Automatic Performance Evaluation of Dewarping Methods in Large Scale Digitization of Historical Documents

Maryam Rahnemoonfar, Beth Plale
School of Informatics and Computing, Indiana University
maryrahn@indiana.edu

ABSTRACT

Geometric distortions are among the major challenging issues in the analysis of historical document images. Such distortions appear as arbitrary warping, folds and page curl, and have detrimental effects upon recognition (OCR) and readability. While there are many dewarping techniques discussed in the literature, there exists no standard method by which their performance can be evaluated against each other. In particular, there is not any satisfactory method capable of comparing the results of existing dewarping techniques on arbitrary wrapped documents. The existing methods either rely on the visual comparison of the output and input images or depend on the recognition rate of an OCR system. In the case of historical documents, OCR either is not available or does not generate an acceptable result. In this paper, an objective and automatic evaluation methodology for document image dewarping technique is presented. In the first step, all the baselines in the original distorted image as well as dewarped image are modelled precisely and automatically. Then based on the mathematical function of each line, a comprehensive metric which calculates the performance of a dewarping technique is introduced. The presented method does not require user interference in any stage of evaluation and therefore is quite objective. Experimental results, applied to two state-of-the art dewarping methods and an industry-standard commercial system, demonstrate the effectiveness of the proposed dewarping evaluation method.

Categories and Subject Descriptors

H.3.4 [**Systems and Software**]: Performance evaluation (efficiency and effectiveness); I.4.1 [**Digitization and Image Capture**]: *Imaging geometry*; I.4.3 [**Enhancement**]: *Geometric correction*

Keywords

Performance Evaluation, Arbitrary Warping, Large Scale Digitization

1. INTRODUCTION

Commercial OCRs are tailored for flat pages with straight text lines. Even moderate warping can cause most current OCR systems to fail. For modern documents, the importance of geometric distortion correction is placed on extracting the printed text and making it recognizable by the commercial OCR. However, for historical manuscripts the goal of digitization is to produce digital facsimiles that represent the original printed content as faithfully as possible. Here "the correction procedure" aims at producing the original flat document; that is, a document image which resembles the original page as much as possible.

While there are many dewarping techniques in the literature [2; 4; 6; 7; 9; 13; 14] there exists no standard methodology for the comparison of their performance. Most of the evaluations done so far concentrate on subjective and qualitative visual comparison of the output and the input image[3; 7]. In some cases, the recognition rate of an OCR system is used for the evaluation of dewarping techniques [10]. According to this approach, the increase of the recognition rate which OCR achieves between the original image and the dewarped image indicates the improvement of the dewarping method. Therefore the improvement of OCR as a performance evaluation metric is highly dependent upon the performance of the OCR itself. In the case of historical documents, however, often OCR does not generate acceptable results or no OCR system exists. In recent years, some quantitative dewarping evaluation methodologies [5; 11; 12] have been proposed in the literature. The proposed evaluation methodology in [5] uses the original camera-captured image previous to any distortion. In most cases and especially in historical documents, we do not have access to the perfect original image of any document that belongs to previous centuries. In the evaluation methodologies proposed in [11; 12], in the first stage, the user is supposed to mark several points on a number of text lines of the original warped image. The points should be selected on the long text lines which are considered to be most representative of the image. The rest of the method essentially involves matching a polynomial curve fitting and an area approximation by integration. The advantage of this method is that it does not depend on OCR for evaluation. However this approach is strongly dependent upon the user. Therefore, the evaluation is inevitably very subjective. Moreover in many historical documents, arbitrary warping is common, which is to say that each line has its own distortion. In such a case, selecting a few lines (for example, 3 or 6 lines) in a page with approximately 30 arbitrary text lines cannot guarantee a comprehensive evaluation. In addition, if a dewarping method fails to work properly and ends up distorting shorter lines, its performance cannot be evaluated by this evaluation methodology. It also has to be noted that polynomial curve fitting is not suitable in the arbitrary page situation, as it is capable of modelling only very smooth page curls.

