3D Geometric and Optical Modeling of Warped Document Images from Scanners

1Li Zhang, 1Zheng Zhang, 1Chew Lim Tan, 2Tao Xia
1School of Computing, National University of Singapore
3 Science Drive 2, Singapore 117543
2Department of Mathematics, National University of Singapore
2 Science Drive 2, Singapore 117543
Email: 1{zhangli,zhangz,tancl}@comp.nus.edu.sg, 2iesxiat@nus.edu.sg

Abstract

When one scans a document page from a thick bound volume, the curvature of the page to be scanned results in two kinds of distortion in the scanned document images: i) shade along the ‘spine’ of the book, ii) warping in the shade area. In this paper, we propose an efficient restoration method based on the discovery of the 3D shape of a book surface from the shading information in a scanned document image. We first build practical models namely a 3D geometric model and a 3D optical model for the practical scanning conditions to reconstruct the 3D shape of book surface. We next restore the scanned document image using this shape based on de-shading and de-warping models. Finally, we evaluate the restoration results by comparing the OCR (Optical Character Recognition) performance on the original and restored document images. The experiments show that the geometric and photometric distortions are mostly removed and the OCR results are improved markedly.

1. Introduction

A curved page results in two kinds of distortions in its scanned image as shown in Figure 2: i) shade along book spine (photometric distortion), and ii) warping of the book surface in shade areas (geometric distortion). In this paper, we address the problem of recovering the 3D book surface shape from the shading information in its scanned skew document image.

There have been some related techniques reported in the literature. Weng and Zhu [1] propose a nonlinear shape restoration algorithm for document images based on a linear interpolation theory that is able to detect and restore nonlinear shape distortions in any irregular quadrilateral-shaped patterns. However, this method can only handle binary document images, and many parameters have to be set manually to achieve good restoration results. Pilu [2] presents a novel method based on the physical modeling of paper deformation with an applicable surface. A relaxation algorithm is then used to fit the applicable surface to noisy data and flatten it to produce the final undistorted image. The quality of the restored image largely depends on the density of the polygon mesh used to represent the applicable surface. Brown et al [3] propose a general de-skewing algorithm for arbitrary warped documents based on the 3D shape. They first obtain the depth of each point in the image by some stereo vision method to construct a depth map, and then de-warp the image according to the depth map. However, how to map the points on the rough, noisy surface defined by the depth map to the points on the plane is still a problem. Cao et al [4] introduce a method to rectify the warping of a bound document image captured by camera. They build a general cylindrical model and use the skeleton of horizontal text lines in the image to help estimate the model parameters. This method cannot remove the shading and would not work if there are no text lines or very few text lines. Wada et al [5] develop a model to reconstruct the 3D book surface shape incorporating interreflections. This method requires the book spine being strictly parallel to the scanning light and assumes the book surface is cylindrical. These are often not true under real scanning conditions. Another limitation is its high computational cost in dealing with interreflections even with the tessellation method they propose. In fact, the interreflections may be ignored, since they mainly affect the illumination on the space around the book spine and thus have little effect on the estimated surface shape. We developed an efficient restoration system [6] based on Wada’s method by ignoring interreflection, but with the spine alignment constraint. As a further work of [6], our goal is to provide an efficient solution that can work with any scanner or photocopier in practice. Thus, our method here does not require the alignment of the book spine with the scanning light axis as [5] and [6], which is unrealistic under normal scanning conditions. Basically we use two models to uncover the geometric and lighting information with respect to the book surface and the scanner’s structure, which enable us to transform the warped image into its flattened rendition.
We first build the practical models for the real scanning conditions. Then, we restore the image by performing de-shading and de-warping using the two models. Finally, as a measure of success, we present experimental results to show the improvement of OCR.

2. Practical Model

Figure 1(a) shows the structure of the image scanner and the coordinate system of the book surface. The image scanner consists of a light source $L$, a linear CCD sensor $C$, a mirror $M$ and a lens $E$. The sensor $C$ takes a 1D image $P(i)\ (0 \leq i < \text{image height} I)$ along the scanning line $S$ and moves with $L$, $M$, and $E$. The sequence $P(i)$ forms a 2D image $P(i,j)\ (0 \leq j < \text{image width} J)$. Note that in real scanning conditions, the book being scanned may not be aligned strictly parallel to the scanning light. Let $\varepsilon$ be the angle between the scanning line $S$ and the projection of book spine on the scanning plane. We define the 3D coordinate system of the book surface as shown in Figure 1(a). In Figure 1(b), the relation between the 3D $x$-$y$-$z$ coordinates and 2D $i$-$j$ image indices are computed by:

