Unwarping Images of Curved Documents Using Global Shape Optimization

Jian Liang, Daniel DeMenthon, David Doermann
Language And Media Processing Laboratory
University of Maryland, College Park, MD, 20770
{lj,daniel,doermann}@cfar.umd.edu

Abstract

The unwarping of curved document images is a crucial problem for camera-based document analysis since most of current OCR techniques can not handle distortion due to perspective and warping. In previous work we have shown how to recover the page shape from a single image using an iterative procedure without camera calibration, and using the shape information to restore a frontal view of a flat document. In this paper we report our recent progress using a global optimization method to do shape estimation. Experimental results show a clear improvement over our previous method.

1 Introduction

Digital cameras have become more and more popular not only among consumers but also business and technical professionals. For the OCR community, they provide a potential alternative to scanners as document imaging devices. Current OCR techniques are, however, designed with digital scans of flat documents in mind, and cannot handle general camera-captured documents due to both perspective and warping.

One way of removing the added 3D distortion is to use special 3D scanning equipments such as structured light. A mesh can be built to represent the 3D surface and directly flattened [1] or transformed to a developable mesh [8]. Alternatively, the shape can be estimated from the image. The problem of removing only the perspective from images of planar documents is addressed in [3, 8, 4]. For warped documents, there are parametric approaches [2, 5, 12] that estimate the 3D shape and non-parametric ones [10, 11] that bypass shape estimation. Among them, [11, 12] are designed only for scans of bound books; [2, 10] require a straight frontal view of page with cylinder shape; [5] is proposed for general images, but needs camera calibration and a prior knowledge of a closed contour (e.g., page boundaries) on the page which may be difficult in practice. Overall, current methods have various restrictions that keep them from being applied to general images.

Our goal is to handle general warped documents with fewer restrictions. Our method falls in the parametric category. It is based on two key observations: 1) curved document pages form developable surfaces which can be approximated by planar strips, and 2) the projected image of printed textual content on the page constrains the underlying surface shape by the parallelism, geodesic, and equidistant properties of text lines (see [6] for a discussion on geodesic texture flow and developable surface under perspective projection). Compared to other's work, our method does not require special equipment or camera calibration, can be applied to general warped documents, and can work on partially occluded documents.

In [7] we have discussed the details of image processing, shown that page shape can be estimated, and obtained much higher OCR rates from unwarped images. However, shape information is not explicitly expressed in [7], which makes it difficult for evaluation, nor is the estimation process globally optimal with regard to developable property and text property. In this paper we present our recent progress using global shape optimization which gives significant improvement over [7].

Section 2 and Section 3 briefly recalls the work in [7]. In Section 4 we describe the initialization and optimization of shape estimation. Section 5 discusses experiment results and finally Section 6 concludes the paper.

2 Problem Modeling

The shape of a smoothly rolled document page can be modeled by a developable surface. In [7] we show that a developable surface can be approximated by planar strips that come from the family of its tangent planes (see Fig 1), which can be fully described by a set of reference points $\{P_i\}$, and surface normals $\{\mathbf{N}_i\}$.

For documents that are covered by printed text we can define two texture flows on the surface, both in 3D space and in 2D projected images. One corresponds to the text line direction, which we call *3D(2D) major texture flow* and denote by $\mathbf{T}$(or $\mathbf{t}$); the other corresponds to the vertical char-

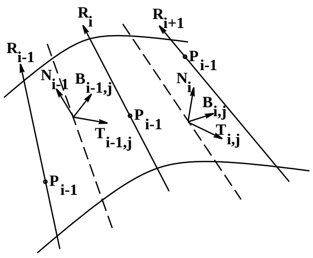
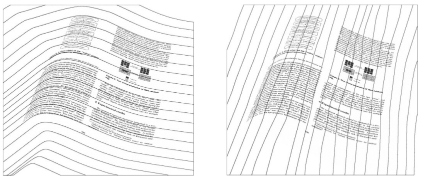


Figure 1. A developable surface can be approximated by planar strips (for variable definitions see Section 4)



(a) (b)

Figure 2. Texture flow detection: (a) major texture flow (b) minor texture flow

acter stroke direction, which we call *3D(2D) minor texture flow*, denoted by **B**(or **b**). For a 3D ruling **R** we also define a *2D projected ruling* **r** in the image.

