Pattern Recognition 109 (2021) 107621

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

# A theoretical justification of warping generation for dewarping using CNN

Arpan Garai*, Samit Biswas, Sekhar Mandal

*Department of Computer Science and Technology, Indian Institute of Engineering Sciences and Technology, Shibpur, Hawrah, West Bengal, 711103, India*

ARTICLE INFO

*Article history:*
Received 10 February 2020
Revised 21 August 2020
Accepted 29 August 2020
Available online 30 August 2020

*Keywords:*
Dewarping
Artificial neural netwroks
Synthetic image generation

ABSTRACT

Dewarping is a necessary preprocessing step to recognize text from a distorted camera captured document image. According to recent literature, deep learning-based approaches perform with higher accuracy in similar domains. The deep learning-based neural networks are not yet fully explored in the domain of dewarping. To fill this gap, we propose a dewarping approach based on the convolutional neural network. A large number of images are required to train such networks. However, it is a tedious job to capture such a large number of images. Hence, it is required to generate synthetic warped images for the training phase of the deep learning-based neural network. The existing synthetic warped image generation methods are heuristic-based. In this paper, we propose a novel mathematical model for the generation of warped images. The proposed model takes some parameters such as depth of the surface, camera angle, and camera position and generates the corresponding warped image. These parameters are the ground truth for that particular warped image. We use a Convolutional Neural Network (CNN) based model to estimate the warping parameters from a 2*D* warped image for dewarping. In the training phase of CNN based model, the synthetic images and their corresponding ground truth are used. Next, the trained model is used to dewarp the unknown warped images. The performance of the proposed warping model is analyzed. Finally, the proposed dewarping method is compared with existing approaches. In both cases, the results are encouraging.



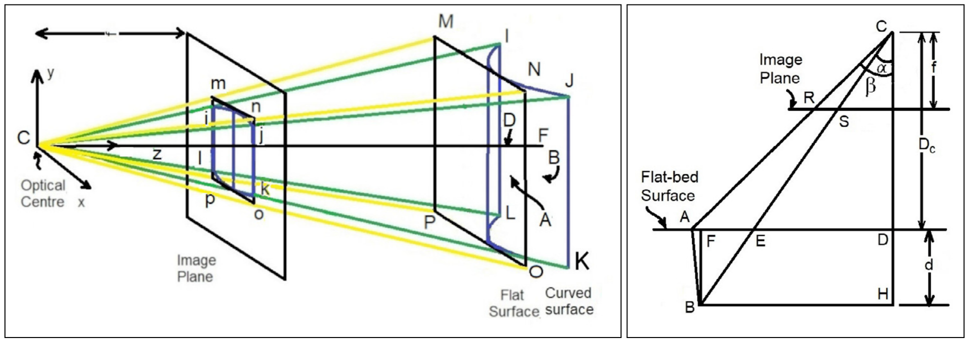
## 1. Introduction

In the current era, digital cameras associated with smart mobile phones and similar devices, are frequently used to capture the paper documents. As a result, various types of distortions are often found in those images. Warping is one of the important distortions in these camera captured images. The traditional systems for optical character recognition [1], optical character detection [2], text recognition [3], and other document processing tasks often fail to perform with a significant accuracy if warped document images are used as input. So, the documents must be restored to improve the performance of OCR/other document processing tasks. Most of the existing dewarping methods are traditional learning-free approaches and tested on Alphabetic scripts like English. They often fail to produce accurate results for Alpha-syllabary scripts like Bangla. In the recent past, the Convolutional Neural Network (CNN) [4,5] and Generative Adversarial Network (GAN) [6,7] based techniques perform well in the domain of Document Image Analysis (DIA). So, we propose a deep CNN based approach for dewarping and test the algorithm on both Alphabetic and Alpha-syllabary scripts. Generally, deep learning-based approaches need a large number of images during the training phase. The following warped document image datasets are publicly available: (i) 'DFKI document image contest dataset' [8] and (ii) 'Doc3D dataset' [4]. DFKI document image contest dataset contains 102 and 130 images present in the 'Doc3D dataset'. We create another dataset 'WDID' that contains 258 images. The number of images in the datasets mentioned above is not sufficient to train a deep CNN model. Manually capturing such a large number of warped document images and preparing the ground truths is difficult. A procedure for synthetically generated warped document image is an alternative. In this paper, a synthetic warped image generation model is also proposed.

To generate synthetic images, similar to a real warped image, we need a suitable mathematical model. The model should incorporate all the necessary geometric parameters like the curvature of the surface, camera angle, the position of the camera over the document, distance from the document surface to the camera, etc. In this paper, we propose such a mathematical model, which is discussed in Section 3.

* Corresponding author.
*E-mail address:* ag.rs2016@cs.iiests.ac.in (A. Garai).

https://doi.org/10.1016/j.patcog.2020.107621
0031-3203/




(a) (b)

**Fig. 1.** Model for warping generation: (a) Three-dimensional view; (b) Two-dimensional view.

The rest of the paper is organized as follows. Section 2 briefly describes the related works. In Section 3, the proposed warping generation model is discussed. Next, Section 4 describes the process of the generation of synthetic images using the proposed warping model. The deep CNN based framework is described in Section 5. Experimental results and evaluation of the proposed method are shown in Section 6. Finally, the concluding notes are put in Section 7.

## 2. Related works

The problem of dewarping is solved in several ways which can be roughly classified into two categories: 'hard' and 'soft' methods. The methods which use external 'hardware' [9,10] are termed as 'hard' methods, and the rest are 'soft' methods [11].

Again, the 'soft' methods can be classified into two classes. They are termed as traditional approaches and deep learning-based approaches [12]. The traditional approaches generally consist of two sub-steps, namely implicit or explicit warp estimation and distortion correction. Depending on the warp estimation, the existing traditional dewarping methods are of two types: 2-Dimensional (2D) and 3-Dimensional (3D) [13]. In 3D methods, the shape of a page is estimated, and it is represented using the geometric models, e.g., cylindrical surface model by Cao et al. [14], shape-from-shading model by Zhang et al. [15], General Cylindrical Surface (GCS) model by Meng et al. [16], etc. Liang et al. [17] and Meng et al. [16] proposed 3D methods based on texture flow and 'line convergent symmetry', respectively. On the other hand, You et al. [18] and He et al. [19] proposed models based on the smoothness of the image and the boundary of the book, respectively.

In 2D methods, the warping is estimated using the flow of the text lines and the slant-ness of the characters present in the document image. Ezaki et al. [20] and Lu et al. [21] proposed 2D methods based on an elastic curve model and vertical stroke boundary (VSB), respectively. The method given by Ulges et al. [22] rectifies of both perspective and page curl distortion. In the method proposed by Gatos et al. [23], a word-wise lower and upper baseline is estimated. Also, the slant of each word is calculated. A two-stage rectification technique is proposed by Stamatopoulos et al. [24]. In this method, a coarse rectification is done using page borders, followed by fine rectification, using word baseline. Each pixel of the document image is transformed according to the estimated of that boundary. This boundary supposes to be a rectangle in the method given by Fu et al. [11]. A discrete representation of the text line is used to estimate the flow of text line in the method proposed by Kim et al. [25]. In the method given by Liu et al. [13], the warping is estimated based on the shape of the baselines and slant of the characters. Kil et al. [26] measured the warping from text lines and text line segments and used an iterative approach to dewarp the image. Recently, a mesh-based technique is proposed by Yang et al. [27] to dewarp the historical documents.

The general idea of these methods is that the baseline is calculated and it is used to compute the amount and direction of the warping. This idea often fails on documents containing text mixed with figures. Moreover, most methods contain time-consuming and complicated optimizations [12]. Recently, Ma et al. [4] proposed a deep neural-based approach called 'DocUNet' to dewarp the documents. It has used only the 2D information of a warped document image but performs well only on synthetic images. Das et al. [12] have also proposed a dewarping technique that used deep neural networks. The method included the 3D geometry of the document as input to their network. In [28], Liu et al. proposed another deep neural network called Adversarial Gated Unwarping Network (AGUN) to rectify geometric distortions. The method consists of three gated-based modules. They are spatial gated module (SGM), element-wise gated module (EGM), and channel gated module (CGM). The method also performs well for synthetic images and not satisfactory for highly crumpled images.

## 3. Warping: a surface model

Kieu et al. [29] have proposed a semisynthetic method that is based on a mesh, and the mesh is generated by 'the KrĘon Aquilon laser 3D scanner'. In [4], a warping method is proposed. Here, a mesh is used to get the control points for warping. The strength and direction of deformation at a point are randomly generated. Kim et al. [25] have generated synthetic images to evaluate the dewarping method proposed by them. The methods mentioned above do not provide proper theoretical justification for their approaches. Here, a mathematical model is proposed for warped image generation.

