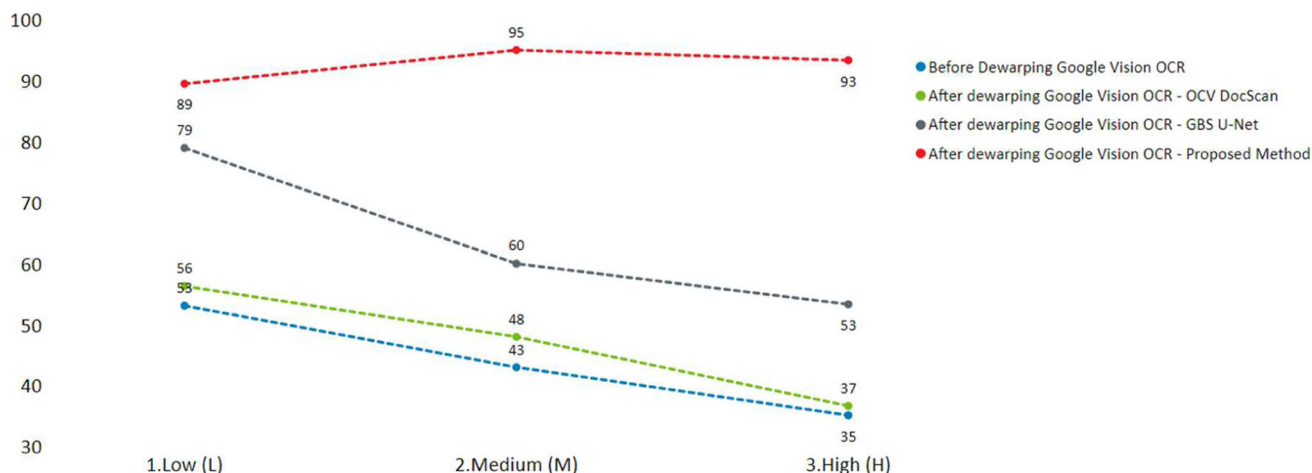


Fig. 14 Graph representing the OCR accuracy obtained on documents from AVV-CD234 dataset on original and enhanced images with various techniques

OCR accuracy results obtained from outputs of the proposed algorithm for various image complexity classes are presented in Col. 4 of Table 4. It may be observed that the proposed algorithm can enhance the images to generate similar OCR accuracies irrespective of the input image complexity. The output image OCR accuracies vary between 89 to 95%. This is against the accuracy of 35–53% observed for the original images. The proposed algorithm has significantly enhanced the accuracy for low, medium, and high-complex image classes, 36, 52, and 58% respectively. Thus, the proposed technology has brought extraordinary accuracy improvements (doubled the OCR accuracies) and reduced the range of accuracies across the input image complexity classes. This accuracy range reduction suggests that the technology can generate similar quality output images irrespective of the input image quality.

It may be observed from both Table 4 and Fig. 14 that the OCR accuracy obtained from the output images of the proposed algorithm is lesser than those for medium and highly complex image classes. This is counter-intuitive. On closer examination, it was observed that the quality of handwriting in the low complex images is poor compared to the medium and high-complex images as demonstrated in Fig. 15. Thus, it is inferred that the lower OCR accuracy obtained for low complex images is not an adverse effect induced by the proposed

algorithm but associated with the ability of the OCR engine to deal with poor handwriting.

The proposed system has been aptly demonstrated to be superior to its competitive methods used for performance evaluation on the AVV-CD234 dataset. The experiment was extended to observe the algorithm's behavior on various other datasets described in Sect. 5.1. The results of this experiment are presented in Table 5 and a graphical representation of the same is provided in Fig. 16.

