

# Exploiting Vector Fields for Geometric Rectification of Distorted Document Images

Gaofeng MENG<sup>1,3</sup>, Yuanqi SU<sup>2</sup>, Ying WU<sup>3</sup>, Shiming XIANG<sup>1</sup>,  
Chunhong PAN<sup>1</sup>

<sup>1</sup> National Laboratory of Pattern Recognition, Institute of Automation,  
Chinese Academy of Sciences, Beijing, China

{gfmeng, smxiang, chpan}@nlpr.ia.ac.cn

<sup>2</sup> Xi'an Jiaotong University, Xi'an, Shaanxi, China

yuanqisu@mail.xjtu.edu.cn

<sup>3</sup> Northwestern University, Evanston, USA

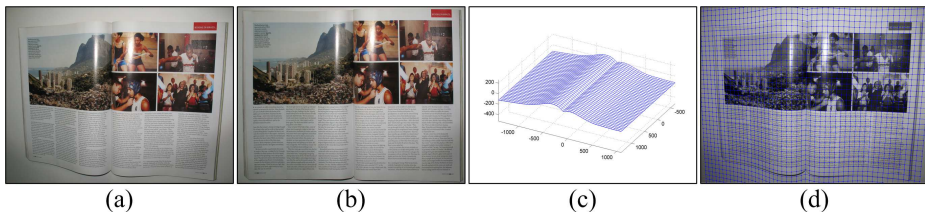
yingwu@northwestern.edu

**Abstract.** This paper proposes a segment-free method for geometric rectification of a distorted document image captured by a hand-held camera. The method can recover the 3D page shape by exploiting the intrinsic vector fields of the image. Based on the assumption that the curled page shape is a general cylindrical surface, we estimate the parameters related to the camera and the 3D shape model through weighted majority voting on the vector fields. Then the spatial directrix of the surface is recovered by solving an ordinary differential equation (ODE) through the Euler method. Finally, the geometric distortions in images can be rectified by flattening the estimated 3D page surface onto a plane. Our method can exploit diverse types of visual cues available in a distorted document image to estimate its vector fields for 3D page shape recovery. In comparison to the state-of-the-art methods, the great advantage is that it is a segment-free method and does not have to extract curved text lines or textual blocks, which is still a very challenging problem especially for a distorted document image. Our method can therefore be freely applied to document images with extremely complicated page layouts and severe image quality degradation. Extensive experiments are implemented to demonstrate the effectiveness of the proposed method.

**Keywords:** Document Image processing, Geometric rectification, Vector fields, 3D shape recovery, OCR

## 1 Introduction

Recent decades have witnessed the increasing popularity of using portable cameras in the digitization of paper documents [1]. In comparison with the traditional flat-bed scanners, the use of portable cameras, *e.g.*, smartphone cameras and compact cameras, provides many great advantages. For example, they are portable, fast response and can be applied flexibly to documents with different sizes. However, the images captured by a hand-held camera often suffer from



**Fig. 1.** An example of geometric rectification of a distorted document image by our method. (a) the distorted document image, (b) the rectification result, (c) the reconstructed 3D page surface, (d) the constructed mesh grid for image dewarping.

serious geometric distortions due to page curl and perspective of camera. This generally happens when one captures the images of an opened thick and bound book. Many sophisticated methods for document image analysis and recognition, *e.g.*, OCR and page layout analysis, are vulnerable to the geometric distortions in images. Therefore, removing the geometric distortions is often a critical and indispensable preprocessing step for many tasks related to camera-based document images recognition [1, 2].

Up to now, many efforts have been made to address this challenging issue. According to how a dewarping mapping is derived, the existing methods can be roughly categorized into two groups: the image-based methods [3–8] and the model-based methods [9–19]. The former adopts a local [6–8] or global dewarping transform [3–5] directly derived from image visual cues, *e.g.*, the curved horizontal text lines or the document boundaries, to rectify the distortions. These methods can generally produce a desirable result with straightened text lines, which is quite OCR friendly. However, the methods cannot completely remove the distortions in images, since no 3D shape model is introduced to account for the physical page distortion.

Later works pay more attention to the model-based methods and some significant advances have been made [10, 9, 13, 14, 16, 19, 18]. The model-based methods introduce a 3D page surface model together with a camera model to explain the geometric distortions in a document image. These methods differ in how the 3D page shape is accurately recovered. Some early methods use shape-from-shading techniques [16, 19] to estimate the 3D page shape. These methods work desirably for scanned document images, in which the illumination is well controlled in the scanning process. However, they generally fail to images captured by hand-held cameras, since the environmental illumination is extremely complicated in reality.

The more robust approaches for page shape recovery are to use a 3D scanner [9, 20, 15] or multiple images captured from different viewpoints [11, 18]. The advantage of these methods is that they can be applied to document images with much complicated geometric distortions. However, the use of additional hardware and images makes them less attractive in the research. In contrast, the methods based on a single document image exploit the extracted visual

cues, *e.g.*, the curved text lines, to recover the 3D page shape. Representative works include [10, 13, 17, 14]. One disadvantage of these methods is that they have to first detect and segment the curved text lines in the images. However, this is generally quite challenging especially for a document image with severe distortions. Moreover, these methods are also vulnerable to complicated page layouts. This is especially true for document images with sparse text lines or large areas of non-textual contents.