Consider a flat-bed surface (*MNOP*) parallel to the image plane on the warped surface (*IJKL*) and a camera having an optical center at *C* (see Fig. 1(a)). The images obtained from the flat-bed and the warped document surface are *mnop* and *ijkl*, respectively. The image generated from the warped surface is distorted. Here, we estimate the amount of distortion in *ijkl* with respect to *mnop*. Let *A* be a point on the flat-bed surface and *CD* is the perpendicular to the flat surface. If we warp the flat-bed surface, then the point *A* on the flat-bed surface may shift to *B* on the curved surface (see Fig. 1(b)). The lines *CA* and *CB* intersect the image plane at *R* and *S*, respectively. Next, we draw a perpendicular from *B* to the flat-bed surface. Let the length of *BF* and *AF* are *d* and $\gamma$, respectively. Here, $\angle DCA = \beta$ and $\angle DCE = \alpha$. It is evident from Fig. 1(b) that the point *B* will appear at point *E* and *S* on the flat-bed surface and the image plane (if the viewpoint is *C*), respectively. So, the amount of lateral shift of the point *A* on the flat-bed surface

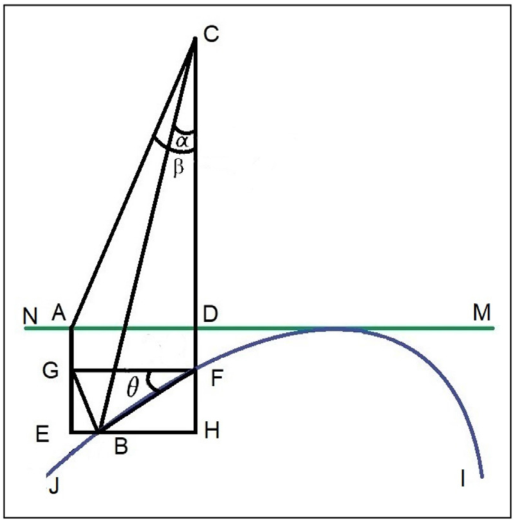



**Fig. 2.** Two dimentional view of the warping model for calculation of $\alpha$.

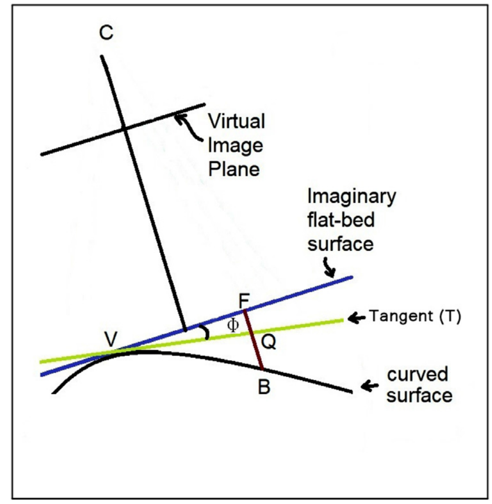



**Fig. 3.** Two dimensional view of the warping model for calculation of depth.

is *AE* and the same on the image plane is *RS*. Now, from $\triangle RCS$ and $\triangle ACE$, we get Eqs. (1) and (2), respectively.

$$\frac{RS}{\sin \angle RCS} = \frac{CS}{\sin \angle CRS} \implies \frac{RS}{\sin(\beta - \alpha)} = \frac{f.\cos\alpha}{\sin(90 - \beta)} \tag{1}$$

$$\frac{AE}{\sin \angle ACE} = \frac{CE}{\sin \angle CAE} \implies \frac{AE}{\sin(\beta - \alpha)} = \frac{D_c.\cos\alpha}{\sin(90 - \beta)} \tag{2}$$

Here, $f$ is the focal length of the camera, i.e., the perpendicular distance between optical center and image plane and $D_c$ is the perpendicular distance from the optical center to the flat-bed surface. Let the values of AE and RS are $\Delta_s$ and $\Delta_i$, respectively, then from Eqs. (1) and (2) we get the following:

$$\frac{RS}{AE} = \frac{f}{D_c} \Rightarrow RS = \frac{f}{D_c} \times AE \Rightarrow \Delta_i = \frac{f}{D_c} \times \Delta_s \tag{3}$$

*RS* is proportional to *AE* as the value of $\frac{f}{D_c}$ for an image is a constant. From Fig. 1(b), we calculate $AE = \Delta_s = AF + FE = \gamma + d\tan\alpha$. Here, $\angle DCB = \angle CBF = \alpha$ (alternate interior angles) and $BF = d$. Note that both $\alpha$ and $\beta$ are the two variables which can vary both in the x-direction as well as y-direction. So, the warping appears in both x-axis and y-axis. Here, $d$ is a 2D matrix that contains the perpendicular distance from each point in the curved surface to the flat-bed surface and the matrix, $d$ is called depth of the curved surface. So, the values of $d$ depend on the curve-ness of the surface and the angle between camera and the surface. The $\alpha$ and $\gamma$ depend upon other external features which are discussed in the following sub-sections.

### 3.1. Calculation of $\alpha$

The perpendicular drawn from *C* on the flat-bed surface (*MN*) intersects it at *D* and the extended part of *CD* cuts the curved surface (*IJ*) at *F* (see Fig. 2). Here, *A* is the point on the flat-bed surface and $\angle ACD = \beta$ and $\angle BCD = \alpha$. The length of *AD* and *BF* are $a$ and $p$, respectively. The following line segments are drawn *AE, FH, HB, BE* and *FG* such that $AE||CF$, $FH||AE$, $FG \perp AE$, $HB||FG$ and $BE||FG$. Considering Fig. 2, the following equation is obtained.

$$\tan(\beta - \alpha) = \frac{\tan\beta - \tan\alpha}{1 + \tan\beta.\tan\alpha} = \frac{\frac{AD}{CD} - \frac{BH}{CH}}{1 + \frac{AD}{CD}\cdot\frac{BH}{CH}} \tag{4}$$

The length of *FD* is denoted by $d_1$ and the angle $\angle BFG = \theta$. The $\angle FBH$ and $\angle BFG$ are alternate interior angles. So, $\angle FBH = \angle BFG = \theta$. Hence, $BH = p.\cos\theta$ and $FH = p.\sin\theta$. Substituting these values in Eq. (4) we get:

$$\tan(\beta - \alpha) = \frac{\frac{a}{D_c} - \frac{p.\cos\theta}{D_c + d_1 + p.\sin\theta}}{1 + \frac{a}{D_c}\cdot\frac{p.\cos\theta}{D_c + d_1 + p.\sin\theta}}$$

$$\Rightarrow \alpha = \beta - \tan^{-1}\left[\frac{\frac{a}{D_c} - \frac{p.\cos\theta}{D_c + d_1 + p.\sin\theta}}{1 + \frac{a}{D_c}\cdot\frac{p.\cos\theta}{D_c + d_1 + p.\sin\theta}}\right] \tag{5}$$

### 3.2. Computation of $\gamma$

The length of *EB* is $\gamma$ and $EB = EH - BH$ (see Fig. 2). Also, $EH = AD = a$ and $BH = p.\cos\theta$. Therefore, $\gamma = a - p.\cos\theta$.

### 3.3. Computation of the depth matrix (d)

The depth of a particular point on the curved surface is the perpendicular distance from the flat-bed surface. It depends upon the curvy-ness of the surface and the angle ($\phi$) between *T* and the flat-bed surface, as shown in Fig. 3. Here, *T* is the tangent drawn at the point (*V*) on the curved surface where the flat surface touches the curved surface. The depth due to the curvy-ness ($\rho$) of the curved surface depends on the real-world environment. Now consider an arbitrary point *B* on the curved surface, and a perpendicular (*BF*) is drawn from *B* on the flat-bed surface. *BF* intersects *T* at point *Q*, then the depth at this point (*B*) is $BF = BQ + QF$. Here, *BQ* and *QF* are the depths due to the curvy-ness of the surface and $\phi$, respectively. So, $QF = VF \times \tan\phi$. Now, consider that the value of *VF* is $b$. So, the depth at any point in the curved document is given by $d = \rho + b \times \tan\phi$. The next section describes the generation of the synthetic warped images from a flat-bed scanned document image using this warping model.

## 4. Synthetic image generation

It is evident from the warping model that the distortion of a warped document depends on certain parameters such as $\alpha$, $\gamma$, and $d$. The value of $\alpha$ depends on $\beta$, $a$, $D_c$, $p$ and $\theta$, whereas we get the value of $\gamma$ from $a$, $p$ and $\theta$. $d$ is dependent on $\rho$ and $\phi$. Here, $\beta$ and $a$ can be estimated from the position of the optical center. $\theta$ and $p$ individually is a function of the position of optical center and curvature of the document surface. $\rho$, as well as $b$, depends on the curvature of the curved surface. $D_c$ is the distance between the camera and the flat-bed surface.

