Performance evaluation methodology for document image dewarping techniques

N. Stamatopoulos[1,2] B. Gatos[1] I. Pratikakis[3]

[1]National Centre of Scientific Research 'Demokritos', Institute of Informatics and Telecommunications, GR-15310 Athens, Greece
[2]Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece
[3]Department of Electrical and Computer Engineering, Democritus University of Thrace, GR-67100 Xanthi, Greece
E-mail: nstam@iit.demokritos.gr

Abstract: The performance evaluation of dewarping techniques is currently addressed by concentrating in visual pleasing impressions or by using optical character recognition (OCR) as a means for indirect evaluation. In this study, the authors present a performance evaluation methodology that calculates a comprehensive evaluation measure which reflects the entire performance of a dewarping technique in a concise quantitative manner. The proposed evaluation measure takes into account the deviation of the dewarped text lines from a horizontal straight reference which is considered to be the optimal result. This measure is expressed by the integral over the dewarped text line curves. To reduce the manual effort for identifying the text lines in the dewarped image, the authors propose a point-to-point matching procedure that finds the correspondence between the manually marked warped document image and the dewarping counterpart. This enables an evaluation for unlimited number of methodologies addressing a marking procedure which is applied only once. The validity of the proposed performance evaluation methodology is demonstrated by a concise experimental work that comprises four state-of-the-art dewarping techniques as well as the involvement of different users in the interactive part of the procedure.

1 Introduction

Document digitisation by a flatbed scanner or a digital camera often results in document images which suffer from distortions caused by warping and perspective distortions (see Fig. 1). These distortions not only reduce document readability but also affect the performance of subsequent processing and recognition as most of modern optical character recognition (OCR) systems are based on the assumption that the text lines in a document are straight and horizontal. Many recent approaches have addressed these problems and can be classified into two main categories based on three-dimensional (3D) document shape reconstruction [1–3] and two-dimensional (2D) document image processing [4–11]. Approaches of the former category obtain the 3D information of the document image using specialised hardware. The approaches in the latter category rely on 2D information from camera document images.

However, although various techniques have been proposed for document image dewarping, no standard performance evaluation methodology exists. In particular, most of the evaluations concentrate in visual pleasing impressions [9–11]. As a result, the performance of these techniques is based on perceptual, subjective and qualitative human vision evaluation, hence objective evaluations or quantitative comparisons among the different dewarping techniques cannot be obtained. Furthermore, the use of OCR as a means

for indirect evaluation is widely used in the evaluation of dewarping techniques by comparing the OCR performance on warped and dewarped document images [1–8]. However, in many cases, such as in handwritten or historical documents, a meaningful OCR is not always feasible.

In this paper, an evaluation measure is proposed that takes into account the deviation of the dewarped text lines from a horizontal straight reference which is considered to be the optimal result. This measure is expressed by the integral over the dewarped text line curves. To reduce the manual effort for identifying the text lines in the dewarped image, we propose a point-to-point matching procedure that finds the correspondence between the manually marked warped document image and the dewarping counterpart. This enables an evaluation for unlimited number of methodologies addressing a marking procedure which is applied only once.

This work is an extension of our previous work presented in [12]. Compared to our previous work, the proposed work presents a new, more efficient and more accurate method for marking the warped image. Furthermore, each selected text line is subdivided into multiple connected cubic polynomial curves instead of being approximated by only one so that higher accuracy is obtained. Furthermore, an additional evaluation measure is proposed that takes into account the degree of distortions for every particular text line. Finally, as the proposed evaluation methodology
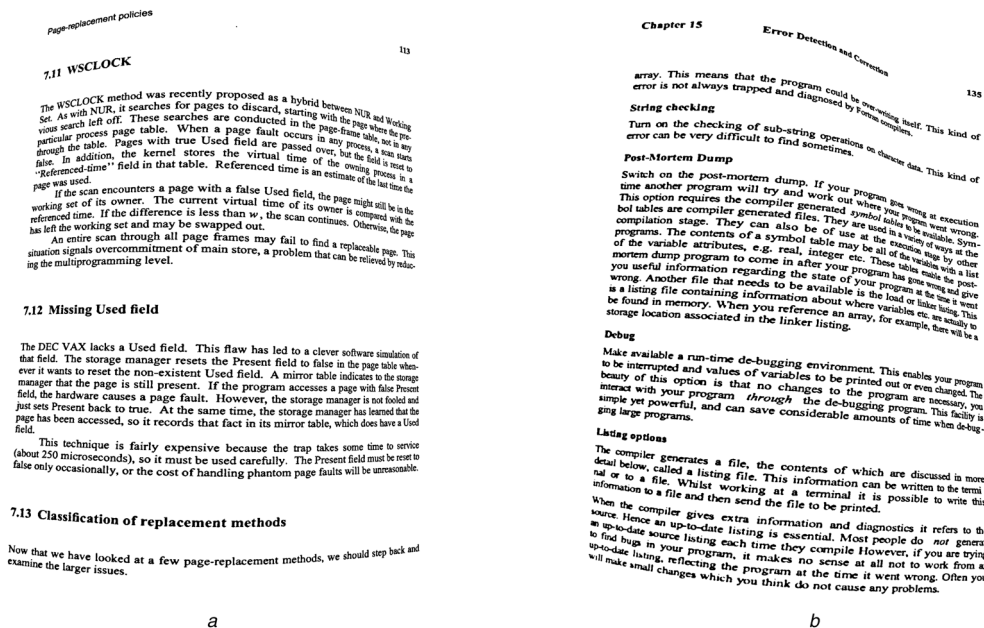
Fig. 1 *Examples of document images which suffer from distortions caused by warping and perspective distortions*