\[
x(i,j) = \sqrt{i^2 + j^2} \cdot \cos(\arctan \frac{j}{i} - \varepsilon)
\]

\[
y(i,j) = \sqrt{i^2 + j^2} \cdot \sin(\arctan \frac{j}{i} - \varepsilon)
\]

\[
x(i) = \sqrt{i^2 + y^2} \cdot \cos(\arctan \frac{y}{x} + \varepsilon)
\]

\[
y(i) = \sqrt{i^2 + y^2} \cdot \sin(\arctan \frac{y}{x} + \varepsilon)
\]

Figure 1. The practical scanning conditions

The document skew $\varepsilon$ is detected by Hough transform on shade boundary. Figure 1(c) shows the cross section shape of the book surface in the $j$-$z$ plane, where $V$ is vertical to the scanning plane, $L_0$ the light source direction, $\psi$ the tilt angle between the light source and the normal to the scanning plane, $z$ the distance from $A$ to the scanning plane, $d$ the distance from scanning plane to light source.

We specify our problem by two practical models: a 3D geometric model and a 3D optical model. We introduce the following assumptions in our models:

**Assumption 1:** The cross section shape of the book surface is smooth on the $y$-$z$ plane (except for the points on the book spine) and constant along the $x$-axis.

**Assumption 2:** The book surface is Lambertian, i.e. no specular reflections and uniform brightness of reflected light in all directions.

2.1. Geometric Model

In order to reduce the 3D book surface to a unique 2D cross section shape, we define four lines in $x$-$y$ plane: $x = x_0$, $x = x_M$, $y = y_0$, and $y = y_N$, where $x_0 = \max(0, x(I - J - 1))$, $x_M = \min(x(I - 1), x(I - 1, J - 1))$, $y_0 = \max(0, y(J - 1, 0))$, $y_N = \min(y(0, J - 1), y(I - 1, J - 1))$. $I$ and $J$ are the image height and width, function $x$ and $y$ refer to equation (1) and (2) respectively. We define the processing area of the image to be the rectangle bounded by the above four lines. Figure 2 shows a typical scanned grayscale document image and its processing area. Note the shade along the book spine area and the warping in the shade.

Figure 2. A grayscale image scanned from a skew bound document, and its processing area

Suppose the 3D point $A$ in Figure 1(b) corresponds to the pixel $P(i, j)$ in the scanned document image. By assumption 1, the $z$ values are constant along $x$-axis.
Thus the 3D coordinates of $A$ in the $x$-$y$-$z$ space can be represented as $(x(i, j), y(i, j), z(y(i, j)))$, where $x(i, j)$ and $y(i, j)$ are calculated from (1) and (2) respectively. Figure 3(a) shows a cross section shape $P’AQ’$ of the book surface at $(x(i, j))$ in $y$-$z$ plane. Note that plane $PAQ$ in Figure 1(b) is perpendicular to the $i$-axis (also the light source), while plane $P’AQ’$ here is perpendicular to the $x$-axis (also the book spine). Figure 3(b) shows the cross section shape in $x$-$y$ plane. Mathematically, we have $z(y(i, j)) = \int_{y(i)}^{y(j)} z(y)dy$.

Since $y(i, j)$ and $y_N$ are integers, and $z(y(i)) = 0$, in our discrete $y$-$z$ coordinate system, the depth map $z(y(i, j))$ is represented as:

$$
z(y(i, j)) = \sum_{y=y(i)}^{y(j)} z(y) = \sum_{y=y(i)}^{y(j)} \tan\theta(y_k) = \sum_{y=y(i)}^{y(j)} \theta(y_k), \quad (y_k \leq y(i, j) \leq y_N, \theta \in [0, \pi])$$

(5)

where $\theta(y_k)$ the slant angle at $y_k$.

![Figure 3. The cross section shape of the book surface in (a) x-y-z space (b) y-z plane](image)

### 2.2. Optical Model

We formulate our problem by considering the following factors: 1) proximal and moving light source, 2) Lambertian reflection, and 3) nonuniform albedo distribution.

We first consider an ideal shape-from-shading problem satisfying 1) a distant and fixed light source, 2) Lambertian reflection, and 3) uniform albedo. The problem under these ideal conditions is formulated as:

$$I_s(p) = I_i \cdot k \cdot \cos \varphi(p)$$

(6)

where $p$ denotes a 2D point in the scanned image, $I_s(p)$ the reflected light intensity observed at $p$, $I_i$ the illuminate intensity, $k$ the albedo on the surface, and $\varphi(p)$ the angle between the light source direction and the surface normal at the 3D point on the book surface corresponding to $p$.