Four parameters are needed to be specified to warp of a given image having flat-bed surface and they are (i) Distance from the optical center of the camera to the flat surface ($D_c$); (ii) Position of the optical center (say $C_p$); (iii) The angle ($\phi$); and (iv) Depth at each position due to curvature at each position of the curved surface ($\rho$). In our experiment, we take a flat-bed captured image as input and use different values of these four parameters to generate different warped images. The selection of these four parameters is discussed next.

**Measurement of distance from the optical center to the flat surface ($D_c$):** We measure the perpendicular distance from the flat-bed surface to the lens of the camera ($d_z$) while capturing that flat-bed image. If $f$ is the focal length of camera which can be estimated from a given image, then $D_c = d_z + f$.

**Specification of the position of the optical center ($C_p$):** The camera can be held at any place over the document surface. The $C_p$ is a weighted function of height ($H$) and width ($W$) of the input image. The weights can take any values within the range between 0 and 1. If the values of the weights are 0.5 and 0.5, then the camera is placed in the middle of the document. In our experiment, we select three different values of $C_p$. They are such that the camera is placed on the (i) top (0.3) (ii) bottom (0.7) and (iii) center (0.5) position of the document surface.

**Specification of angle ($\phi$):** Generally, people try to keep the angle $\phi$ to $0^o$ while capturing the document surface. But sometimes it varies. So, three different values of $\phi$ are applied. They are $-1.5^o$, $0^o$ and $1.5^o$.

**Specification of depth due to curvature of the surface ($\rho$):** Using different values of $\rho$ various types of depths are generated. $\rho$ is a 2D matrix, and the size of this matrix is the same as image size. The value of each cell of $\rho$ is obtained in the following way.

(i) For each row of the image, five knots are selected. The position each knot is determined by position parameters ($P_1$ and $P_2$) and the values of $P_1$ and $P_2$ are user-defined.

(ii) The value of each knot is determined by a set of eight control parameters ($P_3, \ldots, P_{10}$) which are also user-defined.

(iii) To obtain the values of other cells of a given row, we use a smoothing cubic spline interpolation function, which fits the five knots.

*Position of Knots:* The positions of the first and last knots of each row are the first and last positions of that row, respectively. The positions of the third knot of the first row and the last row are obtained as $P_1 \times W$ and $P_2 \times W$, respectively. Here, $W$ is the width of the input image and the parameters ($P_1$, $P_2$) take values within the range $0 - 1$. If the values of $P_1 \times W$ and $P_2 \times W$ are real, then we take their nearest integers. Now, a straight line ($Z_1Z_2$) is drawn through the third knots of the first and last rows, which are shown in Fig. 4. The intersection of this line with any row (except fist and last row) gives the position of the third knot of that corresponding row. The positions of the second and fourth knots of other rows are the middle point of first and third knots and the middle point of third and last knots, respectively.

*Estimation of $\rho$ matrix:* The values at first and last knots of the first row of $\rho$ are $P_3$, and $P_6$, respectively and the user will supply these values. The values at first and last knots of the last row of $\rho$ are $P_7$ and $P_{10}$, respectively, which are also specified by the user. $P_4$ and $P_5$ are the values of 2nd and 4th knots of the first row respectively. $P_8$ and $P_9$ denote the values of 2nd and 4th knots of the last row, respectively.

In our experiment, we set $P_4 = \frac{1}{2} \times P_3,\ P_5 = \frac{1}{2} \times P_6,\ P_8 = \frac{1}{2} \times P_7$ and $P_9 = \frac{1}{2} \times P_{10}$. The value of the third knot for all rows is set to zero. Consider the values of 1st, 2nd, 4th and 5th knots of the $i$th row are $K_1^i$, $K_2^i$, $K_4^i$ and $K_5^i$, receptively. The following equations are used to estimate their values. Here, $H$ denotes the height of the input image.

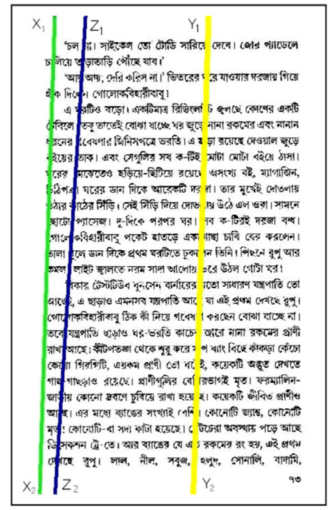



**Fig. 4.** Positions of 2nd(along $X_1X_2$/green line), 3rd( along $Z_1Z_2$ / blue line) and 4th(along $Y_1Y_2$ / yellow line) knots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**
Possible values of PP and CP used to generate different types of warping.

| Types of warping | $P_1$ | $P_2$ | $P_3^*, P_6^*, P_7^*, P_{10}^*$ |
|---|---|---|---|
| Type-I | 0.2 | 0.2 | $(0.04 \times D), \ldots, (0.06 \times D)$ |
| Type-II | 0.8 | 0.8 | $(0.04 \times D), \ldots, (0.06 \times D)$ |
| Type-III | 0.5 | 0.5 | $(0.04 \times D), \ldots, (0.06 \times D)$ |

[*] The step size for $P_3$, $P_6$, $P_7$, $P_{10}$ is $(0.01 \times D)$ where $D$ is the length of the diagonal of the image.

$$K_1^i = \frac{P_7 - P_3}{H - 1} \times (i - 1) + P_3 \tag{6}$$

$$K_2^i = \frac{P_8 - P_4}{H - 1} \times (i - 1) + P_4 \tag{7}$$

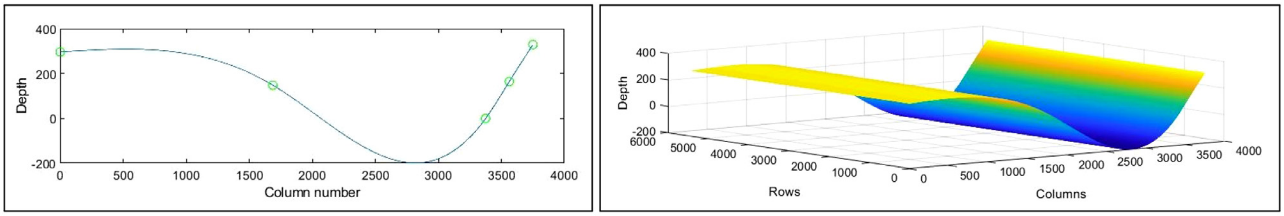
$$K_4^i = \frac{P_9 - P_5}{H - 1} \times (i - 1) + P_5 \tag{8}$$

$$K_5^i = \frac{P_{10} - P_6}{H - 1} \times (i - 1) + P_6 \tag{9}$$

Consider an example where the image size is ($5401 \times 3751$). Length of the diagonal is $D = 6576$. The value of both $P_1$ and $P_2$ is 0.1. We set the values of $P_3 = P_6 = P_7 = P_{10} = 0.04 \times D$. The depth corresponding to the first row and the 3D representation $\rho$ for the entire image are shown in Fig. 5(a) and (b), respectively

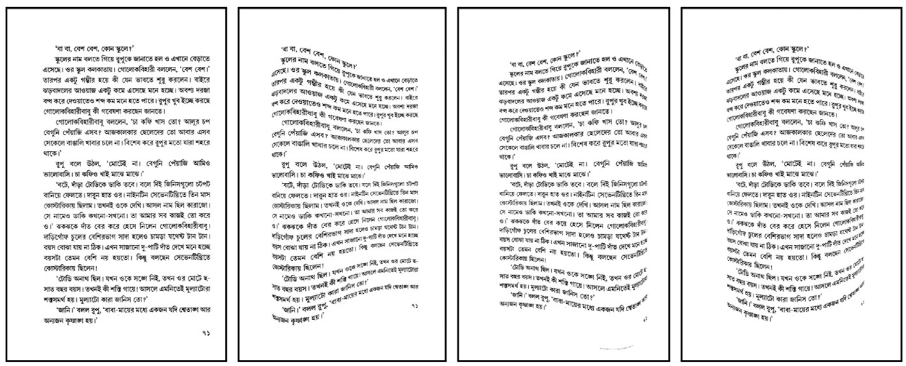
*Warped Image formation:* The input is a document image having a flat-bed surface. From the focal length of the camera and the distance between the lens and the document surface, $D_c$ is calculated. Next, the camera position ($C_p$), camera angle ($\phi$) are specified. The depth matrix ($\rho$) is obtained, as mentioned above. Warping factor ($\Delta_i$) at each point of the input image is estimated from the model parameters $\alpha$, $\gamma$, and $d$, and the corresponding image point is translated by an amount $\Delta_i$.

Generally, three types of warped documents are mostly available in the real world. They are (i) Type-I: Open book page having hump at left; (ii) Type-II: Open book page having hump at right; (iii) Type-III: Document pasted on Lamp-post. Our proposed model also generates these three types of warped images. Fig. 6 shows the outputs of the proposed warped model and used position and control parameters are tabulated in the Table 1. These parameters are set empirically. The source codes of the proposed synthetic image generation technique are available at https://github.com/ArpanGarai/Snthetic_warped_Image_generation.




