Global and Local Document Degradation Models

Tapas Kanungo, Robert M. Haralick and Ihsin Phillips

Intelligent Systems Laboratory
Department of Electrical Engineering, FT-10
University of Washington
Seattle, WA 98195

Email: {tapas,haralick,yun}@ee.washington.edu

Abstract

Two sources of document degradation are modeled – i) perspective distortion that occurs while photocopying or scanning thick, bound documents, and ii) degradation due to perturbation in the optical scanning and digitization process: speckle, blurr, jitter, threshold. Perspective distortion is modeled by studying the underlying perspective geometry of the optical system of photocopiers and scanners. An illumination model is described to account for the non-linear intensity change occuring across a page in a perspective-distorted document. The optical distortion process is modeled morphlogically. First a distance transform on the foreground is performed and followed by a random inversion of binary pixels where the probability of flip is a function of the distance of the pixel to the boundary of the foreground. Correlating the flipped pixels is modeled by a morphological closing operation.

1 Introduction

There are many reasons for modeling document degradation. First, in order to study the performance of any OCR algorithm, it is necessary to characterize the perturbation in the output performance as a function of the perturbation in the input. This is possible only if we have a perturbation/degradation model for the input document. Second, a degradation model permits the evaluation of an algorithm for a continuum of degradation levels, from low degradation levels to high degradation levels. This in turn allows us to locate the 'break-down' point or the 'knee' of the algorithm which is not available in the commonly used evaluation methods, e.g., confusion matrices. Third, a knowledge of the degradation model can enable us to design algorithms for restoring degraded documents. Furthermore, OCR algorithm designers can make use of these degradation models explicitly rather than implicitly as is usually done in current literature.

In this paper we model two sources of document degradation – i) perspective distortion that occurs while photocopying or scanning thick, bound documents, and ii) degradation due to perturbation in the optical process: speckle, blur, jitter, threshold, etc.. Perspective distortion is modeled by studying the underlying perspective geometry of the optical system of photocopiers and scanners. An illumination model is proposed to account for the non-linear intensity change occurring across a page in a perspective-distorted document. The local optical distortion process is modeled morphologically. First a distance transform on the foreground is performed and followed by a random inversion of binary pixels where the probability of flip is a function of the distance of the pixel to the boundary of the foreground. Correlating the flipped pixels is modeled by a morphological closing operation.

Baird [Bai90] discusses a model for character degradation. His model does not account for the non-linear distortions produced from perspective distortions. Loce [LL90] models the perturbation introduced due to mechanical disturbances in high-end Xerox photocopiers. Our paper models the distortions in geometry and illumination due to perspective. This is a page level distortion model as opposed to the noise model proposed by Baird which works at character and pixel level. Furthermore, our model of the optical process is morphological and more conducive for degradation parameter estimation. Maltz [MS88] has done a transfer function analysis of the Xerography process.

2 A Global Distortion Model for Perspective and Non-linear Illumination Effects

A typical setup for scanners and photocopiers is shown in the figure 1. In the figure we have shown a book that is to be photocopied. The page to be photocopied is not flat on the document glass since the book is tightly bound and the 'spine' of the book is thick. We model four sources of degradation in the following subsections.

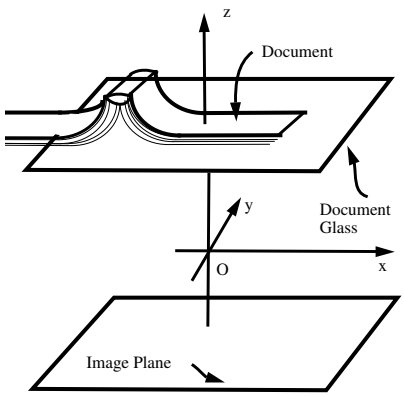

Figure 1: The setup while photocopying a thick, bound document. The center of perspectivity is at O, which is also the origin of the coordinate frame.

2.1 Deformation Model for the Physical Page Bending Process

First the page it self undergoes a physical deformation where the document page goes through a 'bending' process near the 'spine' of a thick, bound document. The page is no longer a flat surface on the document glass but a curved surface bending away from the glass near the spine of the book. We model this curved portion of the document page as a circular arc segment along the x axis and assume that there is no such deformation along the y axis. The global rotation and translation can be modeled in another stage. Figure 2 illustrates this deformation phenomenon.

