Document Image Dewarping Contest

Faisal Shafait
German Research Center for Artificial Intelligence (DFKI),
Kaiserslautern, Germany
faisal.shafait@dfki.de

Thomas M. Breuel
Department of Computer Science
Technical University of Kaiserslautern, Germany
tmb@informatik.uni-kl.de

Abstract

Dewarping of documents captured with hand-held cameras in an uncontrolled environment has triggered a lot of interest in the scientific community over the last few years and many approaches have been proposed. However, there has been no comparative evaluation of different dewarping techniques so far. In an attempt to fill this gap, we have organized a page dewarping contest along with CBDAR 2007. We have created a dataset of 102 documents captured with a hand-held camera and have made it freely available online. We have prepared text-line, text-zone, and ASCII text ground-truth for the documents in this dataset. Three groups participated in the contest with their methods. In this paper we present an overview of the approaches that the participants used, the evaluation measure, and the dataset used in the contest. We report the performance of all participating methods. The evaluation shows that none of the participating methods was statistically significantly better than any other participating method.

1 Introduction

Research on document analysis and recognition has traditionally been focused on analyzing scanned documents. Many novel approaches have been proposed over the years for performing page segmentation [1] and optical character recognition (OCR) [2] on scanned documents. With the advent of digital cameras, the traditional way of capturing documents is changing from flat-bed scans to capture by hand-held cameras [3, 4]. Recognition of documents captured with hand-held cameras poses many additional technical challenges like perspective distortion, non-planar surfaces, low resolution, uneven lighting, and wide-angle-lens distortions [5]. One of the main research directions in camera-captured document analysis is to deal with the page curl and perspective distortions. Current document analysis and optical character recognition systems do not expect these types of artifacts, and show poor performance when applied directly to camera-captured documents. The goal of page dewarping is to flatten a camera captured document such that it becomes readable by current OCR systems.

Over the last decade, many different approaches have been proposed for document image dewarping [5]. These approaches can be grouped into two broad categories according to the acquisition of images:

1. 3-D shape reconstruction of the page using specialized hardware like stereo-cameras [6, 7], structured light sources [8], or laser scanners [9].

2. reconstruction of the page using a single camera in an uncontrolled environment [10, 11, 12]

The first approaches proposed in the literature for page dewarping were those based on 3-D shape reconstruction. One of the major drawbacks of the approaches requiring specialized hardware is that they limit the flexibility of capturing documents with cameras, which is one of the most important features of camera-based document capture. Therefore, the approaches based on a single camera in an uncontrolled environment have caught more attention recently. The approach in [12] claims to be the first dewarping approach for documents captured with hand-held cameras in an uncontrolled environment. It is interesting to note that the approaches in [10, 11, 12], which were all published in 2005, actually served as a trigger for research in analyzing documents captured with a hand-held camera and many other approaches like [13, 14, 15] have emerged in the following years. Despite the existence of so many approaches

for page dewarping, there is no comparative evaluation so far. One of the main problems is that the authors use their own datasets for evaluation of their approaches, and these datasets are not available to other researchers.

As a first step towards comparative evaluation of page dewarping techniques, we have organized a page dewarping contest along with CBDAR 2007. For this purpose we have developed a dataset of camera captured documents and have prepared ground-truth information for text-lines, text-zone, and ASCII text for all documents in the dataset (Section 2). Three groups participated in the contest. The dataset was given to the participants, and they were given a time frame of two weeks to return flattened document images, along with a brief summary of their methods. The description of the participating methods is given in Section 3. The documents returned by the participants were processed by an OCR system to compare and evaluate their performance. The results of the participating methods are discussed in Section 4 followed by a conclusion in Section 5.

2 DFKI-1 Warped Documents Dataset

To compare different dewarping approaches on a common dataset, we have prepared a ground-truthed database of camera captured documents. The dataset contains 102 binarized images of pages from several technical books captured by an off-the-shelf digital camera in a normal office environment. No specialized hardware or lighting was used. The captured documents were binarized using a local adaptive thresholding technique [11]. Some sample documents from the dataset are shown in Figure 8.