In this paper a flexible, accurate, quantitative and objective performance evaluation methodology which is independent of OCR or human interference is presented. In the first step, the precise baselines are modelled with a robust mathematical formula presented in [9] both on the original image and on the dewarped one. The proposed baseline detection method avoids using any curve fitting or energy minimization concept such as the snake method and therefore is both more accurate and more

JCDL'13, July 22–26, 2013, Indianapolis, Indiana, USA.

ACM 978-1-4503-2077-1/13/07...$15.00.

flexible and can be applied on a wide range of curves. In methods which use curve fitting algorithms such as spline or polynomials [8], the types of distortion (smooth curl, cylindrical, fold, etc.) have to be known to be able to fit the best curve to the baseline. Given the fact that arbitrary baselines can take any form, it is thus impossible to model them with polynomials or splines. Moreover, spline and polynomial curve fitting or energy minimization techniques such as active contours and snake [4] fail to give precise baselines and there is always an inclination towards descenders in the above methods. In the second step of our evaluation methodology, based on the mathematical function of each line, a comprehensive metric which calculates the performance of a dewarping technique is introduced.

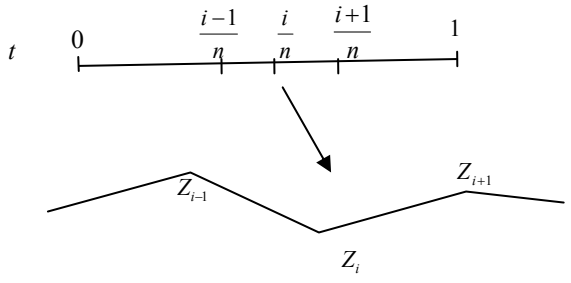
The proposed evaluation methodology has several advantages in comparison to existing evaluation methodologies. Firstly, it does not depend on OCR or the perfect original image such as the method in [5]. Secondly, it evaluates and examines all the lines in the image and not only a selected number of lines such as the method in [12]. Thirdly, it models all kinds of lines including curl and arbitrary warping and not only smooth curves in a document with known distortions. Fourthly, it is the first evaluation methodology which works successfully on historical document images and it is completely automatic.

The remainder of this paper is structured as follows: in Section 2, the proposed technique is described in detail. In Section 3, experimental results are presented and discussed. Finally, in Section 4, conclusions are drawn.

2. Performance Evaluation Methodology

The main purpose of a dewarping method is to have consistent, horizontal, straight lines in the final dewarped image. Therefore any deviation from the straight line in the dewarped image is considered an error and reflects some degree of failure of the dewarping technique. To define a metric for the automatic evaluation and to measure the deviation of each line from the straight line, it is necessary to detect precise baselines.

The precise baselines were detected using our proposed text line methodology in [9]. The integral I of the final parametric equation of each line after having been updated with new coordinate points when t varies between 0 and 1 indicates the area under the curve:



$$I = \int_0^1 \left| (1 - \{nt\}) Z_{[nt]} + \{nt\} Z_{[nt+1]} \right| dt \qquad (1)$$

where n is the number of partitions in the line, [] represents the *floor* function, and {} represents the *fractional part* function. The area under the curve which is a potential metric for evaluating a dewarping technique indicates the deviation of each line from the horizontal straight line. The closer the line is to the horizontal straight line, the less is the value of the integral. According to this approach, the integral or the area under a horizontal straight line is 0.

The integral of a curve calculates the area between the curve and the horizontal axis (x-axis). In the image, the origin of all coordinate points is calculated from the top-left corner of the image. Obviously the area between each text line and the upper boundary of the image is not a good metric for dewarping evaluation because with this approach, even the integral of a straight horizontal line is not zero but rather is the area enclosed by the line and the upper boundary. Therefore, for each text line a different coordinate system would have to be defined. The y-axis of the coordinate system would remain the same for all text lines, which is the left border of the image. However the origin and x-axis would change for each text line.

To determine the best level, where the deviation of the curve from the straight line is minimized, different levels including *minimum, maximum, average and modal level* were examined. The minimum level means the horizontal line equal to the minimum value of points in each curve; we can have something similar for the maximum and average levels. In some cases, the area between the curve and the *minimum level* can produce extra shift and therefore extra error. For example in Figure 1a, there is a minimum point on the curve which is the valley of the curve; therefore the integral of the curve in regard to the *minimum level* produces error at both sides of the valley. The average height of the curve can also create extra shift and extra error for the same reason.