In this paper, we propose a method for geometric rectification of a curved document image by exploiting its intrinsic vector fields. By assuming the page shape is a general cylindrical surface, we recover the parameters of camera and page shape information through weighted majority voting on the vector fields. Finally, the image distortion is rectified by flattening the estimated page surface. Our method is a segment-free approach and does not require to extract the curved horizontal text lines or textual blocks, which is still an open problem due to many challenging factors [1, 21]. The proposed method is thus very robust to document layouts and various types of image quality degradation that commonly happen to a camera-captured document image, including severe image distortion, non-uniform shading, cluttered backgrounds and serious image blur and noises. Fig.1 illustrates an example of the rectification results and the reconstructed 3D page shape of our method for a single distorted document image captured by a hand-held camera.

## 2 Approach

### 2.1 Vector fields and Assumptions

The vector field of a document image essentially originates from the curled page surface under perspective projection of camera [22]. Generally, the document contents, *e.g.*, text lines, figures or tables, are arranged and printed along a group of straight and parallel baselines. These baselines are normally invisible but can be inferred from the contents of the document [21]. Once the document page is curled and projected onto the image plane, the underlying baselines get distorted, yielding a deformed vector field, the vector of which at every point gives the tangent direction of the curved baseline. Consequently, the vector field of a document image encodes the 3D page shape information and the parameters of camera.

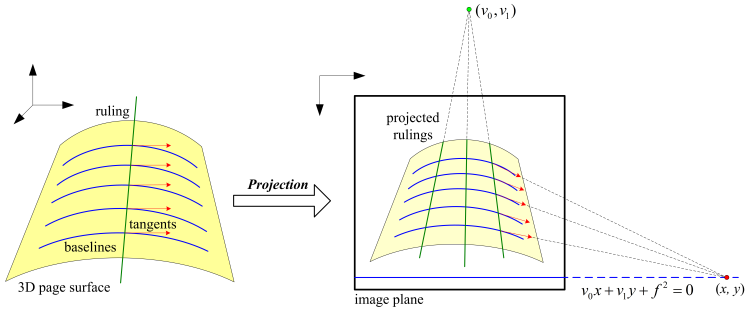
To recover the 3D page shape information from the vector field, we have to first introduce some basic assumptions. Firstly, we assume the curled page shape is a general cylindrical surface (GCS). This assumption, which is very suitable for modeling the page shape of an opened thick and bound book, is also adopted in some previous works, *e.g.*, [14]. Secondly, we require the rulings of GCS to be perpendicular to the baselines. This orthogonality assumption plays an important role on the estimation of model parameters, as will be shown later. Thirdly, we assume the page surface is smooth. This assumption will help the estimation of spatial directrix of GCS against significant noises and outliers.

## 2.2 Estimation of Model Parameters

From the above assumptions, we can derive an important relationship between the page shape and the parameters of camera. For a general cylindrical surface, its rulings are a group of parallel straight lines. Therefore, after perspective projection of camera, these rulings will intersect at a common vanishing point, denoted by  $(v_0, v_1)$ . Similarly, according to the property of cylindrical surfaces, the tangent vectors of the curved baselines along the same ruling are parallel to each other. Hence, these tangent vectors after perspective projection will converge at a common vanishing point, denoted by  $(x, y)$ . Using the orthogonality assumption between the baselines and rulings, we can derive the equation of vanishing line that each  $(x, y)$  is satisfied, *i.e.*,

$$v_0x + v_1y + f^2 = 0, \quad (1)$$

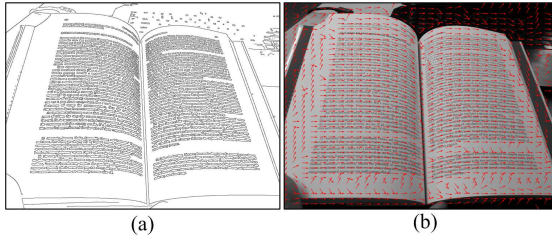
where  $f$  is the focal length of the camera. Fig.2 illustrates the geometric relationship of the vanishing points of the projected rulings and the tangents of baselines. The above equation gives the equality constraint that the vanishing points of the tangent vectors of baselines across the same ruling have to be satisfied. Next, we will exploit this equation to estimate the model parameters and recover the 3D page shape from the vector fields of the document image.



**Fig. 2.** Geometric relationship of the vanishing points of the projected rulings and the tangents of baselines.

**Estimation of vector field** The vector field of a document image consists of the unit tangent vectors of the underlying curved baselines. Hence, the estimation of vector field is actually to recover the local orientations of image. This problem has been extensively studied in the context of skew estimation for the scanned document images with a global skew angle [2]. In this study, we compute the local projections of image across various angles by Radon transform to estimate the local orientations. The angle with the maximum variance of local projections is taken as the estimate of local orientation.

To facilitate the efficient computation of the local projections at every pixel, we also introduce an intermediate integral image. This makes the total computational complexity comparable to that of the Radon transform of image. Fig.3 shows an example of vector field estimation by computing the local projections at each foreground edge pixel. Although the estimated vector field is sparse and noisy due to local projections, it can be well exploited to robustly estimate the model parameters by weighted majority voting.