The following types of ground-truth are provided with the dataset:

1. ground-truth text-lines in color-coded format (Fig 1)
2. ground-truth zones in color-coded format (Fig 1)
3. ground-truth ASCII text in plain text format

Many approaches for dewarping use detection of curved text-lines as a first step [11, 15]. The purpose of providing text-line and text-zone level ground-truth is to assist the researchers in quantitatively measuring the performance of this important intermediate step. ASCII text ground-truth is intended for use as the overall performance measure of a dewarping system by using OCR on the dewarped document. The dataset is publicly available for download from *http://www.iupr.org/downloads/data*.

The dataset is not split into training and test set, because some algorithms need larger training sets as compared to others. It is expected that when other researchers use this dataset, they will split it into test and training sets as per requirements.

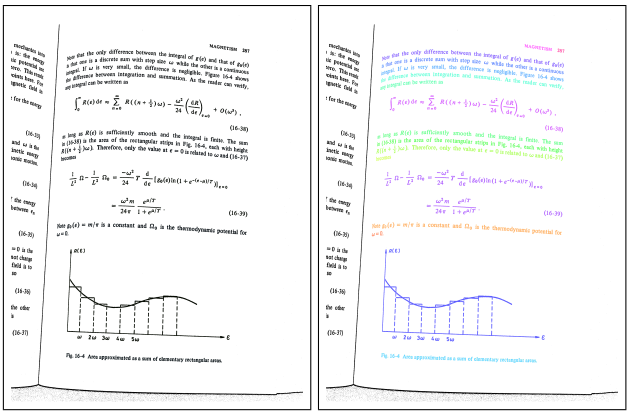


Figure 1. An example image (left) showing the ground-truth text-line and text-zone level information. The green channel contains the label of document zone type (text, graphics, math, table, image), the red channel contains the paragraph information for text-zones in reading order, and the blue channel contains the text-line information. For more information on pixel-accurate representation of page segmentation, please refer to [16]. The right image just replaces all the colors in the original ground-truth image with different visually distinguishable colors for visualization purposes.

3 Participating Methods

Three methods for document image dewarping were presented for participation in the contest by different research groups:

1. Continuous skeletal image representation for document image dewarping[1]
2. Segmentation based document image dewarping[2]
3. Coordinate transform model and document rectification for book dewarping[3]

[1] A. Masalovitch, L. Mestetskiy. Moscow State University, Moscow, Russia. anton_m@abbyy.com, l.mest@ru.net

[2] B. Gatos, N. Stamatopoulos, K. Ntirogiannis and I. Pratikakis. Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos", GR-153 10 Agia Paraskevi, Athens, Greece. http://www.iit.demokritos.gr/∼bgat/, {bgat,nstam,ipratika}@iit.demokritos.gr

[3] W. Li, B. Fu, M. Wu. Department of Computer Science and Technology, Peking University, Beijing 100871, China. {lwx,fubinpku,wuminghui}@pku.edu.cn

The text in the next sub-sections summarizes these methods and is based on the description of the methods provided by the participants.

3.1 Continuous skeletal image representation for document image dewarping (SKEL) [17]

This approach for image dewarping is based on the construction of outer skeletons of text images. The main idea of this algorithm is based on the fact that it is easy to mark up long continuous branches that define inter-linear spaces of the document in outer skeletons. Such branches can be approximated by cubic Bezier curves to find a specific deformation model of each inter-linear space of the document. On the basis of a set of such inter-linear space approximations, the whole approximation of the document is built in the form of a two-dimensional cubic Bezier patch. Then, the image can be dewarped using the obtained approximation of the image deformation.

3.1.1 Problem definition

Consider an image $I(x, y)$, where I is the color of the image pixel with coordinates (x, y). The goal of page dewarping is to develop a continuous vector function $D(x, y)$ to obtain a dewarped image in the form: $\overline{I}(x, y) = I(D_x(x, y), D_y(x, y))$. This function will be the approximation of the whole image deformation.