Let $A = (x_a, y_a, f)'$, $B = (x_b, y_b, f)'$. Furthermore, let ρ be the radius of the deformation circle and let the bent segment subtend an angle θ at the center of the circle D. Let the point A map to the point $A' = (x_{a'}, y_{a'}, z_{a'})'$ after deformation. Then the coordinates of A' are given by

$$x_{a'} = x_a + \rho(\theta - \sin\theta) \quad (1)$$

$$y_{a'} = y_a \quad (2)$$

$$z_{a'} = f + \rho(1 - \cos\theta) \quad (3)$$

Let the point $P = (x_p, y_p, f)'$, be such that $x_a \leq x_p \leq x_b$. and let P map to the point $P' = (x_{p'}, y_{p'}, z_{p'})'$ after deformation. Let the angle subtended by the arc $P'C$ at the center D be ϕ where

$$\phi = \frac{x_a + \rho\theta - x_p}{\rho} = \theta - \left(\frac{x_p - x_a}{\rho}\right). \quad (4)$$

Now the coordinates of P' can be calculated as given below

$$x_{p'} = x_p + \rho(\phi - \sin\phi) \quad (5)$$

$$y_{p'} = y_p \quad (6)$$

$$z_{p'} = f + \rho(1 - \cos\phi) \quad (7)$$

Note that for points P in the original document with $x_p > x_b$, we have no deformation and hence $P' = P$.

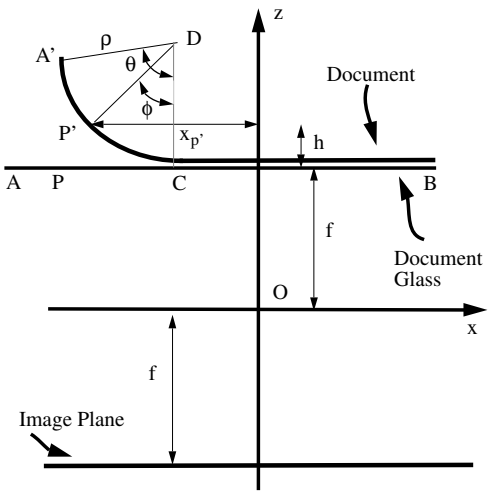

Figure 2: The bending deformation of the document pages. The side view of figure 1 while looking in the positive y direction is shown. The points A',B', and C' on the document page would have been at the points A, B, and C on the document glass if the page had not been curved. The curve $A'P'C$ is modeled as a circular arc segment that subtends an angle θ at the center, D, of the circle, which has a radius ρ. Here $h = \rho(1 - \cos\theta)$ and $x_{p'} = x_p + \rho(\phi - \sin\phi)$ where $\phi = \theta - (x_p - x_a)/\rho$. Rest of the page from C to B along the x axis is not deformed. It is assumed that the page does not undergo any deformation in the y direction either.

2.2 Perspective Distortion Model

The bending deformation is followed by a perspective distortion where the point P' on the document maps to the point P'' on the image. See figure 3. Let the focal length of the optical system be f and let the center of perspectivity, O, be at the origin. Assume that the image plane is at the focal plane at $-f$. Let $P'' = (x_{p''}, y_{p''}, z_{p''})'$ be the perspective projection of the point P' on the document page. The coordinates of P'' are given by the following equations [HS92, Hor86]

$$x_{p''} = -f\left(\frac{x_{p'}}{f + \rho(1 - \cos\phi)}\right) \quad (8)$$

$$= -f\left(\frac{x_p + \rho(\phi - \sin\phi)}{f + \rho(1 - \cos\phi)}\right) \quad (9)$$

$$y_{p''} = -f\left(\frac{y_{p'}}{f + \rho(1 - \cos\phi)}\right) \quad (10)$$

$$= -f\left(\frac{y_p}{f + \rho(1 - \cos\phi)}\right) \quad (11)$$

$$z_{p''} = -f \quad (12)$$

Note that for points P in the original document with $x_p > x_b$, we have no bending or perspective deformation and hence $-P'' = P' = P$.

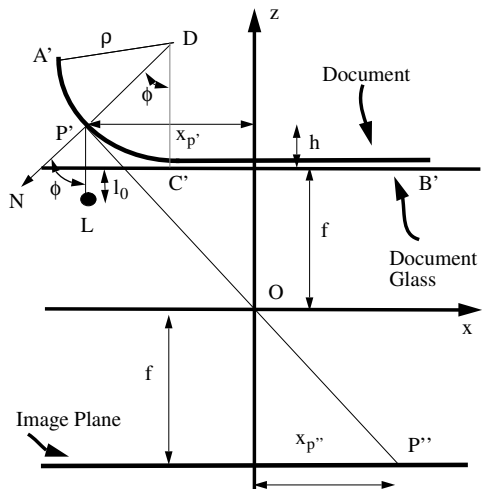

Figure 3: Perspective distortion. The point P' on the bent document page projects to the point P'' on the image plane. The coordinates of P'' are given as $x_{p''} = -f \cdot x_{p'}/(f+h)$, $y_{p''} = -f \cdot y_{p'}/(f+h)$, and $z_{p''} = -f$, where $h = \rho(1-\cos\phi)$ and $\phi = \theta - (x_p - x_a)/\rho$.