(a) (b)

**Fig. 5.** An example of depth estimation: (a) Amount of depth at first row; (b) 3D view of depth for the entire document.



(a) (b) (c) (d)

**Fig. 6.** (a) Input image; Synthetically generated images: (b) Type I; (c) Type II; (d) Type III.

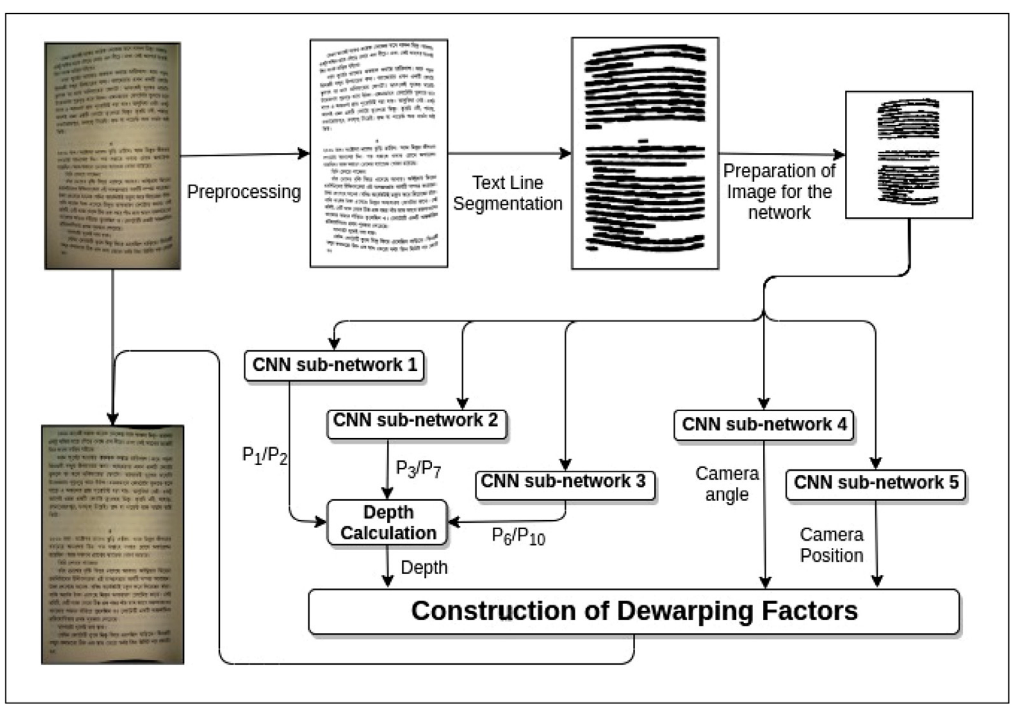



**Fig. 7.** Steps of the proposed dewarping model.

## 5. CNN based framework for dewarping

The warping of a document depends on the camera angle, camera position and the depth of the surface. In this paper, we propose a convolutional neural network (CNN) based dewarping model. During the training phase, the model learns to estimate aforesaid warping parameters. For this type of model, a huge amount of training samples are required, and we use the synthetic warped images (generated from our proposed warped model) for the training of the CNN model.

Experimentally, we have seen that a single network unit is not capable enough to produce the values of all the parameters with higher accuracy. So, we use five different sub-networks for estimating the parameters. Among five sub-networks, Sub-network 4 and 5 (Fig. 7) are used to estimate camera angle and camera position, respectively. Here, the depth is estimated using position parame-

**Table 2**
Configuration of the network.

| Sub-Net 1<br>17 weights Layers | Sub-Net 2/3<br>25 weights Layers | Sub-Net 4<br>19 weights Layers | Sub-Net 5<br>19 weights Layers |
|---|---|---|---|
| Input (150 × 150) | | | |
| 8 - Conv 1 (9 × 9) | 8 - Conv 1 (9 × 9) | 8 - Conv 1 (9 × 9) | 8 - Conv 1 (9 × 9) |
| 8 - Conv 2 (5 × 5) | 32 - Conv 2 (5 × 5) | 8 - Conv 2 (5 × 5) | 8 - Conv 2 (5 × 5) |
| 8 - Conv 3 (3 × 3) | 64 - Conv 3 (3 × 3) | 8 - Conv 3 (3 × 3) | 8 - Conv 3 (3 × 3) |
| . | 64 - Conv 4 (3 × 3) | . | . |
| Max pooling-(2 × 2)- Stride-2 | | | |
| 8 - Conv 4 (5 × 5) | 64 - Conv 5 (5 × 5) | 64 - Conv 4 (5 × 5) | 64 - Conv 4 (5 × 5) |
| 8 - Conv 5 (5 × 5) | 64 - Conv 6 (5 × 5) | 64 - Conv 5 (5 × 5) | 64 - Conv 5 (5 × 5) |
| 32 - Conv 6 (3 × 3) | 64 - Conv 7 (3 × 3) | 64 - Conv 6 (3 × 3) | 64 - Conv 6 (3 × 3) |
| . | 64 - Conv 8 (3 × 3) | . | . |
| Max pooling-(2 × 2)- Stride-2 | | | |
| 64 - Conv 7 (3 × 3) | 64 - Conv 9 (3 × 3) | 64 - Conv 7 (3 × 3) | 64 - Conv 7 (3 × 3) |
| 64 - Conv 8 (3 × 3) | 128 - Conv 10 (5 × 5) | 128 - Conv 8 (5 × 5) | 128 - Conv 8 (5 × 5) |
| 64 - Conv 9 (3 × 3) | 128 - Conv 11 (3 × 3) | 128 - Conv 9 (3 × 3) | 128 - Conv 9 (3 × 3) |
| 64 - Conv 10 (3 × 3) | 64 - Conv 12 (3 × 3) | 64 - Conv 10 (3 × 3) | 64 - Conv 10 (3 × 3) |
| 64 - Conv 11 (3 × 3) | 64 - Conv 13 (3 × 3) | 64 - Conv 11 (3 × 3) | 64 - Conv 11 (3 × 3) |
| . | 64 - Conv 14 (3 × 3) | 64 - Conv 12 (3 × 3) | 64 - Conv 12 (3 × 3) |
| . | 64 - Conv 15 (3 × 3) | 64 - Conv 13 (3 × 3) | 64 - Conv 13 (3 × 3) |
| . | 64 - Conv 16 (3 × 3) | . | . |
| . | 64 - Conv 17 (3 × 3) | . | . |
| Max pooling-(2 × 2)- Stride-2 | | | |
| 64 - Conv 12 (3 × 3) | 64 - Conv 18 (3 × 3) | 64 - Conv 14 (3 × 3) | 64 - Conv 14 (3 × 3) |
| 32 - Conv 13 (3 × 3) | 64 - Conv 19 (3 × 3) | 32 - Conv 15 (3 × 3) | 64 - Conv 15 (3 × 3) |
| 16 - Conv 14 (3 × 3) | 64 - Conv 20 (3 × 3) | 16 - Conv 16 (3 × 3) | 32 - Conv 16 (3 × 3) |
| 8 - Conv 15 (3 × 3) | 64 - Conv 21 (3 × 3) | 8 - Conv 17 (3 × 3) | 8 - Conv 17 (3 × 3) |
| 3 - Conv 16 (3 × 3) | 32 - Conv 22 (3 × 3) | 3 - Conv 18 (3 × 3) | 3 - Conv 18 (3 × 3) |
| . | 16 - Conv 23 (3 × 3) | . | . |
| . | 9 - Conv 24 (3 × 3) | . | . |
| Fully Connected Layer-1 | Fully Connected Layer-1 | Fully Connected Layer-1 | Fully Connected Layer-1 |
| Regression Layer | | | |

ters and control parameters. The sub-network 1 is used to find the value of position parameters, and the sub-networks 2 and 3 are used to estimate the values of control parameters. Next, the trained model is tested on both synthetic and real captured images. Before feeding to the networks, the images are pre-processed, and text-lines are segmented. Steps of the proposed de-warping model are shown in Fig. 7.

### 5.1. Pre-processing

The preprocessing step mainly consists of three tasks, binarization, de-noising, and line segmentation of the input document images. The method proposed in [30] is used for binarization of the input images. After binarization, the next task is to remove the noise from the images. The noises present in the images are mostly border noises, and to remove this type of noise, we use an existing noise removal technique proposed in [31].

There are a few algorithms available for line segmentation of the warped images. Here, we use the line segmentation technique proposed in [13]. The method first calculates the minimum bounding rectangle (MBR) of connected components. These MBRs are further analyzed, and a Minimum Spanning Tree (MST) is generated. The text lines are segmented using the generated MST. The text line segmented image undergoes a morphological closing operation using a line like structuring element (as shown in Fig. 7) such that each connected component represents a text line.