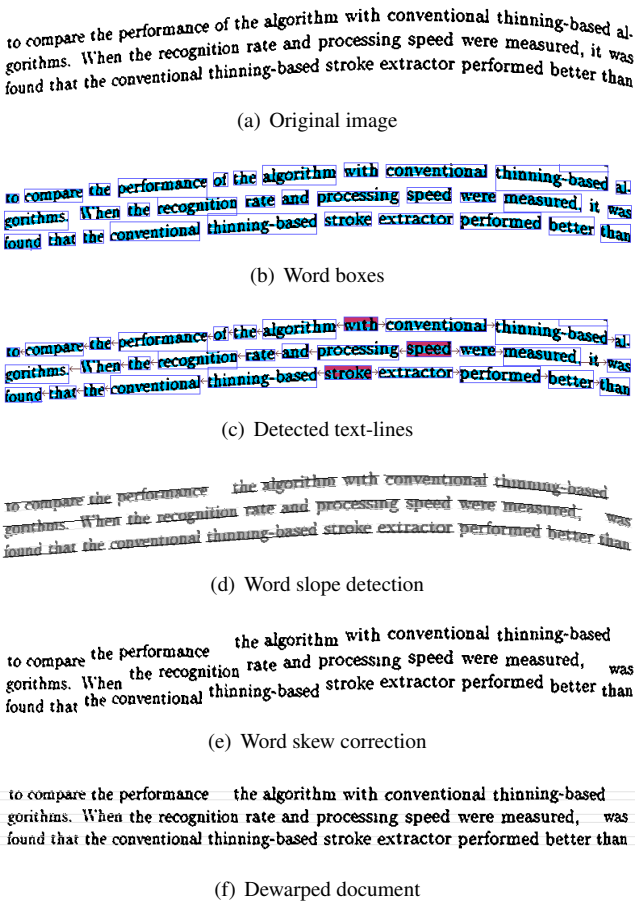
3.1.2 Main idea of the algorithm

The main idea of this algorithm is that in an outer skeleton of a text document image, one can easily find branches that lie between adjacent text-lines. Then, one can use this separation branches to approximate deformation of inter-linear spaces on the image. The proposed algorithm consists of the following steps:

1. A continuous skeletal representation of an image is built. The skeleton of an area is a set of points, such that for each point there exist no less than two nearest points on the border of the area. As border representation, polygons of minimal perimeter that enclose black objects on a picture are used. Methods exist that allow building of a continuous skeleton in time $O(n \log(n))$ [18].

2. The skeleton is filtered (useless bones are deleted).

3. All branches of the skeleton are clustered by their length and angle to find out horizontal and vertical branches.



(a) Original image

(b) Word boxes

(c) Detected text-lines

(d) Word slope detection

(e) Word skew correction

(f) Dewarped document

Figure 2. Example of the intermediate steps of page deskewing with the SEG approach.

4. The list of horizontal branches is filtered to leave only branches that lie between different text-lines.

5. A cubic Bezier approximation is built for each branch.

6. A two-dimensional Bezier patch is built that approximates all obtained curves. The patch is represented in the following form:

$$D(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} P_{ij} b_{i,3}(x) b_{j,3}(y) \qquad (1)$$

where $b_{r,3}(t)$ is a cubic Bernstein polynomial.

The patch thus obtained approximates the deformation function of the whole page.

3.2 Segmentation based document image dewarping (SEG) [19]

This technique enhances the quality of documents captured by a digital camera relying upon

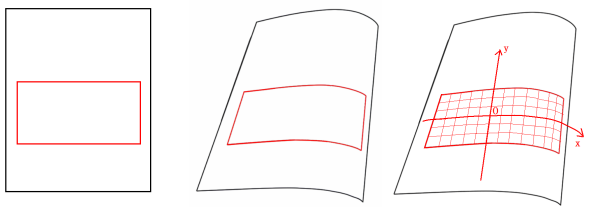


Figure 3. An example of image distortion of a flat area on a page when captured by a hand-held camera. The right-most image shows the curved coordinate net used in the CTM method.