**Fig. 3.** Estimation of the vector field of image by computing the variances of local projections. (a) the image edge map, (b) the estimated vector field of image.

**Vanishing point of projected rulings** The vanishing point  $(v_0, v_1)$  is the common intersection point of the projected rulings. We use the Radon transform to detect the potential projected rulings in the image and then vote them for the vanishing point.

For a curved general cylindrical surface, its rulings are the only linear structures on the surface. After perspective projection, these rulings remain straight in the image. Many related visual cues are available in a document image for the detection of these straight lines, for example, the vertical boundaries of documents, text blocks and inserted photos or vertical lines in tables or some vertical strokes of characters. We adopt the Radon transform on the image edge map to detect these potential lines.

The Radon transform is the line integral of an image along a group of straight lines. Let  $\mathcal{R}(\rho, \theta)$  be the Radon transform of the image edge map, where  $(\rho, \theta)$  defines a straight line, along which the line integral of edge map is computed. We calculate the likelihood of lines being projected rulings as the local variances of Radon transform along the  $\rho$ -axis, *i.e.*,

$$\mathfrak{L}(\rho, \theta) = \int_{\rho-\delta}^{\rho+\delta} (\mathcal{R}(\tilde{\rho}, \theta) - m(\rho, \theta))^2 d\tilde{\rho}, \quad (2)$$

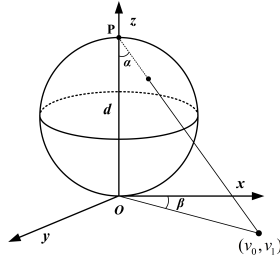
where  $\delta$  is the window size for computing the local variance,  $m(\rho, \theta)$  is the local mean of the Radon transform along the  $\rho$ -axis, defined as:

$$m(\rho, \theta) = \frac{1}{2\delta} \int_{\rho-\delta}^{\rho+\delta} \mathcal{R}(\tilde{\rho}, \theta) d\tilde{\rho}. \quad (3)$$

A weighted majority voting on a sphere is adopted to estimate the vanishing point. To this end, we first introduce the spherical coordinates of  $(v_0, v_1)$  under the stereographic projection. The stereographic projection maps a point on the sphere to a unique point on the plane, as illustrated in Fig.4. Denote the corresponding spherical coordinates of  $(v_0, v_1)$  as  $(\alpha, \beta)$ , where  $\alpha$  and  $\beta$  are two angles, satisfying:

$$\begin{cases} v_0 = d \tan \alpha \cdot \cos \beta, \\ v_1 = d \tan \alpha \cdot \sin \beta, \\ \alpha \in [0, \frac{\pi}{2}), \beta \in [0, 2\pi) \end{cases} \quad (4)$$

where  $d$  is the diameter of the sphere. The benefit of using the spherical coordinates is that they are bounded. Therefore, we can discretize them in a given range and vote them according to the likelihood of projected rulings.



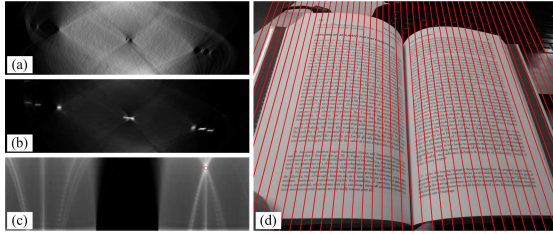
**Fig. 4.** The stereographic projection maps a point on the sphere to a unique point on the plane.

The voting process is as follows: for every candidate projected ruling defined by  $(\rho, \theta)$ , we vote for all pairs of  $(\alpha, \beta)$  satisfying the equation of the projected ruling, *i.e.*,

$$\tan \alpha \cdot \cos(\beta + \theta) = \frac{\rho}{d}. \quad (5)$$

The voting weight is given by the likelihood of  $(\rho, \theta)$  defined in Eq.(2). Actually, Eq.(5) defines a transform from  $\mathcal{L}(\rho, \theta)$  to the voting space  $\mathcal{V}(\alpha, \beta)$ . The point with the maximum votes in  $\mathcal{V}(\alpha, \beta)$  is finally taken as the estimate of the vanishing point. Fig.5 illustrates an example for the estimation of the vanishing point of projected rulings. Once the vanishing point is estimated, we actually recover all the projected rulings in the image.

**Estimating the focal length of camera** Given the estimation of  $(v_0, v_1)$ , we can vote for the focal length  $f$  of the camera according to Eq.(1). To this end, we first sample a sequence of projected rulings. For each projected ruling, we randomly select several pairs of foreground points, say,  $p$  and  $q$ . Then the vanishing point of tangents can be easily computed as the intersection point of



**Fig. 5.** Estimation of the vanishing point of projected rulings. (a) the Radon transform  $\mathcal{R}(\rho, \theta)$  on the image edge map in Fig.3(a), (b) the likelihood of projected rulings  $\mathfrak{L}(\rho, \theta)$ , (c) the voting space  $\mathcal{V}(\alpha, \beta)$ , (d) the estimated projected rulings.