3 Image Processing

Based on the developable surface model, our approach is to segment the surface by a group of rulings, approximate the pieces in between the rulings by planar strips, and unroll the page strip by strip. In this section we will very briefly go over the image processing step that support the shape estimation in next section (for more details see [7]).

The first step is to find the printed text area because we will be using the properties of printed text and any non-text element may cause unexpected results. An adaptive thresholding [9] inside the text area gives us the binary text image, and all following computation is based on the binary text image. Secondly we extract the two 2D texture flows. We divide the image into small blocks, and use projection profile analysis to compute the local texture orientations. A relaxation process resolves any conflict among neighboring blocks, and results in two texture flow fields (Fig. 2).

Projected ruling detection is based on the property that texture flow patterns along a projected ruling is more consistent than that along an arbitrary line. For each ruling, its vanishing point is estimated by the fact that printed text lines are usually equally spaced and the invariance of cross ratio under perspective projection.

4 Page Shape Estimation

In [7] we described how to iteratively optimize the shape of a document under developable property and text property constraints. However, there are two problems. First, we do not get an explicit surface normal for each strip; instead we compute *horizontal* and *vertical* vanishing points. Second, we do not have an explicit objective function and therefore the iterative process does not have an explicit measurement of the progress. In this paper, we address these two problems by formally introducing several constraints defined on surface normals as well as focal length of the camera, and an objective function based on the constraints. By optimizing the objective function we obtain explicit surface normals and focal length.

4.1 Constraints

It is difficult to estimate the normal of each planar strip only using local features. Fortunately there are strong global constraints imposed by the developable property and text properties. First we will define the variables (see Fig. 1), and then introduce the constraints.

• Wanted unknowns:
– 3D normals: $\{\mathbf{N}_i\}_{i=1}^{L}$, where L is the number of strips
– 3D reference points: $\{P_i\}_{i=1}^{L+1}$
– Focal length: f_0.
• Preprocessing results and known variables:
– Projected rulings: $\{\mathbf{r}_i\}_{i=1}^{L+1}$
– Projected reference points: $\{p_i\}_{i=1}^{L+1}$
– Projected texture flow: **t** and **b** all over the image
• Other related variables:
– 3D rulings: $\{\mathbf{R}_i\}_{i=1}^{L+1}$
– 3D texture flow: For the i-th strip, we select a group of J_i sample points inside the strip, and define $\mathbf{T}_{ij}$ as the 3D major texture flow vector at the j-th point, and $\mathbf{B}_{ij}$ as the minor texture flow vector.
– 3D viewing direction vector: For the j-th sample point in the i-th strip, we define $\mathbf{V}_{ij}$ as its viewing direction vector with respect to the camera's optical center.

All the vectors are of unit length.

Suppose $\eta(\cdot)$ represents the normalization operator where $\eta(\mathbf{v}) = \mathbf{v}/|\mathbf{v}|$, then the 3D vectors are related to their projections in the image by the following equations:

$$\begin{aligned}
\mathbf{R}_i &= \eta((\mathbf{r}_i \times \mathbf{V}_i) \times (\mathbf{N}_i + \mathbf{N}_{i-1})/2) \\
\mathbf{T}_{ij} &= \eta((\mathbf{t}_{ij} \times \mathbf{V}_{ij}) \times \mathbf{N}_i) \\
\mathbf{B}_{ij} &= \eta((\mathbf{b}_{ij} \times \mathbf{V}_{ij}) \times \mathbf{N}_i)
\end{aligned}$$

Note that in the equation relating **R** and **r** we use $\mathbf{N}_i + \mathbf{N}_{i+1}$ to approximate the surface normal along $\mathbf{R}_i$.