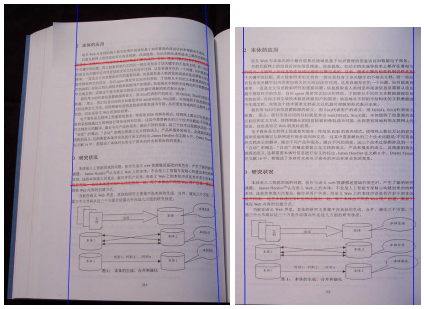


Figure 4. Illustration of document image before and after rectification with the CTM method.

1. automatically detecting and cutting out noisy black borders as well as noisy text regions appearing from neighboring pages

2. text-lines and words detection using a novel segmentation technique appropriate for warped documents

3. a first draft binary image dewarping based on word rotation and translation according to upper and lower word baselines

4. a recovery of the original warped image guided by the draft binary image dewarping result

In this approach, black border as well as neighboring page detection and removal is done followed by an efficient document image dewarping based on text-line and word segmentation [19]. The methodology for black border removal is mainly based on horizontal and vertical profiles. First, the image is smoothed, then the starting and ending offsets of borders and text regions are calculated. Black borders are removed by also using the connected components of the image. We detect noisy text regions appearing from neighboring page with the help of the signal cross-correlation function.

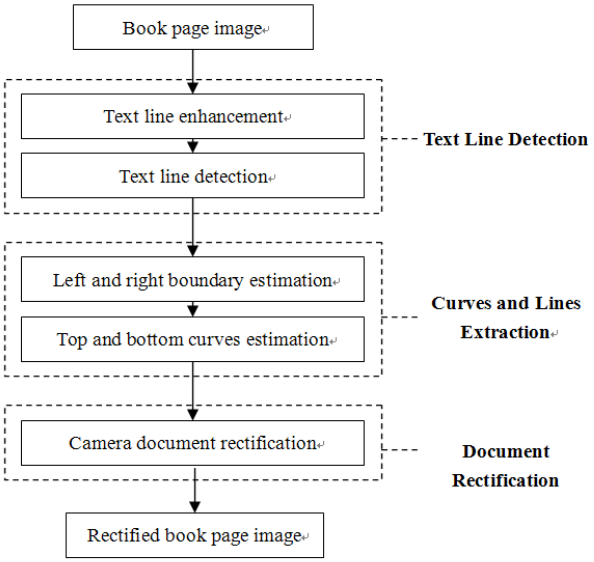


Figure 5. A flowchart of the CTM method.

At a next step, all words are detected using a proper image smoothing (Figure 2(b)). Then, horizontally neighboring words are consecutively linked in order to define text-lines. This is accomplished by consecutively extracting right and left neighboring words to the first word detected after top-down scanning (Figure 2(c)). For every detected word, the lower and upper baselines are calculated, which delimit the main body of the word, based on a linear regression which is applied on the set of points that are the upper or lower black pixels for each word image column [20]. The slope of each word is derived from the corresponding baselines slopes (Figure 2(d)). All detected words are then rotated and shifted (Figure 2(e)) in order to obtain a first draft estimation of the binary dewarped image. Finally, a complete restoration of the original warped image is done guided by the draft binary dewarping result of the previous stage. Since the transformation factors for every pixel in the draft binary dewarped image have been already stored, the reverse procedure is applied on the original image pixels in order to retrieve the final dewarped image. For all pixels for which transformation factors have not been allocated, the transformation factors of the nearest pixel are used.