CCHD124 and RTHD600 datasets consist of camera-captured handwritten document images with near zero warps. Therefore, the OCR accuracy for the original and enhanced images of these datasets is very close to each other (see Table 5 and Fig. 16). Larger accuracy improvement is observed for the CBDAR dataset, which consists of printed documents captured by a camera with slight warps present in them. It is very exciting to observe that as the complexity of document warp increases, the proposed algorithm is able to deliver better quantitative accuracy improvement. This is very clearly observed in Table 4. It is already observed and noted that the accuracy improvement increases with the increase in warp complexity order. This is in line with the observations obtained from Table 4. When this dataset was enhanced with the proposed algorithm, a more significant improvement in accuracy, an increase of 12.5%, was observed. The results

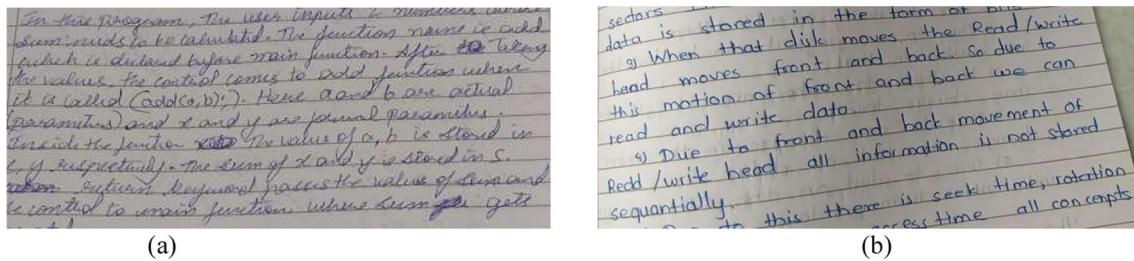


Fig. 15 **a** Low-complexity image with illegible handwriting vs. **b** High-complexity image with good handwriting

Table 5 OCR accuracy measure (in %) for images before and after the enhancement process across various datasets

Dataset	CCHD124	RTHD600	CBDAR 2007 Dataset	DDFCN120	AVV-CD234
Original images	76.1	75.9	82.1	75.4	43.7
Enhanced images	78.4	79.6	90.1	87.8	92.6

Google Vision OCR is used as the recognition engine for this study

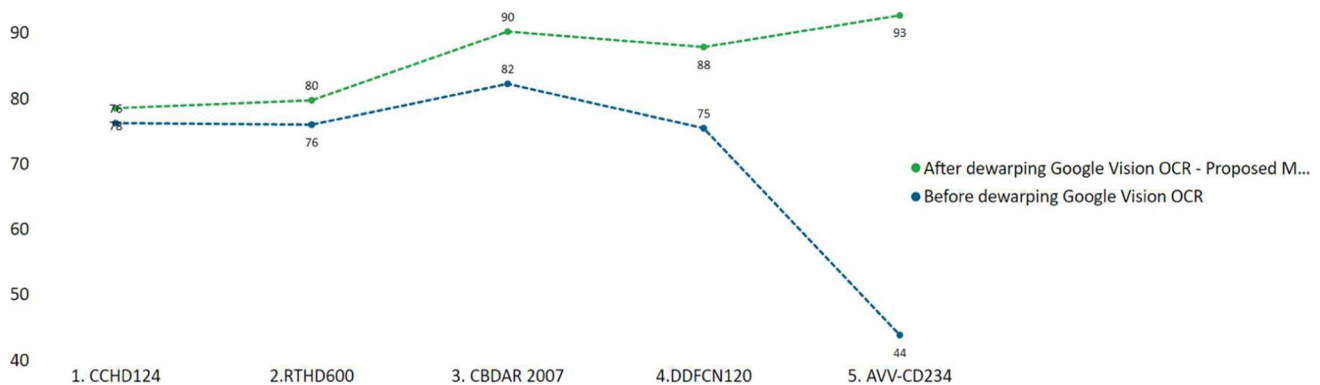


Fig. 16 Comparative analysis of the proposed system with various datasets described in Sect. 5.1. These datasets contain images with different complexities, capture resolutions, sizes & illumination, and noise level variations (details tabulated in Table 2)

were obtained with the AVV-CD234 dataset as column 6 in Table 5 for easier comparison. Earlier in this subsection, it has been explained that the accuracy improvement is proportional to images' complexity increase, and this observation is consistent with the earlier presented inference.

