Restoring Warped Document Images using Shape-from-Shading and Surface Interpolation

Li Zhang
National University of Singapore
School of Computing
3 Science Drive 2, Singapore 117543
zhangli@comp.nus.edu.sg

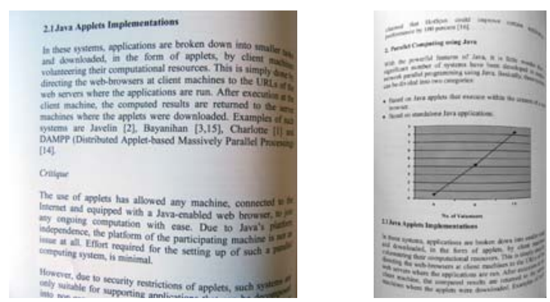
Chew Lim Tan
National University of Singapore
School of Computing
3 Science Drive 2, Singapore 117543
tancl@comp.nus.edu.sg

Abstract

With current high resolution handheld digital devices such as camera phones and PDAs, image capturing of documents like books and posters has become a convenient and efficient way of collecting and disseminating information. Nevertheless, a simple snapshot of such documents in an uncontrolled environment often results in distorted images. One particular example is when capturing documents with non-planar geometric shapes, such as thick bound book pages, rolled posters, etc. The resultant images often exhibit both perspective and geometric distortions. This paper proposes a method to remove these warping distortions through a shape recovery process based on Shape-from-Shading (SFS) followed by restoration using surface interpolation techniques. We evaluated the proposed method using various snapshot images captured by normal handheld digital cameras. The OCR results on the original images and the restored images are compared. The precision is shown to be increased up to 22.3%. The comparison with a 2-D interpolation approach also shows a clear improvement on the restored images near the warping area.

1. Introduction

With the pervasive use of current handheld digital devices such as camera phones, PDAs and video cameras, people have started to capture documents such as books, posters and notes as a means of recording information. In some sense, cameras are now functioning as personal copiers which allow one to gather information whenever and wherever is needed. However, due to the uncontrolled environment of the imaging process, this will soon produce a significant number of document images that are difficult to handle manually. For instance, the physical warping of a document could result in images with both geometric and photometric distortions as shown in Figure 1. This largely affects the quality of the captured images and introduces problems to machine recognition tasks such as OCR. Techniques are thus needed to remove these distortions and produce a flattened image for better human perception and document analysis tasks.



(a) Image1 (b) Image2

Figure 1. Warped document images.

Traditional 2-D de-warping approaches generally make use of 2-D image processing techniques to deal with geometric transformations between the input and the output images as in [1, 9, 6, 13]. Each point in the output image is mapped to a corresponding point in the input image through a spatial transformation function. This function can take different forms such as affine, projective, bilinear and polynomial. However, to find the spatial transformation based on solely 2-D information is difficult and sometimes inaccurate. One main reason is that the polynomials estimated based on text lines or boundary curves cannot precisely represent the geometric and photometric distortions due to pinhole camera's perspective projection. To address the limitations of 2-D based approaches, many attempts have been made to utilize the document geometric shape which can be obtained either through special 3-D setups or from 2-D image analysis as in [4, 8, 11, 2]. Most of the existing 3-D based approaches require special setups such as 3-D scanners or stereo systems to capture the shape information, though their restoration processes are less content dependent. In this paper, we try to address the problem of inaccurate approximation of geometric transformations based on pure 2-D information by incorporating 3-D document shape into a surface interpolation process. First, we obtain the document surface shape using an orthographic SFS technique with an oblique lighting direction. Next, we model the warped surface as an applicable surface by approximating two boundary curves based on the text lines in the document image. Finally, each point in the output flat image is mapped to a corresponding point in the warped image through a surface interpolation function.

2. Shape reconstruction using SFS

In order to model the geometric transformation more accurately, we first try to estimate the surface shape using a SFS technique. The fast marching method proposed by Kimmel and Sethian [5] allows us to derive the surface height relative to the image plane based on the shading variations in the 2-D image. To apply this, we first need to extract the shading image from the original document image. This is done by applying Niblack's local adaptive binarization method [7] to filter out those text and graphics regions and keep the intensity values in the remaining area. The filtered regions can then be filled up with intensities derived from interpolating the neighboring pixels. The resultant image is then smoothed out to give the approximated shading image to be used in the SFS process.