The level which is finally selected as the x-axis for calculating the integral is the level where the most points of the curve are located; the rationale behind this choice is that in this case the vertical distance between the most points on the curve and the horizontal level would be zero and thus the error would be minimal. This level is the level on which the curve sits, or in other words it is the most horizontal part of the curve. We call this level the *modal level*.

Figure 1b shows the case where the integral of the curve is calculated in regard to the *modal level*. It can be seen both from the Figure and as confirmed by numerical values that in this case the integral of the curve will decrease and it does not include any constant shift in the integral.

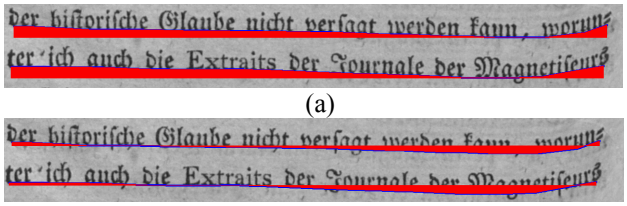


Figure 1. Integral of the baseline curve in regard to (a) *minimum level* and (b) *modal level*

Although, for integral computation, the modal level is the best one which gives the minimum error, the relation between *maximum level*, *minimum level* and *modal level* can be used for interpreting the shape of a curve. For instance, when the integral in regard to the minimum *level* is equal to that of the *modal level*, the curve has a concave shape; on the other hand when the integral in regard to the *maximum level* is equal to that of the modal level, the curve has a convex shape. When the *modal level* integral is between the *minimum level* and *maximum level* integral, the baseline has a

wavy shape. The same interpretation is applicable to the integral value of the function and its absolute value. These measurements can help in determining the type of distortion which exists in each line and whether a dewarping technique has removed the whole distortion, removed a part of it, or has created a new type of distortion.

Two metrics can be defined for dewarping evaluation. The first metric indicates the deviation of each line from the horizontal straight line. Thus each line can have a metric indicating its straightness. Error metric EM_j indicates the deviation of each line from the straight line and is defined independently for each line in the image as follows:

$$EM_j = L_j^{-1} \int_0^1 \left| (1-\{nt\}) Z^{\text{mod}}_{[nt]} + \{nt\} Z^{\text{mod}}_{[nt+1]} \right| dt \qquad (2)$$

In this equation, Z^{mod} are the updated coordinate points calculated in regard to the modal level and L_j is the length of the j^{th} line. Dividing the integral by its length normalized the metric. This is especially necessary for images which do not have text lines with the same length. The lower the value of the EM_j, the less is the deviation of the line from the straight line. This metric can even be defined for the original image, showing how far it is from the straight line.

The second evaluation metric is the accuracy metric (AM) and is defined according to the following equation.It shows the accuracy of each line AM_j or the overall accuracy AM in a dewarped image.

$$AM_j = 100\left(1 - \frac{EM_j}{EM_j^0}\right)$$
$$AM = \frac{1}{N}\sum_{j=1}^{N} AM_j \qquad (3)$$

In this equation, EM_j, is the deviation of line j from the straight line, EM^0_j is the corresponding metric in the original image and N is the number of lines in the image. The higher the value of AM_j, the more superior is the accuracy of the dewarped line. A complete horizontal straight line has the accuracy of 100% while a line which has the same distortion as the original image has the accuracy of 0%.

3. Experimental Results

The proposed evaluation methodology was examined on a dataset of 30 representative historical document images with geometric distortions using two state of the art dewarping methods for Arbitrary Warped Historical document images (AWH) [9] and for Camera Based Document images (CBD) [13] and the Book Restorer commercially available software [1]. The first method [9] uses a novel segmentation and precise baseline model to detect the exact shape of distortion and then using a flow line approach to rectify the image. The second method [13] uses a two step approach, namely coarse and fine dewarping to rectify the image.