3.3 Coordinate transform model and document rectification for book dewarping (CTM) [21]

This method uses a coordinate transform model and document rectification process for book dewarping. This model assumes that the book surface is a cylinder. It can handle both perspective distortion and book surface warping

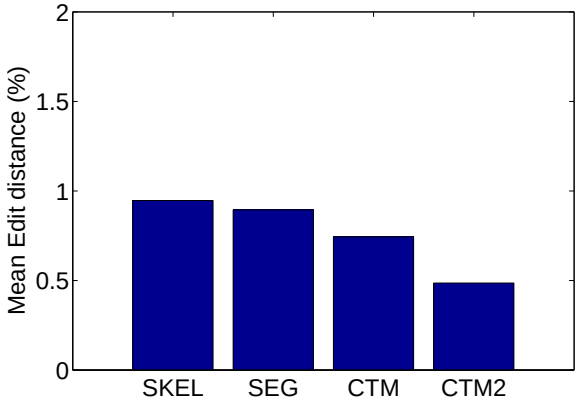


Figure 6. Mean edit distance of the text extracted by running Omnipage on the dewarped documents. Note that CTM2 just adds to CTM some post-processing steps to remove graphics and images from the dewarped documents.

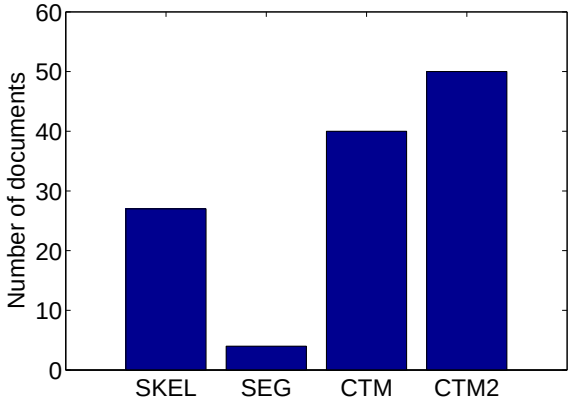


Figure 7. Number of documents for each algorithm on which it had the lowest edit distance among the participating methods.

problems. The goal is to generate a transformation to flatten the document image to its original shape (see Figure 3). The transformation is a mapping from the curved coordinate system to a Cartesian coordinate system. Once a curved coordinate net is set up on the distorted image as shown in Figure 3, the transformation can be done in two steps: First, the curved net is stretched to a straight one, and then adjusted to a well-proportioned square net.

According to the transform model, two line segments and two curves are needed to dewarp a cylinder image. Therefore, the left and right boundaries and top and bottom curves in book images are found for the rectification as shown in Figure 4.

The rectification process involves three steps: 1) the textline detection, 2) left and right boundary estimation and top and bottom curves extraction, and 3) document rectification. The flowchart of the rectification process is illustrated in Figure 5.

As an additional post-processing step, the participants used their programs to remove graphics and images from the processed pages. The results thus produced are referred to as **CTM2**.