2.3 Non-linear Illumination Model

Since the document page is no longer flat but a curved surface, the illumination on the document is not constant. The illumination at a point P' on the document pages is inversely proportional to the distance of point P' from from the light source L. The light source L moves below the document glass from one end to the other. Let the distance between the document glass and the light source L be l_0. See figure 3. At the places where the page is curved the distance between the light source and the document pages is $l = l_0 + \rho(1-\cos\phi)$ where ϕ is the angle arc $P'B$ subtends at B. Note ϕ is also the angle between the normal at P' and the negative z direction. We model the illumination as a diffuse lighting model. Thus the intensity of light is proportional to the cosine of the angle ϕ. Furthermore, after reflection, the diffuse model assumes the intensity of light is same in all directions [HS92, Hor86]. Let I_0 be the intensity at a point where the document is not curved, i.e., the distance between the light and the point under consideration is l_0. Thus

$$I_0 \propto \frac{1}{l_0^2} \tag{13}$$

Next, the intensity at $I_{p'}$ a point on the curved part is proportional to $\cos\phi$ and inversely proportional to $(l_0 + \rho(1-\cos\phi)^2$. Thus

$$I_{p'} \propto \frac{\cos\phi}{(l_0 + \rho(1-\cos\phi))^2} \tag{14}$$

Thus taking a ratio of the above two equations we have

$$I_{p'} = I_0 \left(\frac{l_0}{l_0 + \rho(1-\cos\phi)} \right)^2 \tag{15}$$

Under the assumption of diffuse lighting we have $I_{p'} = I_{p''}$

2.4 Non-linear Optical Point Spread Function

In an imaging process if a point P' is not in the focal plane, it is not in focus in the image plane, if the image plane is at f. In fact, the image of a point geometrically is a disk if the image plane is not in focus [SG88, Pen88, Hor86, HS92]. See figure 4. If Δ is the diameter of the lens, and h is the distance of the image plane from the focal plane, then the diameter of the disk is given by

$$d = \Delta \left(\frac{h}{f} \right) \ . \tag{16}$$

But due to optical irregularities, in reality we do not get a disk as the image but blurred version of a disk. In fact this blurred disk can be modeled as a Gaussian with a standard deviation $\sigma = k \cdot d$, where k is a camera constant.

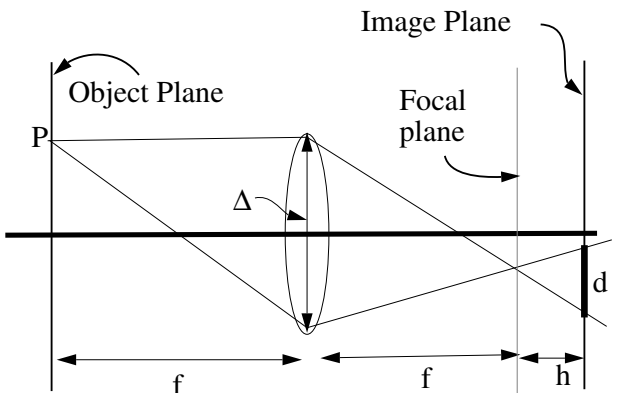

Figure 4: This figure illustrates the fact that if the image plane is not at focus then a point P maps to a disk of radius d. If the diameter of the lens is Δ and the focal length is f, the disk has a diameter $d = \Delta \cdot (h/f)$. Note in the real world the disk becomes blurred and can be approximated by a Gaussian. See text for more details.

Notice that in our case, the distance of a point on the document page is in focus only if the document page is in on the document glass (the focal plane). The curved region, in particular is not in focus since the points in that region are different distances from the focal plane. Thus the amount of blurring that a point goes through is different for the points on the curved segment.

Algorithmically, after performing the bending transformation, perspective distortion, and non-linear illumination, another stage is necessary where the image is convolved with a space-varying Gaussian kernel. The kernel has a standard deviation σ given by

$$\sigma = k \cdot \rho(1-\cos\phi) \ . \tag{17}$$

in the curved regions and constant σ_0. else where.