Figure 2 shows a sample image from the dataset with an arbitrary warping. One of the advantages of the presented dewarping evaluation method is that the error metric--which measures the deviation of each line from the horizontal line--can be calculated for each line in every image and even in the original image. In this way, the most distorted lines both in the original distorted image and in the dewarped images can be detected. Figure 3 shows the error metric (EM_j) for each line and each image. As can be observed from figure 3, for most of the lines the dewarped methods show less error than the original distorted image. But for some lines the error measure is equal to the distorted image or even is worse. For example, figure 3 shows that the CBD [13] method and Book-restorer [1] in lines 22, 23, 24 have more distortion than the original image. Figure 4 shows the qualitative measurements in lines 22-24 for all four images in figure 2. The red regions show the deviation of each line from the horizontal line. The qualitative results in figure 4 confirm the quantitative measurements in figure 3.

The presented dewarping evaluation technique measures the error in each single line and not only in some represetative lines, such as the evaluation method in [12]. Our evaluation technique thus gives more accurate and comprehensive evaluation metrics while also detecting the all the errors in the whole page and not only on selected lines. The accuracy metric (AM) for the above image for the AWH method [9] is 92%%. The Book-Restorer [1] and CBD method [13] reach the average accuracy of 41% and 25%, respectively. Table 1 illustrates the average AM of all dewarping techniques on all images in the dataset.

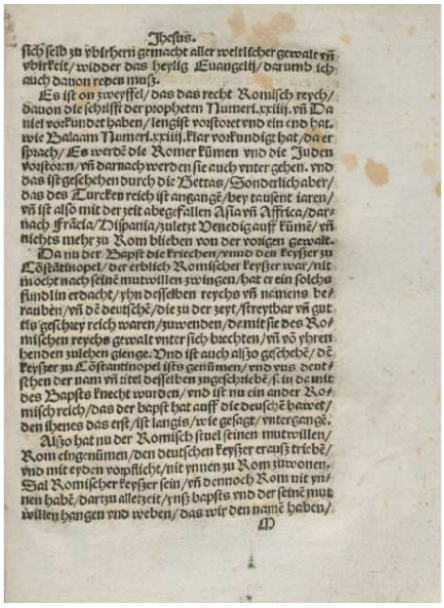


Figure 2. Image with Arbitrary Warping (Copyright: Bavarian State Library)

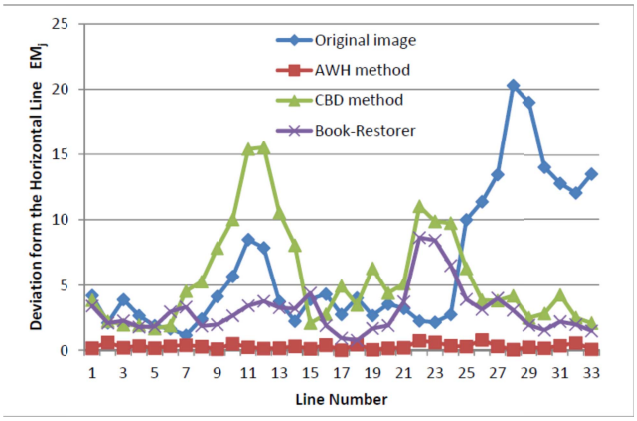
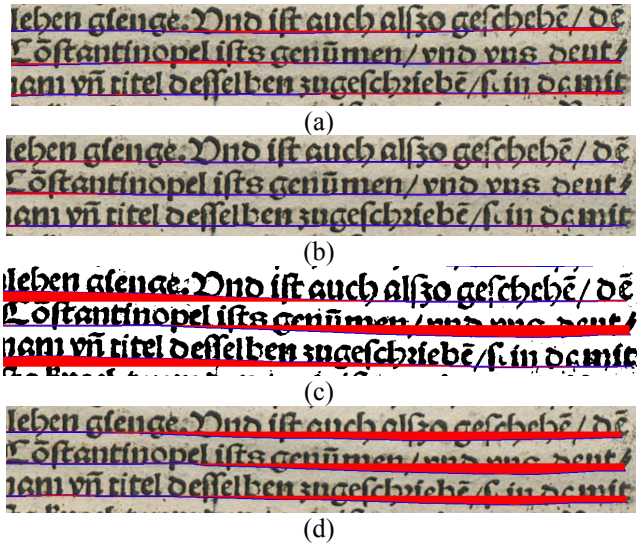


Figure 3. The error metric for each line in original image and dewarped images: AWH [9], CBD [13], Book Restorer [1]