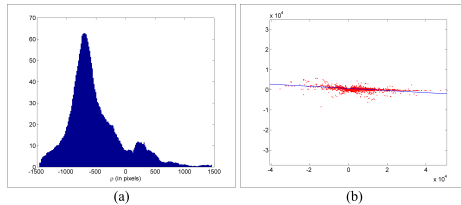
the tangent lines at  $p$  and  $q$ , *i.e.*,

$$\begin{cases} x = \frac{\rho_p \cos \theta_q - \rho_q \cos \theta_p}{\sin(\theta_q - \theta_p)} \\ y = \frac{\rho_p \sin \theta_q - \rho_q \sin \theta_p}{\sin(\theta_q - \theta_p)} \end{cases}, \quad (6)$$

where  $(x, y)$  is the vanishing point of tangents,  $\theta_p$  and  $\theta_q$  are the included angles of the tangent vectors at  $p$  and  $q$  with the  $x$ -axis, respectively.  $\rho_p$  and  $\rho_q$  are computed as below:

$$\begin{aligned} \rho_p &= y_p \cos \theta_p - x_p \sin \theta_p \\ \rho_q &= y_q \cos \theta_q - x_q \sin \theta_q \end{aligned}, \quad (7)$$

where  $(x_p, y_p)$  and  $(x_q, y_q)$  are the coordinates of  $p$  and  $q$  in the image, respectively.



**Fig. 6.** The estimation of focal length of camera by majority voting of points from the projected rulings. (a) the voting histogram of  $\rho$ , (b) the vanishing points of tangents and the estimated vanishing line (the blue line).

We introduce a signed distance  $\rho$  between the origin and the vanishing line in Eq.(1), defined as:

$$\rho = \frac{xv_0 + yv_1}{\sqrt{v_0^2 + v_1^2}}. \quad (8)$$

Instead of directly voting for  $f$ , we vote for  $\rho$  according to the estimated  $(x, y)$  in Eq.(6). One benefit of voting  $\rho$  is that we can avoid the problem caused by

sign flipping of  $(x, y)$  and  $(v_0, v_1)$ . This problem commonly happens when  $(x, y)$  or  $(v_0, v_1)$  is approaching to the infinity. In this case, a small error may cause the reversal of signs in  $(x, y)$  and  $(v_0, v_1)$ .

The focal length  $f$  is finally calculated as:

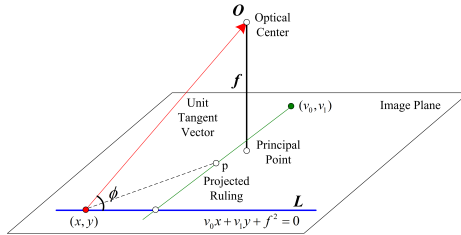
$$f = \sqrt{|\rho^*| \sqrt{v_0^2 + v_1^2}}, \quad (9)$$

where  $\rho^*$  is the  $\rho$  with the maximum votes. Fig.6 illustrates an example of focal length estimation by majority voting of points from the projected rulings. In the figure, we also show the computed vanishing points of tangents and the estimated vanishing line.

### 2.3 3D Reconstruction of Page Surface

**Estimation of unit tangent vectors** A general cylindrical surface is generated by moving a straight line (*i.e.*, the ruling) along a curve called *directrix*. We can estimate the directrix by recovering the unit tangent vectors of spatial baselines.

According to our assumptions, the unit tangent vectors of spatial baselines across the same ruling are uniquely determined by their vanishing point in the image. Moreover, these vanishing points satisfy the vanishing line equation defined in Eq.(1). Therefore, we can parameterize each unit tangent vector with a single parameter  $\phi$ , *i.e.*, the included angle between the unit tangent vector and the vanishing line  $L$ , as illustrated in Fig.7.



**Fig. 7.** The unit tangent vectors of directrix. Given the vanishing line  $L$ , the unit tangent vector is uniquely determined by the included angle  $\phi$  between  $L$  and the unit tangent vector.

The parameter  $\phi$  corresponding to each projected ruling is estimated by majority voting of points on the ruling. For every point  $p$  on a projected ruling, we first compute the intersection point  $(x, y)$  of the tangent line at  $p$  with the vanishing line  $L$  (see Fig.7). Then  $\phi$  is calculated and voted.

The estimated  $\phi$  may be noisy or even erroneous for some projected rulings in page margins or photographic regions due to the sparse and noisy vector field. We further use the smoothness assumption on the page surface to refine



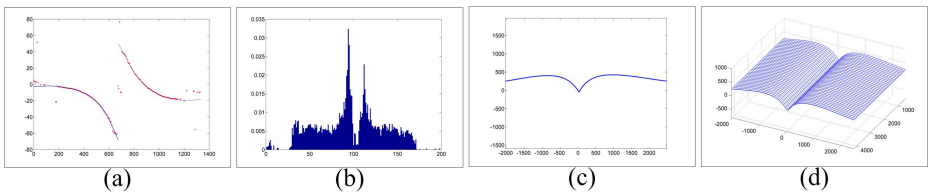
the estimation. To this end, we sample a sequence of projected rulings,  $\ell_k (1 \leq k \leq n)$ , where  $n$  is the total number of sampled rulings. For each projected ruling, its corresponding  $\phi_k$  is estimated by majority voting. Further denote the corresponding maximum votes as  $w_k$ . The smoothness of page surface also means that  $\phi$  is a smooth curve. Therefore, we can fit a smooth curve to the estimated  $\phi_k$  by solving the following 1D optimization problem:

$$\min_{\phi(t)} \sum_{k=1}^n w_k g_{\tau}(|\phi(t_k) - \phi_k|) + \lambda \left| \frac{d^2 \phi(t_k)}{dt^2} \right|^2, \quad (10)$$

where  $g_{\tau}(\cdot)$  is a robust influence function, defined as

$$g_{\tau}(x) = \begin{cases} x^2, & \text{if } x^2 \leq \tau \\ \tau, & \text{else} \end{cases}. \quad (11)$$