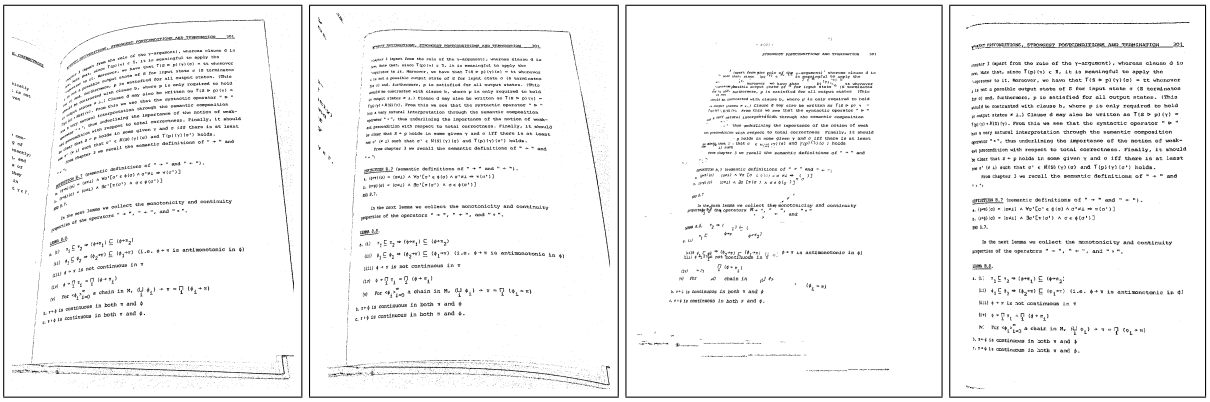
4 Experiments and Results

The results of the participating methods on some example documents from the dataset are shown in Figure 8. The dewarped documents returned by the participants were processed through Omnipage Pro 14.0, a commercial OCR system. After obtaining the text from the OCR software, the edit distance with the ASCII text ground-truth was used as the error measure. Although OCR accuracy is a good measure for the performance of dewarping on text regions, it does not measure how well the dewarping algorithm worked on the non-text parts, like math or graphics regions. Despite this limitation, we used the OCR accuracy because it is the most widely-used measure for measuring performance of dewarping systems [5].

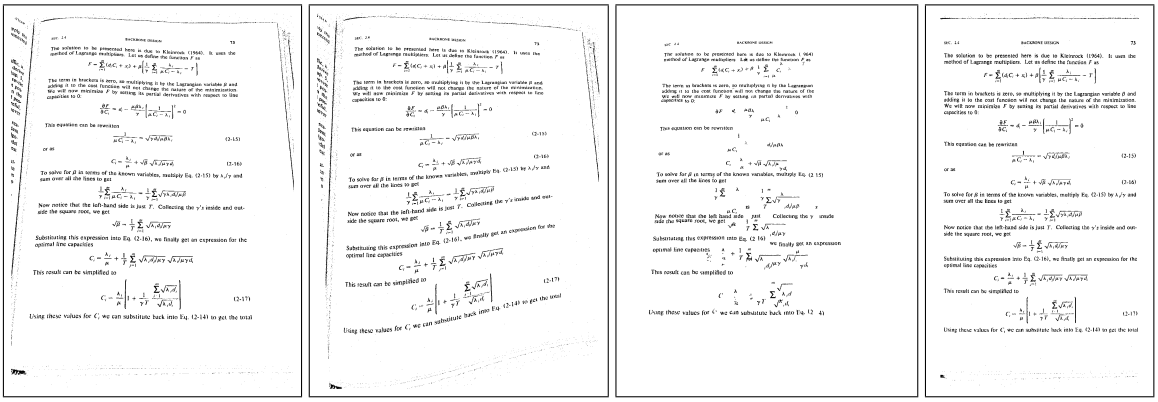
The mean edit distance of the participating methods is shown in Figure 6. The graph shows that the CTM technique performs best on the test data, and its results further improve after post-processing to remove graphics and images. This is because the ground-truth ASCII text contains text coming only from the textual parts of the documents, so the text that is present in graphics or images is ignored. Hence, the dewarped documents that contain text inside graphics regions get higher edit distances.

To analyze whether one algorithm is uniformly better than the other algorithms, we plotted the number of documents for each algorithm on which it had the lowest edit distance on character basis (Figure 7). If there was a tie between more than one methods for the lowest error rate on a particular document, all algorithms were scored for that document. Interestingly, the results show that the SEG method achieves the lowest error rate in only four documents. Here again the CTM2 method proves to be the best for the highest number of documents.