### 5.2. Preparation of input to the network

The size of the input images of all the sub-networks is 150 × 150. So, all the morphological closed images are scaled down such that it satisfies the following equation: $\frac{h}{w} \approx \frac{H}{W}$ and $140 \leq \max(h, w) < 150$. Here, $h \times w$ and $H \times W$ are the sizes of scaled image and original image, respectively. Next, we take a blank image of size 150 × 150. Then, we place the scaled image inside the blank image in such a way that the point of intersections of two diagonals of two images coincides.

### 5.3. Network architecture

Recent literature [32–34] suggests that a modified version of the VGG-16 network is capable enough to solve a variety of problems in the domain of image analysis. Hence, all five sub-networks in the proposed approach are the modified versions of the VGG16 network. The sub-net 1 is used to find the position parameters. Here, we assume the value of two-position parameters is the same. It is also assumed that the values of the depths at every row are the same. In other words, we take $P_3 = P_7$ and $P_6 = P_{10}$. So, two sub-networks (2 and 3) are used to predict parameters $P_3$ and $P_{10}$, respectively. Subnet 4 is used to determine the angle $\phi$. The sub-net 5 is used for predicting the camera position ($C_p$). The detailed description of the subnetworks is given in Table 2. Here, each convolutional layer is followed by a RELU and a batch normalization layer. We have used the SGDM optimizer.

To reduce the number of parameters, we have used only one fully-connected layer, followed by a regression layer. Next, depending on the application, the number of filters in the convolution layer, size of the convolution window, stride size, size of the window in the max-pooling layer, etc. are modified.

In each of this sub-network, we initialize the weights randomly. More than 10000 synthetic images are used to train and validate the network. Among them, 75% used during training, 20% used for validation. Remaining images are used for testing.

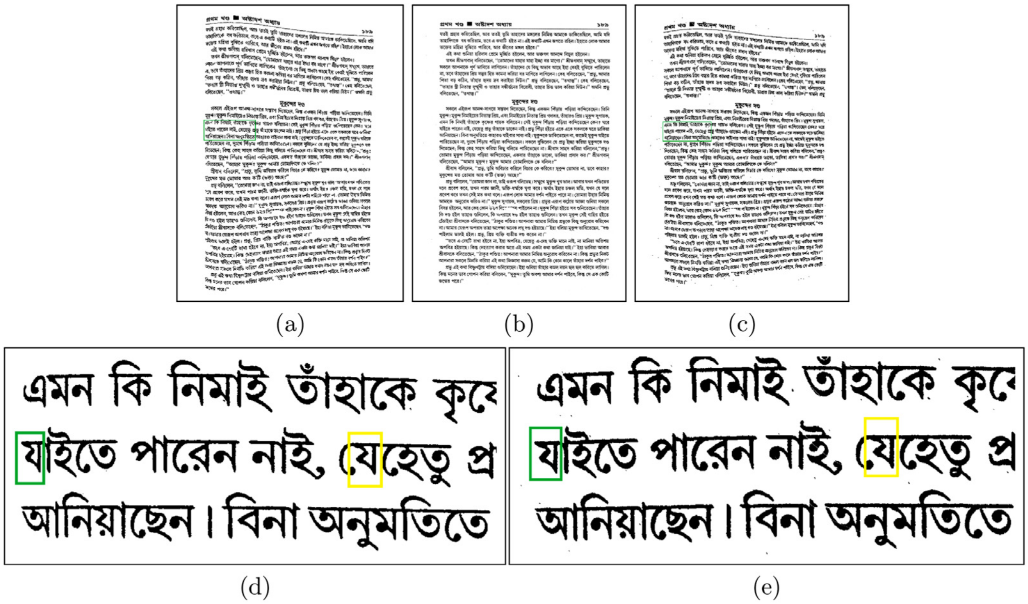



**Fig. 8.** Analysis of proposed model: (a) Camera captured warped image; (b) Flat-bed scanned image; (c) Synthetically warped image; (d) Enlarged crop of Fig. 8(a); (e) Enlarged crop of Fig. 8(c).

### 5.4. Reconstruction of the warping factors

The values of $C_p$, $\phi$ and $\rho$ are estimated using CNN. Most of the documents are captured from a distance varies from 35 cm to 50 cm. In the proposed approach, we assume that this distance as the multiple of the document height. Here, we chose the distance ($D_c$) to be $1.5 \times H$. Here, $H$ is the height of the image. Next, the values $\alpha$, $\gamma$ and $d$ are calculated. The value $\Delta_i$ is calculated from $\alpha$, $\gamma$ and $d$. The dewarping factors ($\Delta'_i$) is obtained using the following equation: $\Delta'_i(i + \Delta_i(i, j), j) = -\Delta_i(i, j)$. The image is dewarped using these dewarping factors. Source codes of the proposed dewarping technique are available at https://github.com/ArpanGarai/dewarpingCNNgeoparam.

## 6. Experimental results and evaluation

At first, we analyze the proposed synthetic image generation technique in both qualitative and quantitative ways. Next, the dewarping approach is evaluated, and the proposed method is compared with the existing techniques.

### 6.1. Dataset

As mentioned earlier, there are two publicly available warped document image datasets, namely, *DFKI document image contest dataset* [8] and *Doc3D* dataset [4]. These datasets consist of warped document images containing *English* script. There is no dataset for Alpha-syllabary script like *Bangla*. So, we have created a warped document image dataset (WDID) which contains 258 warped images. The images of our dataset have different scripts such as *Bangla, Devanagari, Gurumukhi*, and some of the images have mixed scripts. The dataset contains various types of warped document images, like document pasted on a lamppost, document hanging on a notice board, etc. These varieties are not present in the publicly available datasets. The images in WDID are captured using either digital stand-alone camera or camera attached to a smart mobile phone. The images are taken from various distances. The images are captured in various lighting conditions. Some pictures are taken at the presence of sunlight, whereas other images are captured at night, where the document is illuminated by neon/LED light. Different fonts, types, sizes, and styles are present in the text content of the images. The resolution of most of the images is more than $3000 \times 4000$.

### 6.2. Analysis of synthetic warped image generation model

To analyze the proposed synthetic warped image generation model, we take a camera capture pre-processed warped document image (Fig. 8(a)). The same document is scanned with a flat-bed scanner and fed to the model, which generates a warped image similar in nature to the image, as shown in Fig. 8(a). The input and output images are shown in Fig. 8(b) and (c), respectively.

It is evident from Fig. 8 that the nature of the warping of the camera captured the warped image, and the output of our proposed model is very much similar. For better visualization, portions of both the warped images are cropped and enlarged. These cropped and enlarged images are shown in Fig. 8(d) and (e).

To evaluate the approach quantitatively, we use two measures of structural similarity index measure (SSIM) proposed in [35] and multi-scale structural similarity index measure (MS-SSIM) [36]. These measures are used to compare the structural similarities of the two images. In our experiment, we have taken two images, one from the real warped dataset and the other is the output of our warped image generation model to measure SSIM and MS-SSIM. These two images are visually similar structures. The SSIM and MS-SSIM can take any value from the range of $0 - 1$. If the value of SSIM (MS-SSIM) is 1, then the two images under test have an exactly similar structure.

We have considered three types of warped images as in Section 4. In our experiment, 30 pairs of images are considered (10 pairs from each type of the warped image) and we get average values of SSIM and MS-SSIM is 0.956 and 0.913, respectively. The value SSIM (MS-SSIM) is quite encouraging.

### 6.3. Evaluation of dewarping algorithm

The proposed dewarping algorithm is evaluated in both qualitative and quantitative ways. To evaluate the performance of the proposed method, we use three datasets, which are mentioned in Section 6.1.