The above optimization problem can be efficiently solved by the half-quadric splitting technique [23]. Fig.8(a) illustrates an example of the estimation of the unit tangent vectors. In the example, due to the non-smoothness of the book surface along the spine line, we manually split the estimated  $\phi$  at the spine point into two pieces and fit them separately.

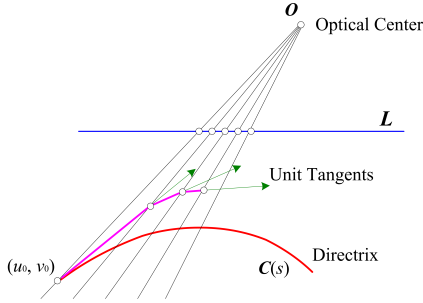


**Fig. 8.** 3D reconstruction of page surface. (a) the fitted  $\phi$ , (b) the weight  $w_k$ , (c) the recovered spatial directrix, (d) the reconstructed 3D page surface.

**Recovery of directrix and 3D page reconstruction** Once the unit tangent vectors of spatial baselines are estimated, we can recover a 3D directrix on the plane determined by the optical center  $O$  of camera and the vanishing line  $L$ . The 3D directrix can be estimated by solving the following ordinary differential equation:

$$\begin{cases} \frac{d\mathbf{C}(s)}{ds} = \mathbf{t}(s) \left| \frac{d\mathbf{C}(s)}{ds} \right| \\ \mathbf{C}(s_0) = (u_0, v_0) \end{cases}, \quad (12)$$

where  $\mathbf{C}(s)$  is the 3D directrix parameterized by  $s$ ,  $\mathbf{t}(\cdot)$  is the estimated unit tangent vector, which is directly determined by the corresponding angle  $\phi$ ,  $(u_0, v_0)$  is the given boundary condition at  $s_0$ . The above ordinary differential equation



**Fig. 9.** Estimation of the 3D directrix by solving the ordinary differential equation with the Euler method. The Euler method iteratively constructs polygonal lines to approximate the solution curve.

can be efficiently solved by using the Euler method. Fig.9 illustrates the process of solving for the directrix by the Euler method.

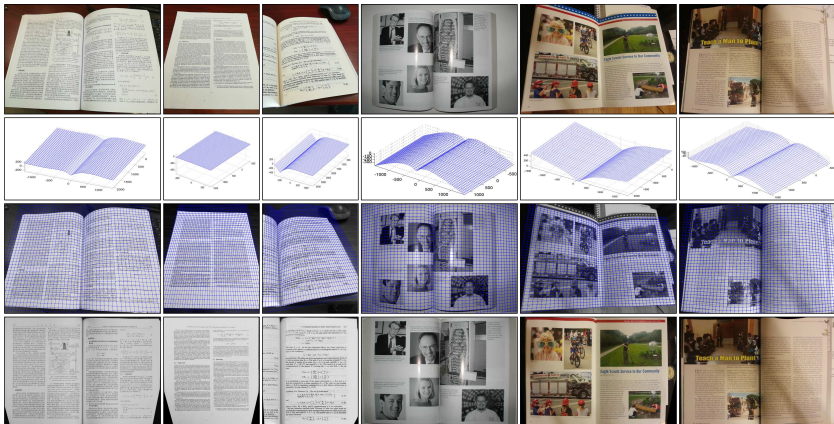
After the directrix is estimated, the 3D page surface can be reconstructed by moving the directrix along the ruling direction. In Fig.8(c) and (d), we illustrate the recovered spatial directrix and the reconstructed 3D page surface of the image in Fig.3(a).

Once the 3D page surface is estimated, we can rectify the geometric distortion in the image through mesh warping. This can be done by fattening the estimated 3D page surface onto a plane to construct a mesh grid. In this process, the correspondence between the points on the plane and the image can be made. Then the dewarping functions for mesh warping can be derived through cubic spline fitting.

## 3 Experiments

### 3.1 Results on real-captured images

To evaluate the performance of our method, we implemented a sequence of experiments on real-captured document images by hand-held cameras and smartphone cameras. Fig.10 illustrates several representative images captured from opened books and documents by using a smartphone camera. The document images have complex page layouts with multiple columns and various types of contents other than text-lines, such as inserted large areas of figures, formulas and cluttered backgrounds. Besides, the documents consist of different languages and are taken from different viewpoints, thus having different levels of perspective distortion and defocusing blur. These images are generally challenging for some text-lines based methods, *e.g.*, [14], since it is quite difficult to accurately extract the curved baselines in distorted document images with complicated layouts. Our method does not require the segmentation of curved text-lines and thereby is robust to complex page layouts. As can be seen from the results, our method work quite well on these images.