With a proximal and moving light source, the illuminant intensity is no longer constant over the object surface. The illuminant intensity on one point $p$ is now a function of the location of $p$ and that of the light source corresponding to $p$. We formulate the problem as follows:

$$I_s(p) = I_i(s(p), l(p)) \cdot k \cdot \cos \varphi(p)$$

(7)

where $s(p)$ and $l(p)$ denote the 3D point on the book surface and light source location corresponding to $p$.

Finally, we can formulate our problem by incorporating Lambertian reflection and nonuniform albedo distribution characteristics into Equation (7):

$$I_s(p) = I_i(s(p), l(p)) \cdot k(s(p)) \cdot \cos \varphi(p)$$

(8)

where $k(s(p))$ denotes the albedo at $s(p)$.

By the coordinate system in Figure 2 and Equation (8), the relationship between the image intensity (pixel value) and the reflected light intensity is represented as follows:

$$P(i, j) = \alpha \cdot I_s(i, j) + \beta$$

(9)

where

- $P(i, j)$: The image intensity at $(i, j)$ in the observed image, i.e., the scanned document image.
- $\alpha, \beta$: The gain and bias of the photoelectric transformation in the image scanner respectively.
- $x(i, j), y(i, j), z(y(i, j))$: The $x$-$y$-$z$ coordinates of the 3D point corresponding to $P(i, j)$ calculated from Equations (1), (2), and (5) respectively.
- $I_i(z(y(i, j)))$: The illuminant intensity distribution on the $y$-$z$ plane when taking the 1D image at $(x(i, j), y(i, j))$, i.e., $I_i(z(y(i, j)))$ is the practical representation of $I_i(s(p), l(p))$. Due to the constant tilt angle $\psi$ of the directional linear light source of the scanner, as shown in Figure 1(b) and (c), the 3D points with same $z$ value have the same distance $(AD + BD)/\cos \psi$, i.e., $(z(y(i, j)) + d)/\cos \psi$, from the light source. Note that the document skew $\varepsilon$ has no effect on this distance, and the book surface depth $z(y(i, j))$ is the only effective factor. Based on the directional linear light source model, $I_i(z(y(i, j)))$ can be represented as follows:

$$I_i(z(y(i, j))) = \frac{I_o(\psi)}{(z(y(i, j)) + d)/\cos \psi} = \frac{I_o(\psi) \cdot \cos \psi}{z(y(i, j)) + d}$$

(10)

where $\psi$ is the tilt angle between the light source direction and the normal to scanning plane, $I_o(\psi)$ the directional distribution of the illuminant intensity.
The cosine of the angle \( \phi \) between the light source direction and the surface normal at the 3D point \( (x(i, j), y(i, j), z(i, j)) \). Since the directional linear light source of the scanner has a constant tilt angle \( \psi \) and from assumption 1, both the light source direction and the surface normal are constant along x-axis, we can represent \( \phi \) as \( \phi(y(i, j)) \). The relation between \( \phi(y(i, j)) \) and \( \theta(y(i, j)) \) in geometric model (Equation (5)) can be derived as follows [7]:

\[
\cos \phi(y(i, j)) = \sqrt{\cos^2 \psi + \cos^2 \psi \cdot \cos \theta(y(i, j)) - \arctan(\cos \cdot \tan \psi)}
\]

In the experiments, parameters \( \alpha, \beta, I_0(\psi) \) are estimated a priori using a calibration image, parameters \( d, \psi \) are estimated by measurement of the geometry of the projected scanner light.

3. Reconstruction of Book Surface Shape and Albedo Distribution

In this section, we discuss how to reconstruct the depth \( z(y(i, j)) \) and discover the albedo distribution \( k(x(i, j), y(i, j)) \) by adopting the two practical models.

3.1. Reconstruction of Book Surface Shape

Since most of book surfaces have a uniformly colored background (typically an unprinted white background), we can assume that pixel values \( P_s(y_s) \) (\( y_0 \leq y_s \leq y_s \)) corresponding to the background can be obtained as the maximum for each \( y \) value in the scanned document image, i.e. we scan the image pixels along the direction of x-axis (in the \( x-y-z \) 3D coordinate system) instead of i-axis (in the i-j 2D image index system), and set \( P_s(y_s) \) as the maximum intensity value of each scan:

\[
P_s(y_s) = \max P(i(x_n, y_s), j(x_n, y_s))
\]

(\( x_0 \leq x_n \leq x_m, y_0 \leq y_s \leq y_s \))

where \( i(x_n, y_s), j(x_n, y_s) \) is the corresponding image index for \( (x_n, y_s) \), which are calculated by (3) and (4).