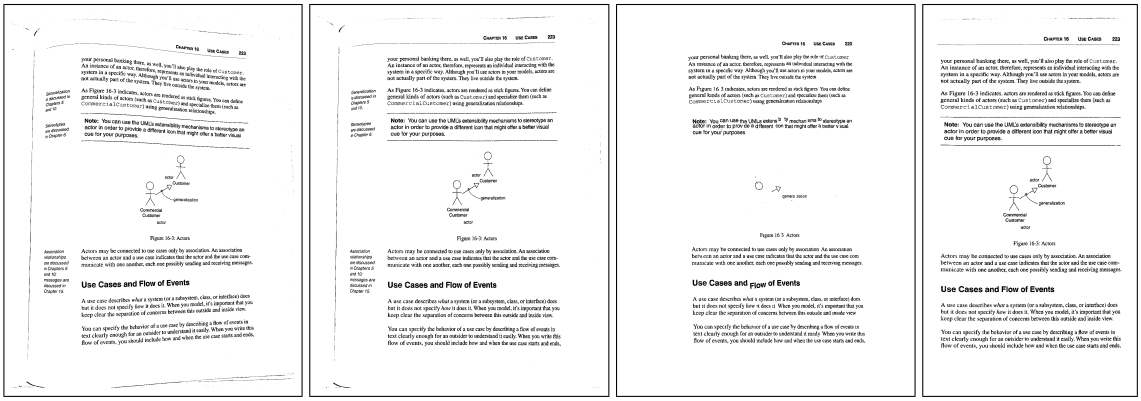
The analysis of the difference in the performance of the participating algorithms was done using a box plot (Figure 9). The boxes in the box plot represent the interquartile range, i.e. they contain the middle 50% of the data. The lower and upper edges represent the first and third quartiles, whereas the middle line represents the median of the data.



(a) Original Image (b) SKEL (c) SEG (d) CTM



(e) Original Image (f) SKEL (g) SEG (h) CTM



(i) Original Image (j) SKEL (k) SEG (l) CTM

Figure 8. Example results of the participants. For image 8(a), the SKEL and SEG methods remove page curl distortion, but could not handle perspective distortion. In image 8(e), the SKEL method was misled by the formulas and did not dewarp it correctly. In image 8(i), the SEG and CTM methods removed some text parts that were present near the left border of the page.

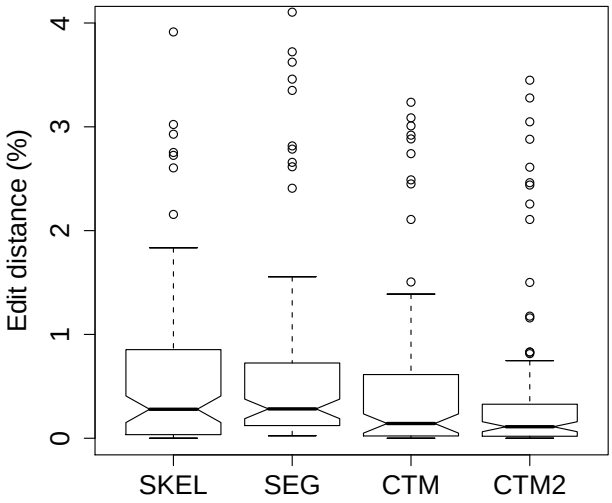


Figure 9. A box plot of the percentage edit distance for each algorithms. Overlapping notches of the boxes show that none of the participating algorithms is statistically significantly better than any other algorithms.

The notches represent the expected range of the median. The 'whiskers' on the two sides show inliers, i.e. points within 1.5 times the interquartile range. The outliers are represented by small circles outside the whiskers. Figure 9 shows that the expected range of medians of the edit distance overlaps for all the algorithms. Hence, it can be concluded that none of the participating algorithms is statistically significantly better than any other algorithm.

5 Conclusion

The purpose of the dewarping contest was to take a first step towards a comparative evaluation of dewarping techniques. Three groups participated in the competition with their methods. The results showed that the coordinate transform model (CTM) presented by Wenxin Li et al. performed better than the other two methods, but the difference was not statistically significant. Overall, all participating methods worked well and the mean edit distance was less than 1% for each of them. We have made the dataset used in the contest publicly available so that other researchers can use the dataset to evaluate their methods.

Acknowledgments

This work was partially funded by the BMBF (German Federal Ministry of Education and Research), project IPeT (01 IW D03).

References

[1] F. Shafait, D. Keysers, and T.M. Breuel. Performance comparison of six algorithms for page segmentation. In *7th IAPR Workshop on Document Analysis Systems*, pages 368–379, Nelson, New Zealand, Feb. 2006.

[2] S. Mori, C.Y. Suen, and K. Yamamoto. Historical review of OCR research and development. *Proceedings of the IEEE*, 80(7):1029–1058, 1992.