**Fig. 10.** Geometric rectification results of real-captured document images. From top to bottom: the distorted document images, the reconstructed 3D page surface, the constructed mesh grids and the rectification results.

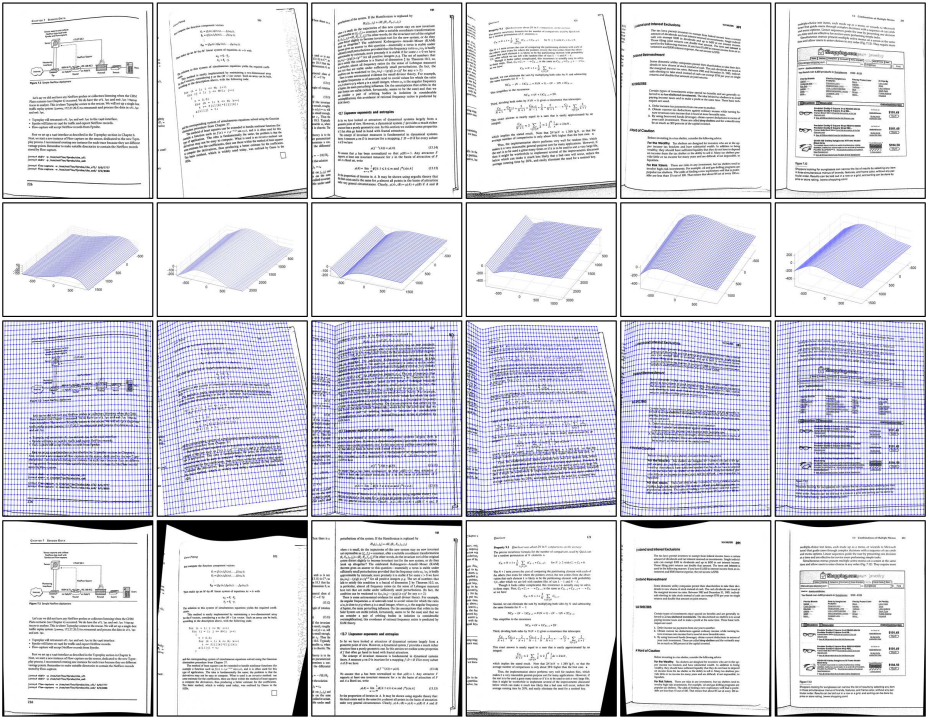
We also tested our method on the DFKI dataset <sup>1</sup> [24]. The dataset is specially designed for the evaluation of methods for curved document images rectification. Fig.11 illustrates several typical examples of the rectification results of our method on the dataset. The dataset consists of binary English document images captured from opened book pages with various types of document contents. All the document pages in the dataset are approximately distorted in general cylindrical shapes and thus are very suitable for the evaluation of our method. As we can see from the results, the proposed method is able to remove all types of distortions in the images quite desirably, including geometric distortions along horizontal text-lines and perspective distortion of camera in vertical direction.

### 3.2 Comparisons with Kim *et al.* 's method

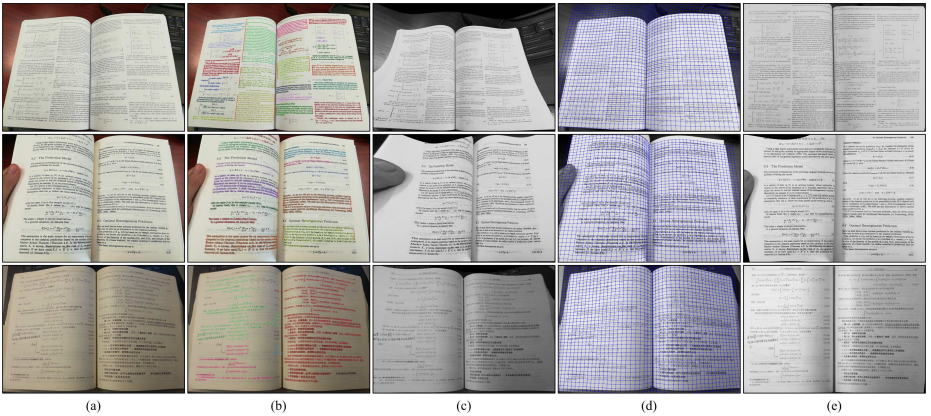
Kim *et al.*'s method <sup>2</sup> [25] adopts a similar cylindrical page shape assumption for geometric rectification. Their method relies on a connected component analysis technique to group characters into text-lines and text blocks. Then the curved baselines is extracted to estimate the parameters of their 3D page model. Fig.12 illustrates the comparisons of rectification results of our method with their method. Kim *et al.*'s method works not well on the given examples mainly due to two reasons. First, Kim *et al.*'s method requires the focal length of camera to be provided manually. An erroneous focal length of camera will result in a large rectification error. Second, accurate segmentation of text-lines and text blocks in

<sup>1</sup> The dataset can be downloaded from <http://staffhome.ecm.uwa.edu.au/00082689/downloads.html>