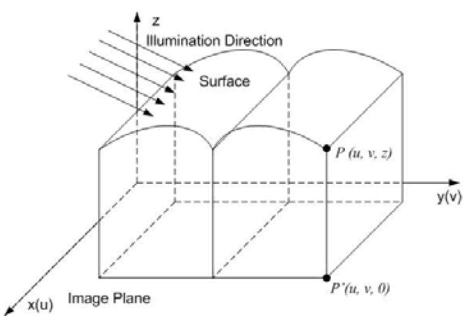


Figure 2. Orthographic SFS formulation.

2.1. Orthographic SFS formulation

For simplicity, we assume an orthographic projection in our SFS formulation as shown in Figure 2. Our goal is to obtain the height z of each surface point $P(u, v, z)$ relative to the u-v image plane Ω. Suppose the projection of the point P in the image plane is denoted as $P'(u, v, 0)$, the document surface can be represented as:

$$S = \{(u, v, z), (u, v) \in \Omega\} \quad (1)$$

Assuming the illumination is a distant point light source with a known direction $L = (\alpha, \beta, \gamma)$, w.l.o.g. we can let $\gamma = -1$. Now, let $p = \frac{\partial z}{\partial u}$ and $q = \frac{\partial z}{\partial v}$, the surface normal at each point x can be derived as $N(x) = (p, q, -1)$. Further assuming that the document surface follows Lambertian reflection, we can obtain the image irradiance equation based on Lambert's cosine law as:

$$I(u, v) = \frac{N(x) \cdot L}{\|N(x)\| \cdot \|L\|} = \frac{p\alpha + q\beta + 1}{\sqrt{p^2 + q^2 + 1} \cdot \sqrt{\alpha^2 + \beta^2 + 1}} \quad (2)$$

To solve the above partial differential equation, we refer to the fast marching algorithm proposed by Kimmel and Sethian [5]. They give the image irradiance equation for the orthographic projection with vertical illumination direction $(0, 0, -1)$ in the form of an Eikonal equation:

$$p^2 + q^2 = \frac{1}{I(u, v)^2} - 1 \quad (3)$$

The solution $z'(u, v)$ of Eq. (3) is obtained using a fast marching algorithm. In our case of an oblique light source, the height value $z'(u, v)$ obtained by considering vertical illumination direction is actually the surface height in a new coordinate system, where the light source $L = (\alpha, \beta, -1)$ is located at $L = (0, 0, -1)$. Therefore, the true height $z(u, v)$ can be obtained through a simple linear transformation:

$$z(u, v) = \alpha + \beta + \|L\| - \|L\|^2 + z'(u, v) \cdot \|L\| \quad (4)$$

An example of the recovered surface shape for Image 1 in Figure 1(a) is shown in Figure 3.

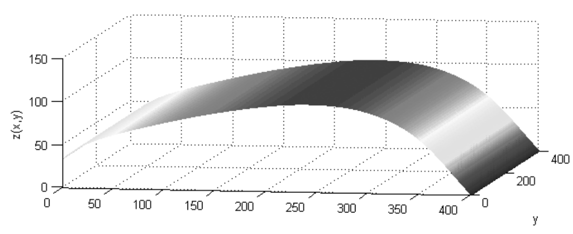


Figure 3. Recovered shape for Image1.

3. Warping restoration

Assuming the book spine is along an iso-parametric line, i.e. its projection in the image plane is a straight line parallel to v-axis as shown in Figure 4, we can then construct a ruled surface model based on the horizontal text lines present in the document. Given a set of text lines $\big(C_1(u), C_2(u), \cdots, C_m(u)\big)$ on the 3-D surface and two vertical boundaries, we can construct a surface g that interpolates all the iso-parametric curves $\{g(u, v_i) \quad i = 0, \cdots, m\}$ and $\{g(u_j, v) \quad j = 0, \cdots, n\}$:

$$g(u, v) = g_1(u, v) + g_2(u, v) - g_{12}(u, v) \tag{5}$$

where

$$\begin{aligned} g_1(u, v) &= \sum_{i=0}^{m} g(u, v_i) L_i^m(v) \\ g_2(u, v) &= \sum_{j=1}^{n} g(u_j, v) L_j^n(u) \\ g_{12}(u, v) &= \sum_{i=0}^{m} \sum_{j=0}^{n} g(u_j, v_i) L_j^n(u) L_i^m(v) \\ L_i^m(v) &= \frac{\prod_{j=0, j\neq i}^{m} v - v_j}{\prod_{j=0, j\neq i}^{m} v_i - v_j} \end{aligned} \tag{6}$$

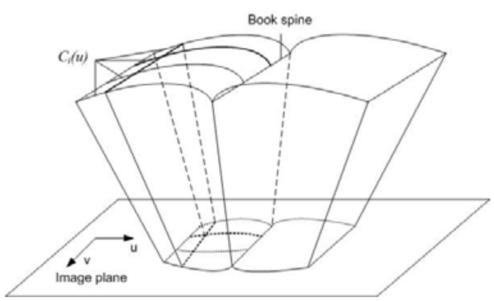


Figure 4. Ruled surface model.