[3] M. J. Taylor, A. Zappala, W. M. Newman, and C. R. Dance. Documents through cameras. In *Image and Vision Computing 17*, volume 11, pages 831–844, September 1999.

[4] T.M. Breuel. The future of document imaging in the era of electronic documents. In *Int. Workshop on Document Analysis*, Kolkata, India, Mar. 2005.

[5] J. Liang, D. Doermann, and H. Li. Camera-based analysis of text and documents: a survey. *Int. Jour. of Document Analysis and Recognition*, 7(2-3):84–104, 2005.

[6] A. Ulges, C. Lampert, and T. M. Breuel. Document capture using stereo vision. In *Proceedings of the ACM Symposium on Document Engineering*, pages 198–200. ACM, 2004.

[7] A. Yamashita, A. Kawarago, T. Kaneko, and K.T. Miura. Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system. In *Proceedings of 17th International Conference on Pattern Recognition (ICPR2004), Vol.1*, pages 482–485, 2004.

[8] M.S. Brown and W.B. Seales. Document restoration using 3d shape: A general deskewing algorithm for arbitrarily warped documents. In *International Conference on Computer Vision (ICCV01)*, volume 2, pages 367–374, July 2001.

[9] M. Pilu. Deskewing perspectively distorted documents: An approach based on perceptual organization. *HP White Paper*, May 2001.

[10] L. Zhang and C.L. Tan. Warped image restoration with applications to digital libraries. In *Proc. Eighth Int. Conf. on Document Analysis and Recognition*, pages 192–196, Aug. 2005.

[11] A. Ulges, C.H. Lampert, and T.M. Breuel. Document image dewarping using robust estimation of curled text lines. In *Proc. Eighth Int. Conf. on Document Analysis and Recognition*, pages 1001–1005, Aug. 2005.

[12] J. Liang, D.F. DeMenthon, and D. Doermann. Flattening curved documents in images. In *Proc. Computer Vision and Pattern Recognition*, pages 338–345, June 2005.

[13] S. Lu and C.L. Tan. The restoration of camera documents through image segmentation. In *7th IAPR Workshop on Document Analysis Systems*, pages 484–495, Nelson, New Zealand, Feb. 2006.

[14] S. Lu and C.L. Tan. Document flattening through grid modeling and regularization. In *Proc. 18th Int. Conf. on Pattern Recognition*, pages 971–974, Aug. 2006.

[15] B. Gatos and K. Ntirogiannis. Restoration of arbitrarily warped document images based on text line and word detection. In *Fourth IASTED Int. Conf. on Signal Processing, Pattern Recognition, and Applications*, pages 203–208, Feb. 2007.

[16] F. Shafait, D. Keysers, and T.M. Breuel. Pixel-accurate representation and evaluation of page segmentation in document images. In *18th Int. Conf. on Pattern Recognition*, pages 872–875, Hong Kong, China, Aug. 2006.

[17] A. Masalovitch and L. Mestetskiy. Usage of continuous skeletal image representation for document images de-warping. In *2nd Int. Workshop on Camera-Based Document Analysis and Recognition*, Curitiba, Brazil, Sep. 2007. Accepted for publication.

[18] L.M. Mestetskiy. Skeleton of multiply connected polygonal figure. In *Proc. 15th Int. Conf. on Computer Graphics and Applications*, Novosibirsk, Russia, June 2005.

[19] B. Gatos, I. Pratikakis, and K. Ntirogiannis. Segmentation based recovery of arbitrarily warped document images. In *Proc. Int. Conf. on Document Analysis and Recognition*, Curitiba, Brazil, Sep. 2007. Accepted for publication.

[20] U.V. Marti and H. Bunke. Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system. *Int. Jour. of Pattern Recognition and Artifical Intelligence*, 15(1):65–90, 2001.

[21] B. Fu, M. Wu, R. Li, W. Li, and Z. Xu. A model-based book dewarping method using text line detection. In *2nd Int. Workshop on Camera-Based Document Analysis and Recognition*, Curitiba, Brazil, Sep. 2007. Accepted for publication.