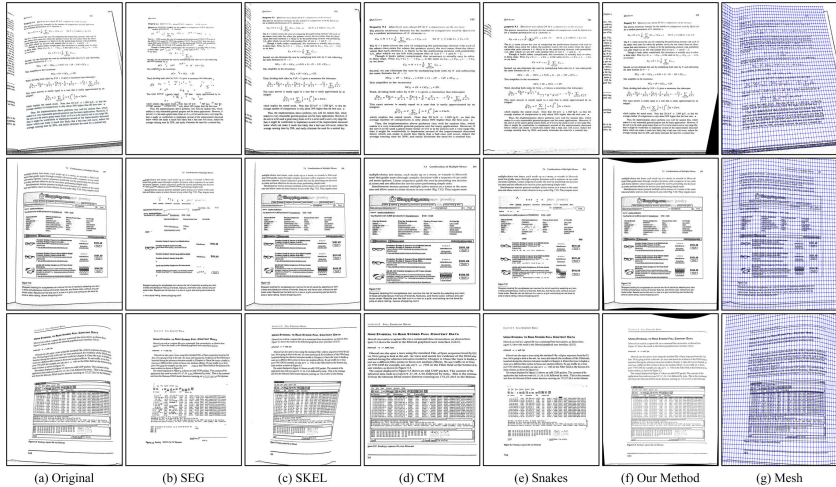
<sup>2</sup> The executable codes can be downloaded from <http://ispl.snu.ac.kr/bskim/DocumentDewarping/>



**Fig. 11.** Geometric rectification results of our method on DFKI dataset. From top to bottom: the distorted document images, the reconstructed 3D page surface, the constructed mesh grids and the rectification results.



**Fig. 12.** Comparisons of our method with Kim *et al.*'s method [25]. (a) the distorted document images, (b) the extracted text-lines by Kim *et al.*'s method, (c) the rectification results of Kim *et al.*'s method, (d) the constructed mesh grid of our method, (e) the rectification results of our method.



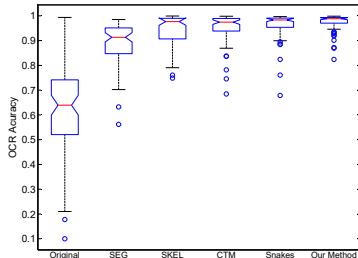
**Fig. 13.** Comparisons of our method with the SEG method [26], the SKEL method [27], the CTM method [28] and the Snakes based method [29] on the DFKI dataset. (a) the distorted document images, (b) results of the SEG method, (c) results of the SKEL method, (d) results of the CTM method, (e) results of the Snakes based method, (f) results of our method, (g) the constructed mesh grids of our method.

a distorted document image is challenging. The proposed connected component analysis technique is sensitive to languages and often fails to document images with sparse text-lines, as can be seen from the results in the figure.

### 3.3 Comparisons on DFKI dataset

We also compared our method with several state-of-the-art methods on DFKI dataset, including the SEG method [26], the SKEL method [27], the CTM method [28] and the Snakes based method [29]. The SEG method [26] segments each word in the image and rotates them onto straight lines to correct local skews. The SKEL method [27] extracts the outer skeletons of a text image and then fits a Bezier surface to the whole page to estimate a deformation mapping. The CTM method [28] extracts the curved horizontal text-lines by a morphological method and adopts a cylindrical shape model to rectify the image. The Snakes based method [29] employs a coupled snake model to extract curved text baselines for image rectification. Fig.13 illustrates the comparisons of our method with the four methods. We also give the box plot of the OCR accuracy of the five methods in Fig.14.

From the results, we can see that the SEG method and the Snakes based method cannot rectify the distortions of non-textual objects in the images, *e.g.*, tables and formulas. The SKEL method fails to estimate the deformation function robustly when an image contains large area of non-textual objects. In comparison, the CTM method and our method work quite well. We point out that



**Fig. 14.** The comparisons of the OCR accuracy of the five methods on the DFKI dataset.

the performance of the CTM method heavily depends on the accuracy of the extracted text lines. The morphological method used in the method is very sensitive to the structure element sizes, which are vulnerable to many challenging factors, including the variations of image resolution and font sizes, image distortion, image blur and document layouts. In comparison, our method does not rely on text-lines detection and segmentation. It can thus be used to document images with complex page layouts and severe quality degradation.

## 4 Conclusion

In this paper, we have proposed a segment-free method for geometric rectification of curved document images captured by a hand-held camera. By assuming that the curved page shape is a general cylindrical surface, we can reconstruct the underlying 3D page shape from a single document image by exploiting its intrinsic vector fields. In comparison with the widely used text-lines based methods, *e.g.*, [14, 25], our method does not require the detection and segmentation of curved horizontal text lines or textual blocks, which is still an open problem especially for severely distorted document images with complex page layouts. The proposed method can exploit various types of available visual cues in images for 3D page shape recovery. It is thus can be applied to document images with complicated page layouts and serious image quality degradation. We also implemented extensive experiments on real-captured document images to test the performance of our method.

## Acknowledgment

We thank the kind area chair and the anonymous reviewers for their valuable comments. This work was supported in part by the National Natural Science Foundation of China under Grants 91646207, National Science Foundation grant IIS-1217302, IIS-1619078, and the Army Research Office ARO W911NF-16-1-0138.