5.3.4 Limitations

The following use cases can hinder the performance of the proposed method:

- The document image content contains graphical elements such as photographs, charts, and graphs.
- The background region of the document image is homogeneous with the foreground region.
- Document images consisting of non-horizontal writings such as billboards and artworks.
- Non-rectangular pages.

6 Conclusion and future work

This paper proposes a robust document detection and dewarping system that can handle warped document images with simple-to-severe perspective and geometric distortions. The proposed automated technique automatically estimates the control points from simple-to-severe distortion scenarios effectively. We have evaluated the performance of our system by computing the OCR accuracy using Google Vision OCR API to establish the reliability of the proposed method for post-dewarping tasks such as recognition. Data acquired through experimental analysis shows that the system can handle images with severe warps and distortions. Moreover, we have also created a robust real-time dataset with images containing warps and other distortions of different complexities.

In the future, a robust method to generate more than four control points for folded documents with severe warps poses a research challenge. Moreover, a completely unsupervised

dewarping technique with the reduced computational cost is the motivation to be taken up to foresee the future developments of dewarping systems.

Author contributions These authors contributed equally to this work

Declarations

Conflict of interest N. Shobha Rani reports administrative support and statistical analysis were provided by Amrita Vishwa Vidyapeetham. Nachappa C H reports a relationship with Amrita Vishwa Vidyapeetham that includes: education and non-financial support. Gokulnath M reports a relationship with Amrita Vishwa Vidyapeetham that includes: employment and non-financial support.