Without loss of generality we can assume that P_i are on the rulings. By the continuity property of the planar strips it is easy to see that once we have obtained surface normals, focal length, and the depth of any particular P_{i_0}, the rest P_i are fully determined.

There are four constraints that we can derive from the developable property of the page and the property of text documents:

- Orthogonality between surface normals and rulings: Ideally, we would want $\mathbf{N}_{i-1} \cdot \mathbf{R}_i = \mathbf{N}_i \cdot \mathbf{R}_i = 0$. Since we have fixed $\mathbf{R}_i$ to be orthogonal to $\mathbf{N}_{i-1} + \mathbf{N}_i$, we only need to check $\mathbf{R}_i \cdot (\mathbf{N}_i - \mathbf{N}_{i-1})$. We define $\mu_1 = \sum_{i=1}^{L-1} (\Delta \mathbf{N}_i \cdot \mathbf{R}_i)^2$ where $\Delta \mathbf{N}_i = \mathbf{N}_i - \mathbf{N}_{i-1}$, and ideally $\mu_1 = 0$.
- Parallelism of text lines inside each strip: Text line directions are represented by $\mathbf{T}_{ij}$. We use $\mu_2 = \sum_i \sum_j |\mathbf{T}_{ij} - \overline{\mathbf{T}}_i|$, where $\overline{\mathbf{T}}_i$ is the average of all $\mathbf{T}_{ij}$ within the i-th strip, to measure their parallelism. Ideally $\mu_2 = 0$.
- Geodesic property of text lines crossing two neighboring strips: The text lines on two neighboring strips form two different angles with the 3D ruling that separates the strips. After unwarping, the angles do not change. If the sum of the two angles is π, it means the text line is straight in the unwarped image. We use $\mu_3 = \sum_i ((\overline{\mathbf{T}}_{i+1} - \overline{\mathbf{T}}_i) \cdot \mathbf{R}_i)^2$ to measure the straightness, which ideally is zero.
- Orthogonality between text line direction and vertical stroke direction: The orthogonality can be measured by $\mu_4 = \sum_i \sum_j |T_{ij}^{T} B_{ij}|$, which in the idea case should be zero.

In our experiment we embedded two additional constraints:

- Smoothness: We use $\mu_5 = \sum_i |\Delta \mathbf{N}_i|$ to measure the smoothness of the surface. A large value indicates abrupt changes in normals of neighboring strips and therefore should be avoided.
- Unit length: Each normal should be of unit length. We measure this by $\mu_6 = \sum_i (1 - |\mathbf{N}_i|)^2$.

The overall optimization objective function is the weighted sum of all constraint measurements,

$$F(\mathbf{X}) = \sum_{i=1}^{6} \alpha_i \mu_i$$

where $\mathbf{X}$ represents all normals and the focal length, and α_i are weights.

Overall, given $\{\mathbf{r}_i\}$, $\{\mathbf{t}_{ij}\}$ and $\{\mathbf{b}_{ij}\}$, the objective function is fully determined by the unknown $\{\mathbf{N}_i\}$ and f_0. The optimal set of $\{\mathbf{N}_i^*\}$ and f_0^* should minimize F.

4.2 Shape Initialization and Optimization

A good initial value of $\mathbf{X}$ is essential for optimizing the highly non-linear objective function. Such initial values can be obtained using the estimated vanishing points of rulings. These vanishing points, when focal length is given, determine the direction of 3D rulings. Since surface normals are orthogonal to 3D rulings, this eliminates one degree of freedom from the unknown normals. The remaining degree of freedom allow a normal to rotate in the plane orthogonal the the ruling. So the objective function is decided by a set of rotation angles. Furthermore, the computation of the objective function involves either each individual normal (μ_2, μ_4, μ_6), or two neighboring normals (μ_1, μ_3, μ_5). Therefore, we can use dynamic programming (DP) search to find the best set of rotation angles that gives the minimum objection function output.