## References

1. Liang, J., Doermann, D., Li, H.: Camera-based analysis of text and documents: a survey. *International Journal of Document Analysis and Recognition* **7**(2-3) (2005) 84–104
2. Nagy, G.: Twenty years of document image analysis in pami. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(1) (2000) 38–62
3. Brown, M.S., Tsoi, Y.C.: Geometric and shading correction for images of printed materials using boundary. *IEEE Transactions on Image Processing* **15**(6) (2006) 1544–1554
4. Stamatopoulos, N., Gatos, B., Pratikakis, I., Perantonis, S.J.: Goal-oriented rectification of camera-based document images. *IEEE Transactions on Image Processing* **20**(4) (2011) 910–920
5. Tsoi, Y.C., Brown, M.S.: Geometric and shading correction for images of printed materials: a unified approach using boundary. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Volume 1. (2004) 240–246
6. Ulges, A., Lampert, C.H., Breuel, T.M.: Document image dewarping using robust estimation of curved text lines. In: *Proceedings of the 8th International Conference on Document Analysis and Recognition*. (2005) 1001–1005
7. Zhang, Z., Tan, C.L.: Straightening warped text lines using polynomial regression. In: *ICIP'02*. Volume 3. (2002) 977–980
8. Zhang, Z., Tan, C.L.: Correcting document image warping based on regression of curved text lines. In: *Proceedings of the 7th International Conference on Document Analysis and Recognition (ICDAR)*. (2003) 589–593
9. Brown, M.S., Sun, M., Yang, R., Yun, L., Seales, W.B.: Restoring 2d content from distorted documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(11) (2007) 1904–1916
10. Cao, H., Ding, X., Liu, C.: A cylindrical surface model to rectify the bound document image. In: *Proceedings of International Conference on Computer Vision (ICCV)*. (2003) 228–233
11. Hyung, I.K., Kim, J., Nam, I.C.: Composition of a dewarped and enhanced document image from two view images. *IEEE Transactions on Image Processing* **18**(7) (2009) 1551–1562
12. Liang, J., DeMenthon, D., Doermann, D.: Flattening curved documents in images. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Volume 2. (2005) 338–345
13. Liang, J., DeMenthon, D., Doermann, D.: Geometric rectification of camera-captured document images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(4) (2008) 591–605
14. Meng, G., Pan, C., Xiang, S., Duan, J., Zheng, N.: Metric rectification of curved document images. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **34**(4) (2012) 707–722
15. Meng, G., Xiang, S., Pan, C., Zheng, N.: Active rectification of curved document images using structured beams. *International Journal of Computer Vision* **122**(1) (2017) 34–60
16. Tan, C.L., Zhang, L., Zhang, Z., Xia, T.: Restoring warped document images through 3d shape modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(2) (2006) 195–208
17. Tian, Y., Narasimhan, S.: Rectification and 3d reconstruction of curved document images. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (June 2011) 377–384

18. You, S., Matsushita, Y., Sinha, S., Bou, Y., Ikeuchi, K.: Multiview rectification of folded documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017)
19. Zhang, L., Yip, A.M., Brown, M.S., Tan, C.L.: A unified framework for document restoration using inpainting and shape-from-shading. *Pattern Recognition* **42**(11) (2009) 2961–2978
20. Meng, G., Wang, Y., Qu, S., Xiang, S., Pan, C.: Active flattening of curved document images via two structured beams. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2014) 3890–3897
21. Meng, G., Huang, Z., Song, Y., Xiang, S., Pan, C.: Extraction of virtual baselines from distorted document images using curvilinear projection. In: *IEEE International Conference on Computer Vision*. (2015) 3925–3933
22. Schneider, D., Block, M., Rojas, R.: Robust document warping with interpolated vector fields. In: *Proceedings of the Ninth International Conference on Document Analysis and Recognition*. Volume 1. (2007) 113–117
23. Nikolova, M., Ng, M.K.: Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM Journal on Scientific computing* **27**(3) (2005) 937–966
24. Shafait, F., Breuel, T.M.: Document image dewarping contest. In: *Proceedings of the 2nd Int. Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil (Sep. 2007)* 181–188
25. Kim, B.S., Koo, H.I., Cho, N.I.: Document dewarping via text-line based optimization. *Pattern Recognition* **48**(11) (2015) 3600 – 3614
26. Gatos, B., Pratikakis, I., Ntirogiannis, K.: Segmentation based recovery of arbitrarily warped document images. In: *The 9th International Conference on Document Analysis and Recognition, Curitiba, Brazil (Sep. 2007)* 989–993
27. Masalovitch, A., Mestetskiy, L.: Usage of continuous skeletal image representation for document images dewarping. In: *The 2nd Int. Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil (Sep. 2007)* 45–52
28. Fu, B., Wu, M., Li, R., Li, W., Xu, Z.: A model-based book dewarping method using text line detection. In: *The 2nd Int. Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil (Sep. 2007)* 63–70
29. Bukhari, S.S., Shafait, F., Breuel, T.M.: Dewarping of document images using coupled-snakes. In: *The 3rd Int. Workshop on Camera-Based Document Analysis and Recognition, Barcelona, Spain (July 2009)* 34–41