(a) (b) (c) (d)

Figure 4. Deviation from the horizontal line for lines 22-24 in figure 2 for (a) original image (b) AWH method [9] (c) CBD method [13] (d) Book-Restorer[1]

Table 1. Comparative results using the proposed evaluation methodology

Dewarping Technique	Accuracy Metric (*AM*)
BookRestorer [1]	66.57%
CBD dewarping method [13]	57.61%
AWH dewarping method method [9]	**94.33%**

4. Conclusion

In this paper a flexible, unsupervised and objective method for the evaluation of document image dewarping is presented. First, the precise baselines are modelled with a robust mathematical formula both on the original image and on the dewarped one. Then based on the mathematical function of each line, a comprehensive metric which calculates the performance of a dewarping technique is introduced. Supervised evaluation methods need a perfect image to evaluate the output image based on the ground-truth one, while unsupervised or fully automatic methods can evaluate the performance independent of the perfect image, based on the simple idea that the final text lines should be straight. Although the process of ground-truthing can produce the possible perfect image, it is a very time-consuming procedure and it can take several hours to make a ground-truth, while for the proposed method the evaluation is quite fast and using the precise baseline detection and a comprehensive metric it gives exact measurement. Another advantage of the proposed method is that it is independent of OCR, which often is not available for historical documents. Experimental results, applied to two leading dewarping methods and an industry-standard commercial system, demonstrate the effectiveness and superiority of the proposed dewarping evaluation method.

5. References

[1] Book-Restorer image restoration software. In *http://www.i2s-bookscanner.com*.

[2] Brown, M., Sun, M., Yang, R., Yun, L.and Seales, W. 2007. Restoring 2D Content from Distorted Documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*, 1904-1916.

[3] Brown, M.and Tsoi, Y. 2006. Geometric and shading correction for images of printed materials using boundary. *IEEE Transactions on Image Processing 15*, 1544-1554.

[4] Bukhari, S., Shafait, F.and Breuel, T. 2009. Dewarping of Document Images using Coupled-Snakes. In *Proceedings of Third International Workshop on Camera-Based Document Analysis and Recognition*, Barcelona, Spain.

[5] Bukhari, S., Shafait, F.and Breuel, T. 2011. An image based performance evaluation method for page dewarping algorithms using SIFT features. In *Camera-Based Document Analysis and Recognition*, 138-149.

[6] Cao, H., Ding, X.and Liu, C. 2003. Rectifying the Bound Document Image Captured by the Camera: A Model Based Approach. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR)*.

[7] Gatos, B., Pratikakis, I.and Ntirogiannis, K. 2007. Segmentation based recovery of arbitrarily warped document images. In *International Conference on Document Analysis and Recognition (ICDAR)*, 989-993.

[8] Lu, S.and Tan, C.L. 2006. Document flattening through grid modeling and regularization. In *Proceedings of the Internationa Conference on Pattern Recognition*2006, 971–974.

[9] Rahnemoonfar, M.and Antonacopoulos, A. 2011. Restoration of Arbitrarily Warped Historical Document Images Using Flow Lines. In *Proceedings of the Eleventh International Conference on Document Analysis and Recognition (ICDAR)*, 905-909.

[10] Shafait, F.and Breuel, T.M. 2007. Document image dewarping contest. In *Proceedings of the 2nd Int. Workshop on Camera-Based Document Analysis and Recognition*2007, 181–188.

[11] Stamatopoulos, N., Gatos, B.and Pratikakis, I. 2009. A Methodology for Document Image Dewarping Techniques Performance Evaluation. In *ICDAR* 956-960.

[12] Stamatopoulos, N., Gatos, B.and Pratikakis, I. 2012. Performance evaluation methodology for document image dewarping techniques. *IET Image Processing*.

[13] Stamatopoulos, N., Gatos, B., Pratikakis, I.and Perantonis, S.J. 2011. Goal-oriented rectification of camera-based document images. *IEEE Transactions on Image Processing 20*, 910-920.

[14] Zhang, Y., Liu, C., Ding, X.and Zou, Y. 2008. Arbitrary Warped Document Image Restoration Based on Segmentation and Thin-Plate Splines. In *ICPR 2008*, 1-4.