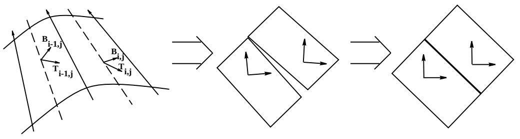


Figure 3. After the surface is developed, post-processing ensures that strips fit each other seamlessly and major texture flow direction is horizontal

The focal length is not covered by the DP search, however. It is independent of the surface normals, and we have to perform an exhaustive search for the initial focal length. More specifically, we select a set of possible focal lengths that are constrained by the physical lens specification, and for each value we find the “best” surface normals, and compute the objective function. We fit a 3rd order polynomial curve to the objective function values vs. the focal lengths, and use the curve to find the “best” focal length. Then we compute the “best” normals for this focal length, and take them as the initial values for non-linear optimization process.

Our non-linear optimization module is based on the optimization toolbox in MATLAB, which is fairly fast and produces good result as long as the initial point is reasonably close to the true solution.

After we have estimated the surface normals and focal length, we can arbitrarily select the depth of any one reference point, which determines the depth of the other reference points, and thus fully determine the 3D position of planar strips. The planar strips can then be mapped to a flat plane, placed side by side to form the flat document. Due to the errors in shape estimation and the fact that the document page in real world may be not perfectly developable, some postprocessing is required to make sure that the strips fit each other and that restored text lines are horizontal and continuous across the whole unwarped image (see Fig. 3).

5 Experiment Results

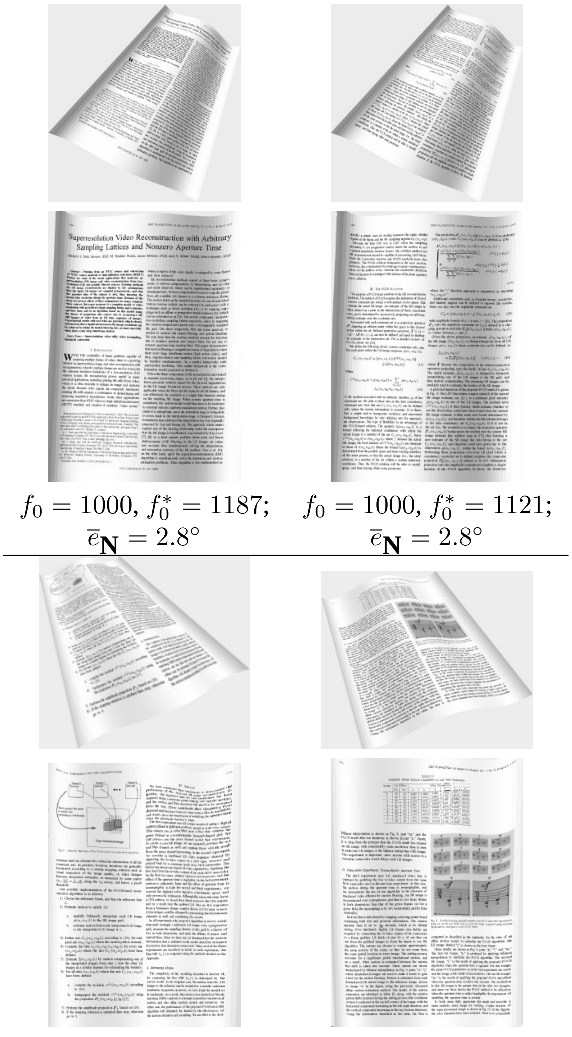
We have applied our method to both synthetic and real images. The synthetic images are generated by warping a flat document image around a predefined developable surface and projecting it onto the image plane of a pinhole camera. With synthetic data, we can evaluate the estimated results such as texture flows, projected rulings, ruling vanishing points, surface normals and focal length against the ground truth.