We then extract the global maximum pixel value \( P_s^{\max} \) over all \( y_s \) columns by:

\[
P_s^{\max} = \max P_s(y_s)
\]

(12)

By Equation (9), the constant albedo \( k_w \) of the white background is calculated as follows:

\[
k_w = \frac{P_s^{\max} - \beta}{\alpha - I_0(0) \cdot \cos \psi}
\]

(14)

The optical model of the white background with the constant albedo \( k_w \) is represented as follows:

\[
P_s(y_s) = \alpha \cdot k_w \cdot I_0(z(y_s)) \cdot \cos \phi(y_s) + \beta
\]

(15)

By substituting Equation (10) into Equations (14) and (15), we derive the following representation of \( z(y_s) \) from the optical model:

\[
z(y_s) = \left[ \frac{(P_s^{\max} - \beta) \cdot \cos \phi(y_s)}{P_s(y_s) - \beta} \cdot \cos \psi - 1 \right] \cdot d
\]

(16)

Replacing \( \cos \phi(y_s) \) by Equation (11), we have:

\[
z(y_s) = \left[ \rho \cdot \cos(\theta(y_s) - \arctan(\cos \cdot \tan \psi)) \right] \cdot \frac{d}{\rho} - 1
\]

where \( \rho \) is a constant represented as:

\[
\rho = \frac{(P_s^{\max} - \beta) \cdot \sqrt{\cos^2 \psi + \cos^2 \psi}}{\cos \psi}
\]

(18)

By substituting \( y_s \) to Equation (5), we derive another representation of \( z(y_s) \) from the geometric model:

\[
z(y_s) = \sum_{\gamma \in y_s} \tan \theta(y_s) \quad (y_0 \leq y_s \leq y_s, \theta \in [0, \frac{\pi}{2}])
\]

(19)

Now we have two representations of \( z(y_s) \), i.e. Equation (17) from the optical model and Equation (19) from the geometric model, and there are only two unknowns, \( z(y_s) \) and \( \theta(y_s) \). Moreover, the initial values of \( z(y_s) \) and \( \theta(y_s) \) are known: \( z(y_s) = 0 \) and \( \theta(y_s) = 0 \).

Thus in theory, the solution of the depth map \( z(y_s) \) can be found. We propose the following algorithm to compute the numerical solutions of \( z(y_s) \) and \( \theta(y_s) \):

Step 1. Rewrite Equation (17) in the following form:

\[
\cos(\theta(y_s) - \arctan(\cos \cdot \tan \psi)) = \frac{P_s(y_s) - \beta}{\rho} \cdot \left[ \frac{z(y_s) + \frac{d}{\rho}}{d} \right]
\]

(20)

Thus if \( \theta(y_s) < \arctan(\cos \cdot \tan \psi) \)

\[
\theta(y_s) = \arctan(\cos \cdot \tan \psi) - \arccos\left[ \frac{P_s(y_s) - \beta}{\rho} \cdot \left( \frac{z(y_s) + \frac{d}{\rho}}{d} \right) \right]
\]

(21)

else

\[
\theta(y_s) = \arctan(\cos \cdot \tan \psi) + \arccos\left[ \frac{P_s(y_s) - \beta}{\rho} \cdot \left( \frac{z(y_s) + \frac{d}{\rho}}{d} \right) \right]
\]

(22)

Step 2. Set \( z(y_s) = 0 \) , \( \theta(y_s) = 0 \) , and Exceed=0 (since \( \theta(y_s) \) is monotonously increasing starting with value 0 from \( y_0 \) to \( y_s \), this control flag “Exceed” indicates whether the \( \theta \) value has exceeded \( \arctan(\cos \cdot \tan \psi) \).

Step 3. For \( (y_s = y_s - 1, y_s \geq y_0, y_s \rightarrow -\infty) \) 

3.1 Compute \( z(y_s) \) by Equation (19).

3.2 If \( \theta(y_s) + 1 \geq \arctan(\cos \cdot \tan \psi) \), set Exceed =1.
3.3 If Exceed = 0, compute $\theta(y_n)$ by substituting $z(y_n)$ to Equation (21), else compute $\theta(y_n)$ by substituting $z(y_n)$ to Equation (22).