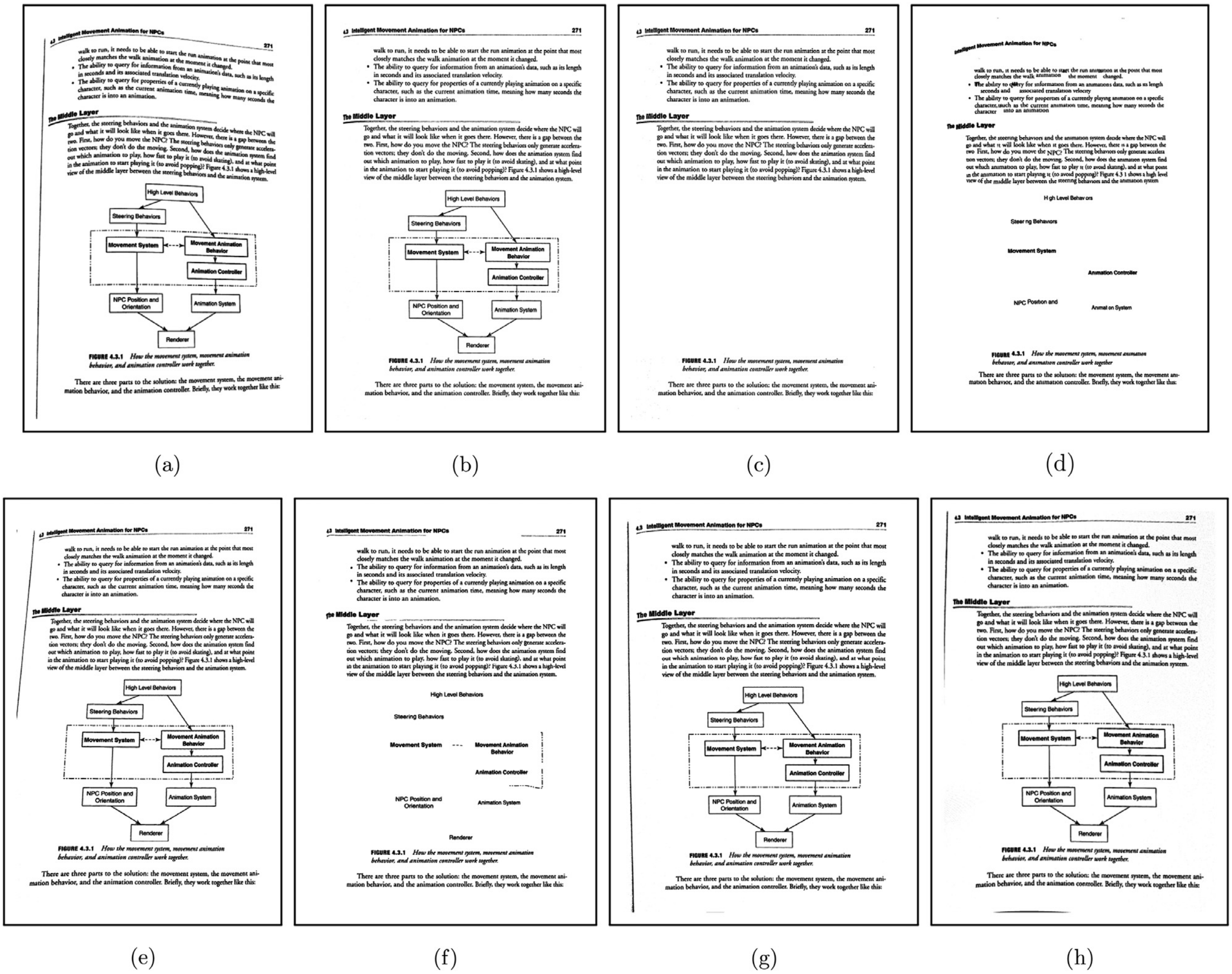


**Fig. 9.** An example of qualitative comparison between proposed method and the existing methods: (a) Input image from DFKI document image contest dataset; (b) Output of the method proposed in [37] (CTM); (c) Output of the method proposed in [37] (CTM2); (d) Output of the method proposed in [23] (SEG); (e) Output of the method proposed in [38] (SKEL); (f) Output of the method proposed in [39] (Coupled Snakes); (g) Output of the method proposed in [40] (by Meng et al.); (h) Output of the proposed Method.

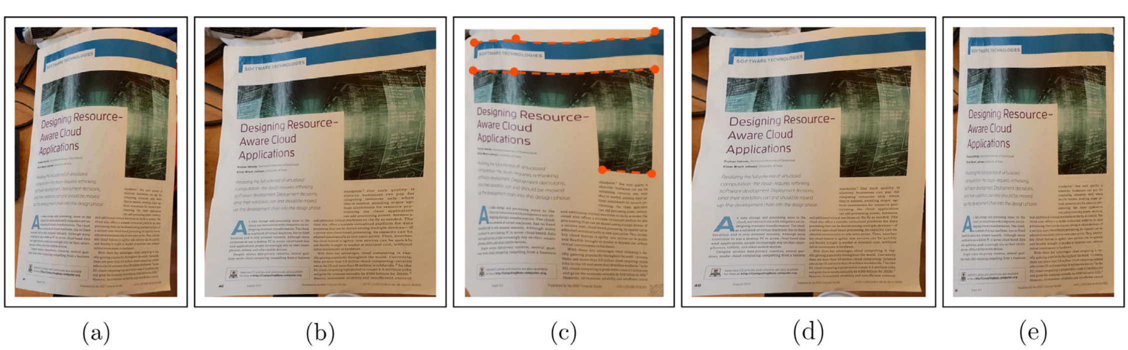
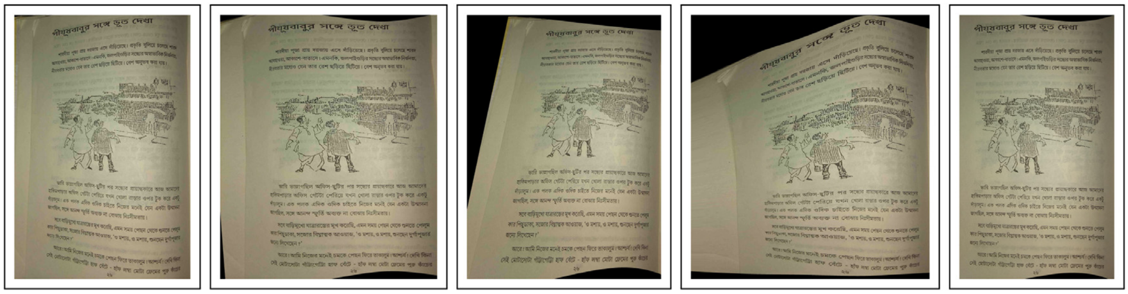


**Fig. 10.** An example of qualitative comparison between proposed method and the existing methods: (a) Input image from Doc3D dataset; (b) Output of the method proposed by Kim et al. [25]; (c) Output of the method proposed by Ma et al. (DocUNet) [4]; (d) Output of the method proposed by Meng et al. [40]; (e) Output of the proposed method.

The performance of the proposed dewarping method is compared qualitatively with the some of the existing methods using a set of three 3 examples as shown in Fig. 9–11.

It is clear from Fig. 9 that the output of the proposed method is either better or similar to other methods. It is evident from the example shown in Fig. 10 that the performance of the proposed method is better than the other mentioned existing methods. It is also clear from Fig. 11 that the proposed method outperforms the other existing methods. To produce the output images of the existing approaches, we have used the executable codes available on the official website of the respective authors.

The quantitative performance evaluation is done generally in two ways, as proposed in [41]. They are the indirect way and the direct way. In the indirect way, the OCR is used to evaluate the performance of the dewaping algorithms. Here, we have used the *Google Doc OCR* as presented in [42]. The accuracy of the OCR for a



(a) (b) (c) (d) (e)

**Fig. 11.** An example of qualitative comparison between proposed method and the existing methods: (a) Input image from WDID dataset; (b) Output of the method proposed by Kim et al. [25]; (c) Output of the method proposed by Kim et al. [26]; (d) Output of the method proposed by Meng et al. [40]; (e) Output of the proposed method.

**Table 3**
Performance analysis using OCR in WDID.

| WDID | Key feature | Number of images | average $\frac{C_1}{C_2}$ | | | |
|---|---|---|---|---|---|---|
| | | | Kim et al. [25] | Kil et al. [26] | Meng et al. [40] | Proposed approach |
| Set 1 | Fold at right | 80 | 0.51 | 0.84 | 0.641 | 0.986 |
| Set 2 | Fold at left | 60 | 0.55 | 0.83 | 0.675 | 0.987 |
| Set 3 | Fold at Centre | 81 | 0.50 | 0.86 | 0.663 | 0.991 |
| Set 4 | Bangla script | 221 | 0.52 | 0.82 | 0.645 | 0.983 |
| Set 5 | Mixed script | 20 | 0.54 | 0.85 | 0.645 | 0.984 |
| Set 6 | Devanagari script | 10 | 0.52 | 0.83 | 0.652 | 0.981 |
| Set 7 | Text-only | 205 | 0.53 | 0.85 | 0.681 | 0.990 |
| Set 8 | Text with non-text | 46 | 0.53 | 0.86 | 0.641 | 0.976 |

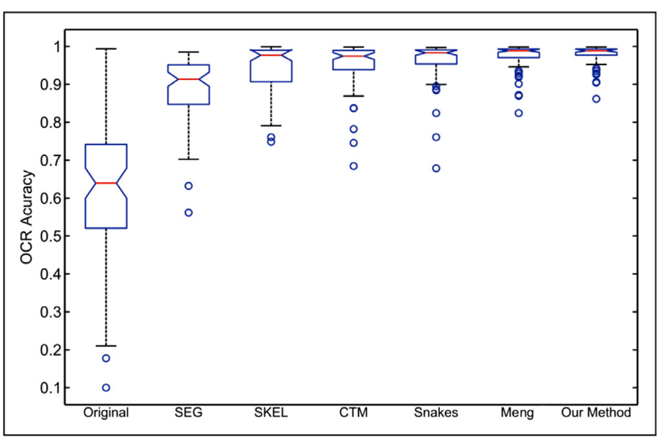



**Fig. 12.** The values of $\frac{C_1}{C_2}$ on 'DFKI dewarping contest dataset'.

particular dewarped image is $C_1$, and the same for the corresponding ground truth image is $C_2$. The ratio of $C_1$ and $C_2$ is an indirect way to measure the performance of the dewarping method. If this ration is high, that means the performance of the dewarping method is good. The performance of the proposed method, along with performances of the other existing methods in terms of the ratio $\frac{C_1}{C_2}$ on 'DFKI dewarping context dataset', is presented using the box plot which is illustrated in Fig 12. It is clear from Fig 12 that the performance of the proposed method is better than the existing methods. In Table 3, the performance of the proposed dewarping method as well as three existing script independent methods like [25,26] and [40] on document images containing Alphasyllabary (*Bangla/ Devanagari*) scripts is presented.