Fig. 4 shows four synthetic images of warped documents and the unwarped images. It also compares the ground truth focal lengths to the estimates, and shows the average error of surface normals. In Fig. 5 two real images of warped documents and their unwarped images are shown. As we can see in both Fig. 4 and Fig. 5, the text lines are mostly straight and horizontal. Some text line still have some curve due to the errors in major texture flow detection, which is more evident around corners or margins.

The errors in estimated surface normals are measured by the angles between them and corresponding true surface normals. We do not, however, measure the focal length estimation by the difference between it and the true value, because when focal length is large, the reconstructed shape is less sensitive to the change in focal length, which means we can tolerate a larger error. To factor that into the evaluation, we use *view angle* defined as following: a view angle for a given focal length f_0 is $2\mathrm{atan}(d/f_0)$ where d is the largest distance from any point in the image to the optical axis (or, roughly half of the image's dimension). The change in view angle w.r.t f_0 vanishes as f_0 increases, and thus is a better performance measurement.

Although we do not have explicit surface normal estimation from the method in [7], in order to compare it with the results obtained by global optimization we construct approximated surface normals from the results of previous method. In Table 1 we list the mean and standard deviation of view angle errors and surface normal errors from the estimation by previous method, by initialization and final optimization of current method. The results of previous method is obtained from 32 images in which documents contain only text, and the results of global optimization is obtained from 44 images in which documents contain figures and tables. Because of the non-text elements, these 44 images are inherently more difficult to process. Nevertheless, our current method has a great lead over the previous one. This is due in part to the refinements we made in other parts of our code but the main reason is still the new optimization method, especially the shape initialization by DP. In [7] without an explicit objective function representing all constraints we initialized the shape based on local information, and it is not surprising that the initial shape is not as good as that obtained by DP that takes into account global information. The benefit of an explicit objective function is also manifested by the improvement from the initial shape to the final result.

Currently, the parameters involved in image processing and shape estimation stages are manually set. The weight factors α_i are set by experiment in such a way that μ_i become comparable to each other. In the future we will address the automatic parameter selection problem. Nevertheless, among several different settings for each procedure we have not found significant changes in the results. In our experiments we used very conservative parameter values in order to ensure accuracy for arbitrary images. In practice with some knowledge of the image, it is possible to tighten some parameters for better speed.



$f_0 = 1000$, $f_0^* = 1187$; $\bar{e}_{\mathbf{N}} = 2.8°$

$f_0 = 1000$, $f_0^* = 1121$; $\bar{e}_{\mathbf{N}} = 2.8°$

$f_0 = 1000$, $f_0^* = 993$; $\bar{e}_{\mathbf{N}} = 2.5°$

$f_0 = 1500$, $f_0^* = 1533$; $\bar{e}_{\mathbf{N}} = 1.6°$

Figure 4. Unwarped synthetic document images: f_0 is true focal length and f_o^* is estimation (both in pixel unit); $\bar{e}_{\mathbf{N}}$ is the average normal error measured in degree

Among the images that have unsatisfactory results, the major problem comes from the text area detection step. If a background object or a picture in the document gets identified as text, it can interrupt the texture flow detection, and

	Previous method [7] (32 tests)	With global optimization (44 tests)	
		Initial estimation	Final optimization
Ave. view angle error	12.7	8.3	7.3
Std. view angle error	20.5	8.0	7.6
Ave. surface normal error	14.0	6.5	4.8
Std. surface normal error	13.9	4.4	3.6

Table 1. Shape estimation evaluation (error measured in degrees)