End of the Algorithm

Figure 4 shows the book surface shape discovered by our algorithm.

3.2 Reconstruction of Albedo Distribution

By Equation (9), the albedo distribution is represented as:

$$k(x(i, j), y(i, j)) = \frac{P(i, j) - \beta}{\alpha \cdot I_z(\theta(y(i, j))) \cdot \cos \theta(y(i, j))}$$

(23)

Substituting Equations (11) into (23), we have:

$$k(x(i, j), y(i, j)) = \frac{P(i, j) - \beta}{\alpha \cdot I_z(\theta(y(i, j))) \cdot \cos \theta(y(i, j))}$$

(24)

where $k(x(i, j), y(i, j))$ is calculated by Equation (24), $I_z(\theta(y(i, j)))$ is calculated by Equation (10). Therefore, our restoration system recalculates the image intensity for each pixel by Equation (25).

4.2. De-warping Model

Taking the image generated by de-shading model as input, we build our de-warping model. Note that distortion along y-axis is only due to orthogonal projection, while perspective distortion along x-axis is due to perspective projection. Thus the de-warping model consists of three parts: 1) Restoration along x-axis; 2) Restoration along y-axis; 3) Correction of document skew $\varepsilon$.

4.2.1. Restoration along x-axis

Figure 5 shows a slice of x-z plane at $y_n$, where $PP'$ denotes a line parallel to x-axis on the book surface, $PP''$ the line on the scanning plane corresponding to $PP'$ viewed by the CCD sensor (i.e. $PP''$ is the projection of $PP'$ in the image plane), $PP'' = z(y_n)$, $PP'' = 2 \times PP'P_0 = x_m - x_0$, the focal length $f = PP'P_0$.

4.2.2. Restoration along y-axis

Distortion along y-axis is only due to orthogonal projection as shown in Figure 6. Thus the correction along y-axis is done by two steps: First we add the estimated true width $w$ calculated from Equation (26), and then we stretch the observed image to its true width along y-axis using similar method discussed in the previous section.
4.2.3. Correction of document skew $\varepsilon$

We simply rotate the image at the center by $\varepsilon$ degree, and the final restored image is shown in Figure 7. The experiment shows that the geometric and photometric distortions are mostly removed. This result demonstrates that the shape is estimated accurately enough for the image restoration task.

5. OCR Results Comparison

Our restoration work does not require the presence of text lines, but we use OCR on the text lines to provide a measure of success. We carried out OCR process before and after the image restoration using three well-known commercial OCR products: ScanSoft OmniPage, IRIS Readiris and ABBYY FineReader. Precision and recall [8] defined in terms of correctly identified characters are used as the comparison metrics for the OCR results.

We scanned 100 document pages with resolution 200, 300 and 400 ppi respectively. Table 1 summarizes the improvements on precision/recall and the average runtime on a PIV 2.6G PC.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>OP Average Precision (%)</th>
<th>RI Average Precision (%)</th>
<th>FR Average Precision (%)</th>
<th>OP Average Recall (%)</th>
<th>RI Average Recall (%)</th>
<th>FR Average Recall (%)</th>
<th>Average Runtime (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 ppi</td>
<td>12.8</td>
<td>13.1</td>
<td>13.8</td>
<td>8.0</td>
<td>14.8</td>
<td>17.4</td>
<td>2.12</td>
</tr>
<tr>
<td>300 ppi</td>
<td>13.4</td>
<td>11.4</td>
<td>13.6</td>
<td>7.8</td>
<td>14.0</td>
<td>16.8</td>
<td>4.71</td>
</tr>
<tr>
<td>400 ppi</td>
<td>13.0</td>
<td>10.6</td>
<td>13.5</td>
<td>7.5</td>
<td>14.1</td>
<td>16.4</td>
<td>8.38</td>
</tr>
</tbody>
</table>

OP: OmniPage; RI: Readiris; FR: FineReader

6. Conclusion

In this paper, we describe the problems of distorted images while scanning thick, bound documents. We reconstruct the book surface shape and discover the albedo distribution based on two practical models. We restore the image by removing the shade and adjusting the warped book surface based on de-shading and de-warping models respectively. We present an experiment to show the improvement of OCR results.

Acknowledgement

The research is supported in part by National University of Singapore URC grant no. R252-000-202-112 and A*STAR, Singapore, grant no. 042 101 0085.

References