The indirect way of measurement does not give a quantitative measure of the visual correctness of the dewarped image. Although an approach to evaluate visual correctness of the dewarping methods [41] it does not check the slant error correction. Therefore, the performance is evaluated in terms of restoration accuracy for the italic and bold type characters ($\alpha_i$ and $\beta_i$), which is also used in [42]. The $\alpha_i$ and $\beta_i$ are defined as $\alpha_i = \frac{\theta_d}{\theta_g} \times 100\%$ and $\beta_i = \frac{B_d}{B_g} \times 100\%$, respectively. Where, slant angles w.r.t vertical axis of the same italic characters in dewarped images and its equivalent ground truth images is given by $\theta_d$ and $\theta_g$, respectively. Here, $B_d$ and $B_g$ are stroke widths of the same characters in dewarped image and its equivalent ground truth image, respectively. The accuracy of the proposed approach along with the other three script independent approaches using ($\alpha_i$ and $\beta_i$), are shown in Table 4. It is clear from the table that the proposed method outperforms the others.

$\alpha_i$ does not give the accuracy of non-italic fonts. So, we have used the mean local slant error percentage ($\gamma_l$), which is introduced in [43]. The index, given as $\gamma_l = \frac{\sum_{i=1}^{N} |\theta_l(i)|}{N} \times 100$, is based based on the angle, $\theta_l$, between right profile and the verticle axis. Only the characters (non-italic) having a linear right profile, which makes an angle of $\pi/2$ with the horizontal direction, are considered. Here, $N$ is the total number of such components in a particular dewarped image. The performance of the proposed approach, along with existing approaches using $\gamma_l$, is shown in Table 4. A smaller value of $\gamma_l$ suggests the better performance of the respective algorithm. It is clear from the table that the performance of the proposed method is better than the others.

### 6.4. Limitation of the proposed method

The proposed method is designed for single folded warped documents. So, it may not give satisfactory results for multiple folded document images. However, this approach can be further extended to dewarp the multiple folded document images. Here, we assume that the depth of the document follows one of the three patterns. The more complex network can be used to dewarp documents with arbitrary depths. Generally, people try to capture documents keeping the virtual image plane parallel to the document surface. So, we assume that the camera angle lies within $-1.5^o$ to $1.5^o$.

**Table 4**
Restoration accuracy using $\alpha_i$, $\beta_i$, $\gamma_l$ and $C_{rms}$.

| Methods | $\beta_i$ | | $\alpha_i$ | | $\gamma_l$ | $C_{rms}$ |
|---|---|---|---|---|---|---|
| | Bold | Bold + Italic | Italic | Bold + Italic | | |
| Kim et al. [25] | 88.1 | 80.5 | 61.1 | 60.5 | 14.5 | 5.2 |
| Kil et al. [26] | 89.3 | 87.3 | 90.3 | 89.6 | 7.5 | 1.8 |
| Meng et al. [40] | 88.7 | 86.4 | 78.4 | 84.4 | 13.4 | 4.1 |
| Proposed Approach | 91.8 | 92.5 | 92.3 | 92.5 | 0.47 | 0.19 |

Hence, our network may not produce good results for images captured in a higher camera angle.

## 7. Conclusion

In this work, we introduce a mathematical model of warping. This model is used to generate different types of synthetic images from an image having a flat-bed surface. We also present a CNN based dewarping method. The synthetic images generated from our proposed warping model are used to train a CNN. The CNN model takes only the 2D image and estimates the warping parameters which are used for dewarping. The performance of both models is evaluated, and the results are encouraging. This dewarping method helps to improve the performance of OCR and other document processing software. In the future, an extended version of the proposed approach may be used to handle core complex types of warping like multiple folded document images, images captured with the higher camera angle, images with a high degree of curl, etc.

Arpan Garai is pursuing a Ph.D. degree from the Department of Computer Science and Technology, Indian Institute of Engineering Sciences and Technology, Shibpur. He received his BE from the department of Computer Science and Engineering, University Institute of Technology, Burdwan University in 2011. Next in 2013, he has done his M Tech from the Department of Computer Science and Engineering, Kalyani Government Engineering College, WBUT. Then he worked as a project linked person in Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata. Next, he was an assistant professor at Pailan College of Management and Technology, WBUT. His research interest includes machine learning, image processing, computer vission and pattern recognition. Samit Biswas is Assistant Professor in the Department of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah. He received his PhD in Computer Science and Technology from Indian Institute of Engineering Science and Technology, Shibpur, Howrah after completing B.E. and M.Tech. He has authored/coauthored several research papers in various International Journals and Conferences. He is an active member of the board of reviewers in various International Journals and Conferences. Currently, his research interests include machine learning, image processing and pattern recognition, computational intelligence, and machine based translation. Sekhar Mandal did his B.Tech. and M.Tech. from University of Calcutta, India, and his PhD from Bengal Engineering and Science University, Shibpur, Howrah, India. He is currently a Professor in Computer Science and Technology Department of Bengal Engineering and Science University, Shibpur, Howrah, India. His research interest mainly lies in digital image processing and pattern recognition. So far he has published 50 research papers in international journals, edited volumes, and refereed conference proceedings.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patcog.2020.107621.

## References

[1] Z. Cao, J. Lu, S. Cui, C. Zhang, Zero-shot handwritten chinese character recognition with hierarchical decomposition embedding, Pattern Recognit. 107 (2020) 107488, doi:10.1016/j.patcog.2020.107488.
[2] W. Sihang, W. Jiapeng, M. Weihong, J. Lianwen, Precise detection of chinese characters in historical documents with deep reinforcement learning, Pattern Recognit. 107 (2020) 107503, doi:10.1016/j.patcog.2020.107503.
[3] M. Yousef, K.F. Hussain, U.S. Mohammed, Accurate, data-efficient, unconstrained text recognition with convolutional neural networks, Pattern Recognit. 108 (2020) 107482, doi:10.1016/j.patcog.2020.107482.
[4] M. Ke, S. Zhixin, X. Bai, W. Jue, S. Dimitris, DocUNet: document image unwarping via a stacked U-Net, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[5] S. Das, G. Mishra, A. Sudharshana, R. Shilkrot, The common fold: utilizing the four-fold to dewarp printed documents from a single image, in: Proceedings of the 2017 ACM Symposium on Document Engineering, in: DocEng '17, ACM, New York, NY, USA, 2017, pp. 125–128, doi:10.1145/3103010.3121030.
[6] Q.A. Bui, D. Mollard, S. Tabbone, Automatic synthetic document image generation using generative adversarial networks: application in mobile-captured document analysis, in: International Conference on Document Analysis and Recognition, 2019.
[7] Z. Zhang, Y. Zeng, L. Bai, Y. Hu, M. Wu, S. Wang, E.R. Hancock, Spectral bounding: strictly satisfying the 1-Lipschitz property for generative adversarial networks, Pattern Recognit. 105 (2020) 107179, doi:10.1016/j.patcog.2019.107179.
[8] S. Faisal, B. Thomas M, Document image dewarping contest, in: 2nd International Workshop on Camera-Based Document Analysis and Recognition, 2007, pp. 181–188.
[9] M.S. Brown, W.B. Seales, Image restoration of arbitrarily warped documents, IEEE Trans. Pattern Anal. Mach. Intell. 26 (10) (2004) 1295–1306, doi:10.1109/TPAMI.2004.87.
[10] A. Yamashita, A. Kawarago, T. Kaneko, K.T. Miura, Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system, in: Proceedings of the 17th International Conference on Pattern Recognition, vol. 1, 2004, pp. 482–485Vol.1, doi:10.1109/ICPR.2004.1334171.
[11] B. FU, W. LI, M. WU, R. LI, Z. XU, A document rectification approach dealing with both perspective distortion and warping based on text flow curve fitting, Int. J. Image Graph. 12 (01) (2012) 1250002, doi:10.1142/S0219467812500027.
[12] S. Das, K. Ma, Z. Shu, D. Samaras, R. Shilkrot, DewarpNet: single-image document unwarping with stacked 3d and 2d regression networks, in: The IEEE International Conference on Computer Vision, 2019.
[13] C. Liu, Y. Zhang, B. Wang, X. Ding, Restoring camera-captured distorted document images, Int. J. Doc. Anal. Recogn. 18 (2) (2015) 111–124, doi:10.1007/s10032-014-0233-8.
[14] H. Cao, X. Ding, C. Liu, A cylindrical surface model to rectify the bound document image, in: Proceedings Ninth IEEE International Conference on Computer Vision, 2003, pp. 228–233 vol.1, doi:10.1109/ICCV.2003.1238346.
[15] L. Zhang, C.L. Tan, Restoring warped document images using shape-from-shading and surface interpolation, in: 18th International Conference on Pattern Recognition, vol. 1, 2006, pp. 642–645, doi:10.1109/ICPR.2006.997.
[16] G. Meng, C. Pan, S. Xiang, J. Duan, Metric rectification of curved document images, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 707–722, doi:10.1109/TPAMI.2011.151.
[17] J. Liang, D. DeMenthon, D. Doermann, Geometric rectification of camera-captured document images, IEEE Trans. Pattern Anal. Mach. Intell. 30 (4) (2008) 591–605, doi:10.1109/TPAMI.2007.70724.
[18] S. You, Y. Matsushita, S. Sinha, Y. Bou, K. Ikeuchi, Multiview rectification of folded documents, IEEE Transactions on Pattern Analysis and Machine Intelligence PP (99) (2017), doi:10.1109/TPAMI.2017.2675980. 1–1
[19] Y. He, P. Pan, S. Xie, J. Sun, S. Naoi, A book dewarping system by boundary-based 3d surface reconstruction, in: 2013 12th International Conference on Document Analysis and Recognition, 2013, pp. 403–407, doi:10.1109/ICDAR.2013.88.
[20] H. Ezaki, S. Uchida, A. Asano, H. Sakoe, Dewarping of document image by global optimization, in: Eighth International Conference on Document Analysis and Recognition, 2005, pp. 302–306Vol. 1, doi:10.1109/ICDAR.2005.87.