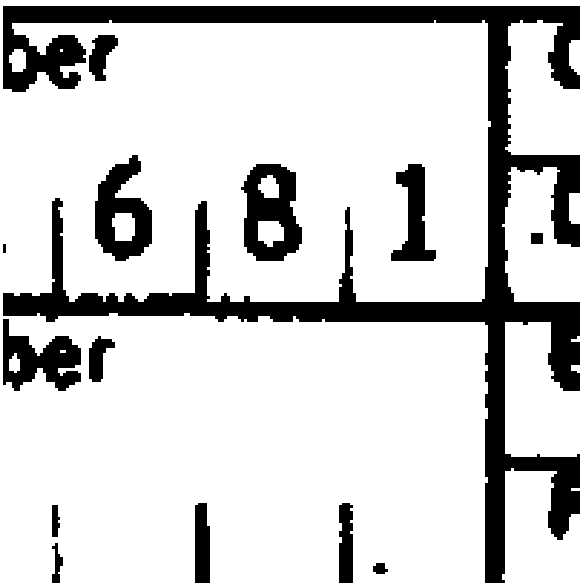

Figure 5: This is the original binary image before undergoing perspective distortion.

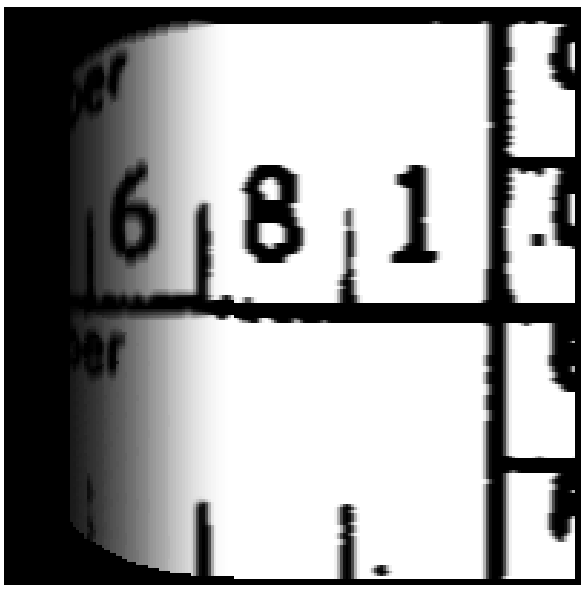

Figure 6: This image is produced after undergoing perspective distortion. Notice that the bend is very gradual and the intensity of light decreases as you go along the curved region. Furthermore, the text is no longer horizontal but curved inward. In addition, the blurring gets progressively worse toward the left edge of the image.

2.5 Simulation of the Perspective Distortion Model

In this section we show some simulation results of the model discussed thus far. The original non-distorted image is shown in figure 5. The dimensions of the image are 201×201. The convolution kernel size used was 5×5. Two perspective deformations are shown in figures 6 and 7. The parameters, in units of pixels, used for generating figure 6 were: $\rho = 152.87$, $\theta = 30^o$, $f = 80$, $\Delta = 20$, $k = 8$, $l_0 = 10$. The parameters used for generating figure 7 were: $\rho = 95.54$, $\theta = 30^o$, $f = 50$, $\Delta = 50$, $k = 1$, $l_0 = 20$.

3 A Morphological Model for Local Distortion

In this section we discuss a document degradation model. This noise model is based on the distance transform [HS92] of the ground truth data and some morphological post-processing. We model the probability of a pixel changing from its ideal value as a function of the distance of that pixel from the boundary of a character. Let d be the distance (four connected or eight connected) of a foreground or background pixel from the boundary of the character and α and β are scale parameter. Let $P(1|d, \beta, f)$ and $P(0|d, \beta, f)$ be the probability of a foreground pixel at a distance d to remain as 1 and to change to a 0, respectively. Similarly, let $P(1|d, \alpha, b)$ and $P(0|d, \alpha, b)$ be the probability of a background pixel at a distance d changing to a 1 and remaining a 0, respectively. The functions $P(1|d, \alpha, f)$ and $P(1|d, \alpha, b)$ could be different. The random perturbation process then proceeds to change pixel values in a pixel by pixel independent manner. This is followed by a morphological closing operation

Figure 7: This image is produced after undergoing perspective distortion which is sharper than the previous image. Also, the intensity variation is not as pronounce as the previous image.

to account for the correlation introduced by the optical point spread function preceding the thresholding operation which produces the noisy image.