References

- Liang, J., Doermann, D., Li, H.: Camera-based analysis of text and documents: a survey. *IJDAR* **7**(2), 84–104 (2005)
- Jung, K., Kim, K.I., Jain, A.K.: Text information extraction in images and video: a survey. *Pattern Recogn.* **37**(5), 977–997 (2004)
- Wolberg, G.: *Digital image warping*, vol. 10662, pp. 90720–91264. IEEE computer society press, Los Alamitos (1990)
- Heckbert, P.S.: *Fundamentals of texture mapping and image warping* (1989)
- Zhang, Z., Tan, C.L.: Correcting document image warping based on regression of curved text lines. In: *Seventh International Conference on Document Analysis and Recognition*, 2003. Proceedings, pp. 589–593. IEEE (2003)
- Wagdy, M., Amin, K., Ibrahim, M.: Dewarping document image techniques: survey and comparative study. *Int. J. Image Graph.* **21**(03), 2150031 (2021)
- Wu, C., Agam, G.: Document image de-warping for text/graphics recognition. In: *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pp. 348–357. Springer, Berlin (2002)
- Kil, T., Seo, W., Koo, H.I., Cho, N.I.: Robust document image dewarping method using text-lines and line segments. In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1, pp. 865–870. IEEE (2017)
- Sutapirat, D., Sookhanaphibarn, K., Lursinsap, C.: Model-based book dewarping method for content-independent document images. In: *2009 International Conference on Digital Image Processing*, pp. 190–194. IEEE 2009.
- Dehbovid, H., Razzazi, F., Alirezaii, S.: A novel method for dewarping in Persian document images captured by cameras. In: *2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, pp. 614–619. IEEE (2010)
- Bukhari, S.S., Shafait, F., Breuel, T.M.: Coupled Snakelet model for curled textline segmentation of camera-captured document images. In: *2009 10th International Conference on Document Analysis and Recognition*, pp. 61–65. IEEE (2009)
- Bukhari, S.S., Shafait, F., Breuel, T.M.: Dewarping of document images using coupled-snakes. In: *Proceedings of Third International Workshop on Camera-Based Document Analysis and Recognition*, Barcelona, Spain, pp. 34–41 (2009)
- Bukhari, S.S., Shafait, F., Breuel, T.M.: Segmentation of curled textlines using active contours. In: *2008 The Eighth IAPR International Workshop on Document Analysis Systems*, pp. 270–277. IEEE (2008)
- Stamatopoulos, N., Gatos, B., Pratikakis, I.: A methodology for document image dewarping techniques performance evaluation. In: *2009 10th International Conference on Document Analysis and Recognition*, pp. 956–960. IEEE (2009)
- Kim, B.S., Koo, H.I., Cho, N.I.: Document dewarping via text-line based optimization. *Pattern Recogn.* **48**(11), 3600–3614 (2015)
- Song, L., Wu, Y., Sun, B.: A robust and fast dewarping method of document images. In: *2010 International Conference on E-Product E-Service and E-Entertainment*, pp. 1–4. IEEE (2010)
- Ramanna, V.K.B., Bukhari, S.S., Dengel, A.: Document image dewarping using deep learning. In: *ICPRAM*, pp. 524–531 (2019)
- Garai, A., Biswas, S., Mandal, S.: A theoretical justification of warping generation for dewarping using CNN. *Pattern Recogn.* **109**, 107621 (2021)
- Xie, G.W., Yin, F., Zhang, X.Y., Liu, C.L.: Dewarping document image by displacement flow estimation with fully convolutional network. In: *International Workshop on Document Analysis Systems*, pp. 131–144. Springer, Cham (2020)
- Ulges, A., Lampert, C.H., Breuel, T.M.: Document image dewarping using robust estimation of curled text lines. In: *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, pp. 1001–1005. IEEE (2005)
- Ezaki, H., Uchida, S., Asano, A., Sakoe, H.: Dewarping of document image by global optimization. In: *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, pp. 302–306. IEEE (2005)
- Kakumanu, P., Bourbakis, N., Black, J., Panchanathan, S.: Document image dewarping based on line estimation for visually impaired. In: *2006 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06)*, pp. 625–631. IEEE (2006)
- Koo, H.I., Kim, J., Cho, N.I.: Composition of a dewarped and enhanced document image from two view images. *IEEE Trans. Image Process.* **18**(7), 1551–1562 (2009)
- Ghods, A.R., Mozaffari, S., Ahmadpanahi, F.: Document image dewarping using Kinect depth sensor. In: *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, pp. 1–6. IEEE (2013)
- Dasgupta, T., Das, N., Nasipuri, M.: Multistage curvilinear coordinate transform based document image dewarping using a novel quality estimator. *arXiv preprint arXiv:2003.06872* (2020)
- Simon, G., &Tabbone, S.: Generic document image dewarping by probabilistic discretization of vanishing points. In: *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 2344–2351. IEEE (2021)
- Wagdy, M., Amin, K., Ibrahim, M.: Detection and correction of multi-warping document image. *Int. J. Image Graph.* **22**50034 (2021)
- Labib, M.W., Amin, K.M.: Warped document image correction based on checkboard pattern and geometric transformation. *IJCI Int. J. Comput. Inf.* **8**(1), 30–54 (2021)
- Xie, G.W., Yin, F., Zhang, X.Y., Liu, C.L.: Document dewarping with control points. In: *International conference on document analysis and recognition*, pp. 466–480. Springer, Cham (2021)
- Shafait, F., Breuel, T.M.: Document image dewarping contest. In: *2nd Int. Workshop on Camera-Based Document Analysis and Recognition*. Curitiba, Brazil, pp. 181–188 (2007)
- Stamatopoulos, N., Gatos, B., Pratikakis, I.: Performance evaluation methodology for document image dewarping techniques. *IET Image Proc.* **6**(6), 738–745 (2012)
- Garai, A., Dutta, A., Biswas, S.: Automatic dewarping of camera-captured comic document images. *Multimed. Tools Appl.* 1–16 (2022)
- Bolelli, F.: Indexing of historical document images: Ad hoc dewarping technique for handwritten text. In: *Italian Research Conference on Digital Libraries*, pp. 45–55. Springer, Cham (2017)
- Rowtula, V., Bhargavan, V., Kumar, M., Jawahar, C.V.: Scaling handwritten student assessments with a document image workflow