It is observed that the iso-parametric lines parallel to the book spine and intersecting with the curves $C_1(u), \cdots, C_m(u)$ must also pass through the corresponding points in the projection image as shown in Figure 4. Therefore, this ruled surface model can be parameterized using the projected text lines in the 2-D image together with the height information obtained through the SFS process.

To extract the text lines, we first identify all the connected components in the 2-D image. A set of boundary points are then selected based on the bottom boundaries of the connected components and interpolate to a Natural Cublic Spline to represent each text line as shown in Figure 5. Given $(n+1)$ points $(P_0, P_1, \cdots, P_n)$ and $(n+1)$ knot values $(t_0, t_1, \cdots, t_n)$ on a curve C_i, the spline is defined as a set of piece-wise functions:

$$s_i(t) = \begin{cases} s_{i,0}(t), & t \in [t_0, t_1] \\ s_{i,1}(t), & t \in [t_1, t_2] \\ \vdots \\ s_{i,n-1}(t), & t \in [t_{n-1}, t_n] \end{cases} \tag{7}$$

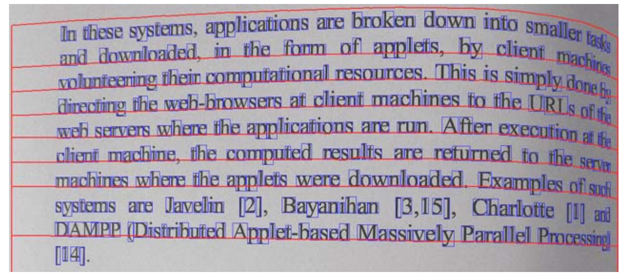

In these systems, applications are broken down into smaller tasks and downloaded, in the form of applets, by client machines volunteering their computational resources. This is simply done by directing the web-browsers at client machines to the URLs of the web servers where the applications are run. After execution at the client machine, the computed results are returned to the server machines where the applets were downloaded. Examples of such systems are Javelin [2], Bayanihan [3,15], Charlotte [1] and DAMPP (Distributed Applet-based Massively Parallel Processing) [14].

Figure 5. A portion of the document with extracted text lines.

While the $(n+1)$ points can be easily obtained from the boundary points, the knot values need to be carefully chosen. Intuitively, the knot values should be defined based on the Euclidean distance between the corresponding 3-D points on the surface. Given the height of each point on the surface, the knot values are defined as follows:

$$t_i = \begin{cases} 0, & \text{if } i = 0 \\ t_{i-1} + \frac{1}{|d|} \cdot |P_i - P_{i-1}|_2, & \text{if } i > 0 \end{cases} \tag{8}$$

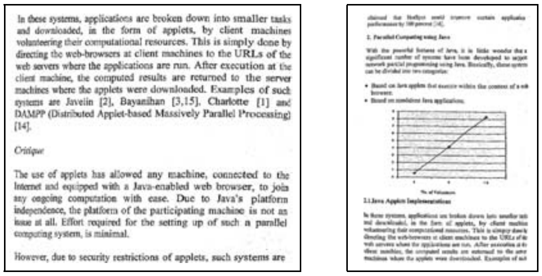
where $|d| = \sum_{i=1}^{n+1} |P_i - P_{i-1}|_2$.

The above parameterization assigns larger t values to those regions with more severe warpings. This allows more pixels being allocated to such regions in the restored image and thus brings back the correct font size and spacing in the warping area.

Finally, the restored image can be obtained based on the 2-D mapping function derived from the ruled surface model defined in Eq. (5). For each point (u_j, v_i) in the restored image, the corresponding point (x, y) is computed accordingly. For simplicity, we can select only the top and bottom text lines and the two vertical boundaries for interpolation using coons patch [3] as applied in the boundary interpolation method on warped art images by Tsoi and Brown [10].