[21] S. Lu, C.L. Tan, Document flattening through grid modeling and regularization, in: 18th International Conference on Pattern Recognition, vol. 1, 2006, pp. 971–974, doi:10.1109/ICPR.2006.458.
[22] A. Ulges, C.H. Lampert, T.M. Breuel, Document image dewarping using robust estimation of curled text lines, in: Eighth International Conference on Document Analysis and Recognition, 2005, pp. 1001–1005Vol. 2, doi:10.1109/ICDAR.2005.90.
[23] B. Gatos, I. Pratikakis, K. Ntirogiannis, Segmentation based recovery of arbitrarily warped document images, in: Ninth International Conference on Document Analysis and Recognition, vol. 2, 2007, pp. 989–993, doi:10.1109/ICDAR.2007.4377063.
[24] N. Stamatopoulos, B. Gatos, I. Pratikakis, S.J. Perantonis, Goal-oriented rectification of camera-based document images, IEEE Trans. Image Process. 20 (4) (2011) 910–920, doi:10.1109/TIP.2010.2080280.
[25] B.S. Kim, H.I. Koo, N.I. Cho, Document dewarping via text-line based optimization, Pattern Recognit. 48 (11) (2015) 3600–3614 https://doi.org/10.1016/j.patcog.2015.04.026.
[26] T. Kil, W. Seo, H.I. Koo, N.I. Cho, Robust document image dewarping method using text-lines and line segments, in: 2017 14th IAPR International Conference on Document Analysis and Recognition, vol. 01, 2017, pp. 865–870, doi:10.1109/ICDAR.2017.146.
[27] P. Yang, Effective geometric restoration of distorted historical document for large-scale digitisation, IET Image Proc. 11 (12) (2017) 841–853.
[28] X. Liu, G. Meng, B. Fan, S. Xiang, C. Pan, Geometric rectification of document images using adversarial gated unwarping network, Pattern Recognit. 108 (2020) 107576, doi:10.1016/j.patcog.2020.107576.
[29] V.C. Kieu, N. Journet, M. Visani, R. Mullot, J.P. Domenger, Semi-synthetic document image generation using texture mapping on scanned 3d document shapes, in: 2013 12th International Conference on Document Analysis and Recognition, 2013, pp. 489–493, doi:10.1109/ICDAR.2013.104.
[30] B. Su, S. Lu, C.L. Tan, Robust document image binarization technique for degraded document images, IEEE Trans. Image Process. 22 (4) (2013) 1408–1417, doi:10.1109/TIP.2012.2231089.
[31] S.S. Bukhari, F. Shafait, T.M. Breuel, Border noise removal of camera-captured document images using page frame detection, in: M. Iwamura, F. Shafait (Eds.), Camera-Based Document Analysis and Recognition, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 126–137.
[32] S. Xie, H. Hu, Y. Wu, Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition, Pattern Recognit. 92 (2019) 177–191, doi:10.1016/j.patcog.2019.03.019.
[33] Z. Tu, W. Xie, Q. Qin, R. Poppe, R.C. Veltkamp, B. Li, J. Yuan, Multi-stream CNN: learning representations based on human-related regions for action recognition, Pattern Recognit. 79 (2018) 32–43, doi:10.1016/j.patcog.2018.01.020.
[34] Q. Zhang, Y. Shi, X. Zhang, Attention and boundary guided salient object detection, Pattern Recognit. 107 (2020) 107484, doi:10.1016/j.patcog.2020.107484.
[35] Zhou Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612, doi:10.1109/TIP.2003.819861.
[36] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment, in: The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003, vol. 2, 2003, pp. 1398–1402, doi:10.1109/ACSSC.2003.1292216.
[37] B. Fu, M. Wu, R. Li, W. Li, Z. Xu, C. Yang, A model-based book dewarping method using text line detection, 2nd International. Workshop on Camera-Based Document Analysis and Recognition, 2007.
[38] A. Masalovitch, L. Mestetskiy, Usage of continuous skeletal image representation for document images dewarping, 2007.
[39] S.S. Bukhari, F. Shafait, T.M. Breuel, TM: Dewarping of document images using coupled-snakes, in: Proceedings of Third International Workshop on Camera-Based Document Analysis and Recognition, 2009, pp. 34–41.
[40] G. Meng, Y. Su, Y. Wu, S. Xiang, C. Pan, Exploiting vector fields for geometric rectification of distorted document images, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision–ECCV 2018, Springer International Publishing, Cham, 2018, pp. 180–195.
[41] N. Stamatopoulos, B. Gatos, I. Pratikakis, Performance evaluation methodology for document image dewarping techniques, IET Image Proc. 6 (7) (2012) 738–745.
[42] A. Garai, S. Biswas, S. Mandal, B.B. Chaudhuri, Automatic dewarping of camera captured born-digital Bangla document images, in: 2017 Ninth International Conference on Advances in Pattern Recognition, 2017, pp. 1–6, doi:10.1109/ICAPR.2017.8593157.
[43] G. Arpan, B. Samit, M. Sekhar, B.B. Chaudhuri, Automatic rectification of warped Bangla document images, IET Image Proc. 14 (9) (2020) 74–83.

**Arpan Garai** is pursuing a Ph.D. degree from the Department of Computer Science and Technology, Indian Institute of Engineering Sciences and Technology, Shibpur. He received his BE from the department of Computer Science and Engineering, University Institute of Technology, Burdwan University in 2011. Next in 2013, he has done his M Tech from the Department of Computer Science and Engineering, Kalyani Government Engineering College, WBUT. Then he worked as a project linked person in Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata. Next, he was an assistant professor at Pailan College of Management and Technology, WBUT. His research interest includes machine learning, image processing, computer vission and pattern recognition.

**Samit Biswas** is Assistant Professor in the Department of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah. He received his Ph.D. in Computer Science and Technology from Indian Institute of Engineering Science and Technology, Shibpur, Howrah after completing B.E. and M.Tech. He has authored/coauthored several research papers in various International Journals and Conferences. He is an active member of the board of reviewers in various International Journals and Conferences. Currently, his research interests include machine learning, image processing and pattern recognition, computational intelligence, and machine based translation.

**Sekhar Mandal** did his B.Tech. and M.Tech. from University of Calcutta, India, and his PhD from Bengal Engineering and Science University, Shibpur, Howrah, India. He is currently a Professor in Computer Science and Technology Department of Bengal Engineering and Science University, Shibpur, Howrah, India. His research interest mainly lies in digital image processing and pattern recognition. So far he has published 50 research papers in international journals, edited volumes, and refereed conference proceedings.