- system. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 2307–2314 (2018)
35. Vinod, H.C., Niranjana, S.K.: Camera captured document dewarping and de-skewing. *J. Comput. Theor. Nanosci.* **17**(9–10), 4398–4403 (2020)
 36. Garai, A., Biswas, S., Mandal, S., Chaudhuri, B.B.: Automatic rectification of warped Bangla document images. *IET Image Process.* **14**(1), 74–83 (2020)
 37. Dulla, A.: A dataset of warped historical Arabic documents (2018)
 38. Feng, H., Wang, Y., Zhou, W., Deng, J., Li, H.: Doctr: Document image transformer for geometric unwarping and illumination correction. arXiv preprint [arXiv:2110.12942](https://arxiv.org/abs/2110.12942) (2021)
 39. Bandyopadhyay, H., Dasgupta, T., Das, N., Nasipuri, M.: A gated and bifurcated stacked U-net module for document image dewarping. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 10548–10554. IEEE (2021)
 40. Ma, K., Shu, Z., Bai, X., Wang, J., Samaras, D.: Docunet: document image unwarping via a stacked U-net. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4709 (2018)
 41. Dutta, A., Garai, A., Biswas, S., Das, A.K.: Segmentation of text lines using multi-scale CNN from warped printed and handwritten document images. *Int. J. Document Anal. Recognit. (IJ DAR)* **24**(4), 299–313 (2021)
 42. Souibgui, M.A., Biswas, S., Jemni, S.K., Kessentini, Y., Fornés, A., Lladós, J., Pal, U.: DocEnTr: an end-to-end document image enhancement transformer. arXiv preprint [arXiv:2201.10252](https://arxiv.org/abs/2201.10252) (2022)
 43. Luisier, F., Blu, T., Unser, M.: Image denoising in mixed Poisson–Gaussian noise. *IEEE Trans. Image Process.* **20**(3), 696–708 (2010)
 44. Chen, J., Benesty, J., Huang, Y., Doclo, S.: New insights into the noise reduction Wiener filter. *IEEE Trans. Audio Speech Lang. Process.* **14**(4), 1218–1234 (2006)
 45. Read, S.M.: The deconvolution problem: an overview. *Proc. IEEE* **74**(1), 82–85 (1986)
 46. Jamil, N., Sembok, T.M.T., Bakar, Z.A.: Noise removal and enhancement of binary images using morphological operations. In: 2008 International Symposium on Information Technology, vol. 4, pp. 1–6. IEEE (2008)
 47. Bradley, D., Roth, G.: Adaptive thresholding using the integral image. *J. Graph. Tools* **12**(2), 13–21 (2007)
 48. Meng, G., Pan, C., Xiang, S., Duan, J., Zheng, N.: Metric rectification of curved document images. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(4), 707–722 (2011)
 49. GitHub Repository: <https://github.com/NachappaCH/Dewarping-Dataset-Annotations.git>
 50. Andrewdcampbell.: Andrewdcampbell/OpenCV-document-scanner: an interactive document scanner built in python using OpenCV featuring automatic corner detection, image sharpening, and color thresholding. GitHub. Retrieved February 22, 2022, from <https://github.com/andrewdcampbell/OpenCV-Document-Scanner> (n.d.)
 51. Shafait, F., Breuel, T.M.: Document image dewarping contest. In: 2nd Int. Workshop on Camera-Based Document Analysis and Recognition, pp. 181–188 (2007)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.