The following forms for the background and foreground conditional probabilities were used in the images shown in figure 8.

$$P(1|d,\alpha,b) = 1 - P(0|d,\alpha,b) = e^{-\alpha d^2} \quad (18)$$

$$P(0|d,\beta,f) = 1 - P(1|d,\beta,f) = e^{-\beta d^2} \quad (19)$$

The closing operation was performed with a 2×2 binary structuring element.

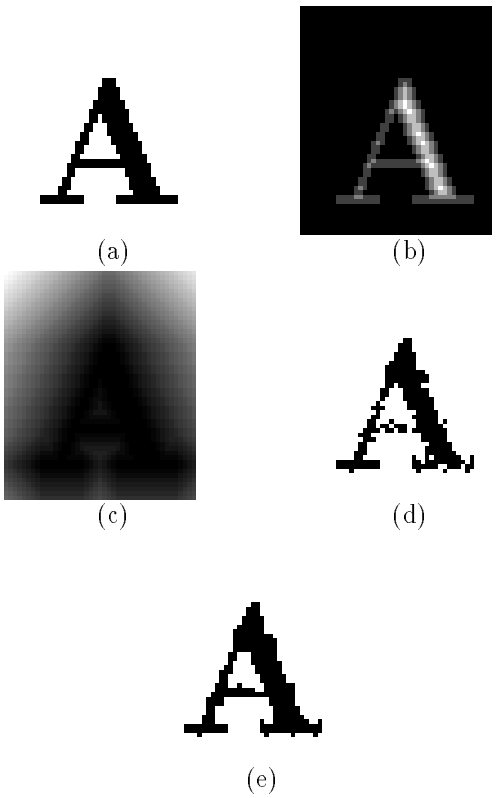

Figure 8: Morphological local distortion model: (a) Ground truth data; (b) Distance transforms of foreground (a); (c) Distance transforms of background (a); (d) The resultant perturbed image with exponential probability distribution $P(0|d,\beta,f) = P(1|d,\alpha,b) = e^{-\alpha d^2}$ and $\alpha = \beta = 2$; (e) closing of result in (d) by a 2×2 binary structuring element

The document degradation model proposed by Baird [Bai90, Bai93] models the physical degradation process. The degradation model is parameterized and some of the parameters are: blur, speckle, jitter, threshold values, size, rotation, etc. Three major noise parameters are blur, speckle, and the threshold value.

4 Conclusion

We describe a model for the perspective distortion occurring during the photocopying and scanning process. This model accounts for the physical deformation of the document page, perspective distortion, non-linear intensity variations, and non-linear optical point-spread function. We also described a morphological model for local distortions in terms of distance transforms and a morphological closing operation. Simulation results were given for both models. Two issues that have not been addressed are degradation model parameter estimation and model validation.

Acknowledgements: Authors would like to thank Henry Baird, Bob Loce and James Harrison for many valuable discussions; Larry Spitz for providing many references; and Jerry Mains for giving a detailed description of the photocopier internals.

References

[Bai90] H. Baird. Document image defect models. In *Proc. of IAPR Workshop on Syntactic and Structural Pattern Recognition*, pages 38–46, Murray Hill, NJ, June 1990.

[Bai93] H. Baird. Calibration of document image defect models. In *Proc. of Second Annual Symposium on Document Analysis and Information Retrieval*, pages 1–16, Las Vegas, Nevada, April 1993.

[FvDFH90] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes. *Computer Graphics.* Addison-Wesley, Reading, Massachesetts, 1990.

[Hor86] B.K.P. Horn. *Robot Vision.* The MIT Press, Cambridge, MA, 1986.

[Hou83] H. S. Hou. *Digital Document Processing.* John Wiley, New York, 1983.

[HS92] R.M. Haralick and L.G. Shapiro. *Machine Vision.* Addison-Wesley Publishing Co., Inc., Reading, Massachusetts, 1992.

[LL90] R. Loce and W. Lama. Half-tone banding due to vibrations in a xerographic image bar printer. *Journal of Imaging Technology*, 16(1):6–11, 1990.

[MS88] M. Maltz and J. Szczepanik. Mtf analysis of xerographic development and transfer. *Journal of Imaging Science*, 32(1):11–15, 1988.

[Pen88] A.P. Pentland. A new sense of depth of field. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 523–531, 1988.

[SG88] M. Subbarao and N. Gurumoorthy. Depth recovery from blurred edges. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 498–503, Ann Arbor, MI, 1988.