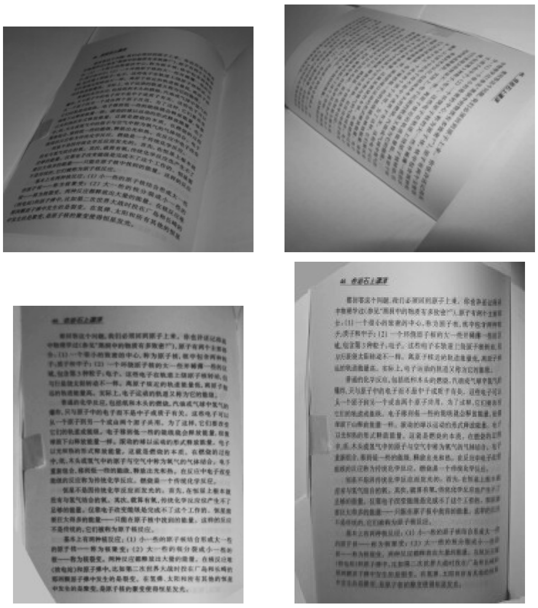


Figure 5. Unwarped real document images

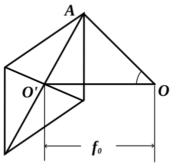


Figure 6. View angle definition: OO' is optical axis; f_0 is focal length; A is the farthest point in the image from OO'.

break the following procedures. Since many researchers have proposed various techniques for identifying text in images, we believe that we can solve this problem by choosing one of them.

6 Conclusion

In this paper we describe how we optimize the page shape estimation globally for unwarping images of curved documents captured by cameras. The document surface is modeled by a developable surface, and we show that the textual content (text lines in particular) provides enough information for recovering the page shape. Compared to the results of our previous method, improvement is obtained by introducing a global optimization into the shape estimation process. From the OCR point of view, the geometry of the reconstructed page is definitely within the acceptable tolerance. However, other challenges still exist, including varying shade, non-uniform blur, fusion of multiple views. We will address them in our future work.

References

[1] M. S. Brown and W. B. Seales. Image restoration of arbitrarily warped documents. *IEEE Trans. PAMI*, 26(10):1295–1306, October 2004.

[2] H. Cao, X. Ding, and C. Liu. Rectifying the bound document image captured by the camera: A model based approach. In *Proc. ICDAR*, pages 71–75, 2003.

[3] P. Clark and M. Mirmehdi. On the recovery of oriented documents from single images. In *Proc. Adv. Concepts for Intelligent Vision Sys.*, pages 190–197, 2002.

[4] C. R. Dance. Perspective estimation for document images. In *Proceedings of SPIE Document Recognition and Retrieval IX*, volume 4670, pages 244–254, 2002.

[5] N. Gumerov, A. Zandifar, R. Duraiswarni, and L. S. Davis. Structure of applicable surfaces from single views. In *Proc. ECCV*, pages 482–496, 2004.

[6] D. C. Knill. Contour into texture: Information content of surface contours and texture flow. *J. Opt. Soc. Am. Ass.*, 18(1):12–35, Jan 2001.

[7] J. Liang, D. DeMenthon, and D. Doermann. Flattening curved documents in images. In *Proc. CVPR*, pages 338–345, 2005.

[8] M. Pilu. Undoing paper curl distortion using applicable surfaces. In *Proc. CVPR*, volume 1, pages 67–72, 2001.

[9] Ø. D. Trier and T. Taxt. Evaluation of binarization methods for document images. *IEEE Trans. PAMI*, 12(3):312–315, 1995.

[10] Y.-C. Tsoi and M. S. Brown. Geometric and shading correction for images of printed materials a unified approach using boundary. In *Proc. CVPR*, pages 240–246, 2004.

[11] Z. Zhang and C. L. Tan. Correcting document image warping based on regression of curved text lines. In *Proc. ICDAR*, volume 1, pages 589–593, 2003.

[12] Z. Zhang, C. L. Tan, and L. Fan. Restoration of curved document images through 3D shape modeling. In *Proc. CVPR*, pages 10–15, 2004.