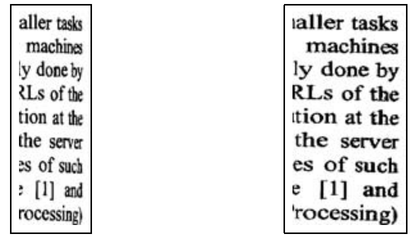
4. Experimental results

We have tested the proposed method on images captured using a normal digital camera with various resolutions and different warping distortions. Figure 6 shows the restored images for the two warped images in Figure 1 with resolutions 1200×1200 and 748×1064 respectively. The restored images are fed into OmniPage OCR engine for recognition. Table 1 shows the OCR precisions on four input images with different levels of distortion as well as the results on their corresponding restored images. Clearly, the precisions on the restored images are much higher than those on the original images. We also compared the present method with our earlier attempt based on 2-D interpolation [12].



(a) Restored Image1 (b) Restored Image2

Figure 6. Binarized version of the restored images in Figure 1.



(a) (b)

Figure 7. (a) Restored image portion using 2-D interpolation; (b) restored image portion using 3-D interpolation.

Table 1. Comparisons of OCR results.

Images	Resolution	OCR word precision	
		Original	Restored
Image1	1200 × 1200	89.6%	99.4%
Image2	748 × 1064	69.4%	91.7%
Image3	1024 × 768	79.4%	94.5%
Image4	1280 × 960	84.7%	96.4%

It is observed that the 2-D method is unable to restore the character size and spacing near the warping area as in Figure 7(a), while the current method does straighten out the text lines and restore the character font size and spacing as in Figure 7(b). Moreover, OCR experiments also show a better performance on the restored images comparing to those of the 2-D method.

5. Conclusion

In this paper, we present a method of restoring warped document images captured using handheld devices. The method is based on surface interpolation with text lines and surface height information derived from the given 2-D image. The restored image improves the OCR performance up to 22.3% in terms of word precision. The current method allows the presence of graphics contents in the document image, though it assumes the existence of two vertical boundaries and a single column text. As a future work, the SFS technique can be further improved to produce more accurate surface shapes with a relaxed condition on the light source direction.

6. Acknowledgement

This research is supported in part by National University of Singapore URC grant R252-000-202-112 and Agency for Science, Technology and Research(A*STAR) grant R252-000-206-305.

References

[1] H. Baird. Document image defect models and their uses. *International Conference on Document Analysis and Recognition*, pages 730–734, 1993.

[2] H. Cao, X. Ding, and C. Liu. A cylindrical model to rectify the bound document image. *International Conference on Computer Vision*, 2:228–233, 2003.

[3] S. Coons. Surfaces for computer aided design. *Technical report, MIT*, 1968.

[4] T. Kanungo, R. M. Haralick, and I. Phillips. Nonlinear global and local document degradation models. *International Journal of Imaging System and Technology*, pages 220–230, 1994.

[5] R. Kimmel and J. A. Sethian. Optimal algorithm for shape from shading and path planning. *Journal of Mathematical and Imaging and Vision*, 14(3):237–244, 2001.

[6] O. Lavaille, X. Molines, F. Angella, and P. Baylou. Active contours network to straighten distorted text lines. *International Conference on Image Processing*, pages 1074–1077, 2001.

[7] W. Niblack. *An Introduction to Digital Image Processing*. Prentice Hall, 1986.

[8] M. Pilu. Undoing page curl distortion using applicable surfaces. *IEEE International Conference on Computer Vision and Pattern Recognition*, 1:67–72, 2001.

[9] Y. Y. Tang and C. Y. Suen. Image transformation approach to nonlinear shape restoration. *IEEE transactions on Systems, Man, and Cybernetics*, 23(1):155–171, 1993.

[10] Y. C. Tsoi and M. S. Brown. Geometric and shading correction for images of printed materials c a unified approach using boundary. *IEEE International Conference on Computer Vision and Pattern Recognition*, 1:240–246, 2004.

[11] A. Yamashita, A. Kawarago, T. Kaneko, and K. T. Miura. Shape reconstruction and image restoration for non-flat surface of document with a stereo vision system. *IEEE International Conference on Pattern Recognition*, 2004.

[12] L. Zhang and C. L. Tan. Warped image restoration with applications to digital libraries. *8th International Conference on Document Analysis and Recognition*, pages 192–196, 2005.

[13] Z. Zhang and C. L. Tan. Straightening warped text lines using polynomial regression. *International Conference on Image Processing*, pages 977–980, 2002.