involves human intervention as well as the selection of a particular number of text lines, a dedicated experimentation to the robustness of those factors is presented.

The remainder of the paper is organised as follows. In Section 2, the proposed evaluation methodology is detailed whereas experimental results are discussed in Section 3. Finally, conclusions are drawn in Section 4.

2 Performance evaluation methodology

The flowchart of the proposed evaluation methodology is shown in Fig. 2. First, in the procedure of 'manual marking', which is the only stage that requires user's intervention, the user defines N set of points on the warped image which correspond to N selected text lines of the document image. Then, the procedure of 'point-to-point

matching' is implemented by using the scale-invariant feature transform (SIFT) [13]. The N set of points of the warped image are matched to the corresponding points of the dewarped image. Finally, at the 'evaluation' procedure, the cubic polynomial curves that fit to the selected text lines are estimated and the dewarping evaluation measure (DM) is calculated based on the integral of each curve. The performance evaluation of a dewarping technique considers that the expected optimal result should be constituted only from horizontal straight text lines. A detailed description of the distinct stages of the proposed evaluation methodology is described in the sequel.

2.1 Manual marking

In this stage, the user marks N set of points on the warped image that correspond to N selected text lines by defining a few points where the curvature of the selected text lines changes. It is not practical to select all, therefore text lines with representative deformation are selected. Fig. 3 depicts an example of a warped document image with six selected text lines $(N = 6)$ that correspond to representative deformation.

In order to mark a selected text line, the user must define only a few points where the curvature of the text line changes. It has to be pointed out that the user should define the points in the middle of the main body of the words (see Fig. 4a). Then, the straight line segments between two consecutive points are defined (see Fig. 4b) and a sampling of each line segment with a step size of five pixels is applied. Applying this procedure, we obtain the set of points that corresponds to the selected text line of the warped image (see Fig. 4c).

This is the only stage of the whole evaluation process that requires user's intervention. However, this stage should be done only once for each warped image. Then, we can evaluate unlimited number of dewarping techniques using the same marked image; as a result, we have a fair comparison among the different dewarping techniques. In

Fig. 2 *Flowchart of the proposed evaluation methodology*

1st selected text line

2nd selected text line

3rd selected text line
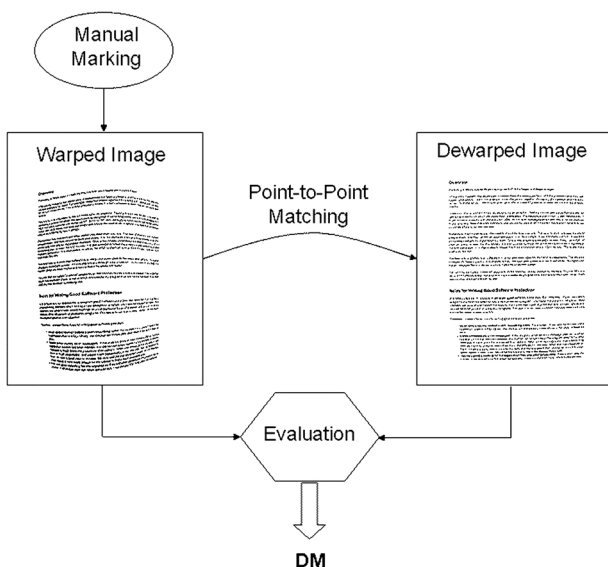
4th selected text line

5th selected text line

6th selected text line
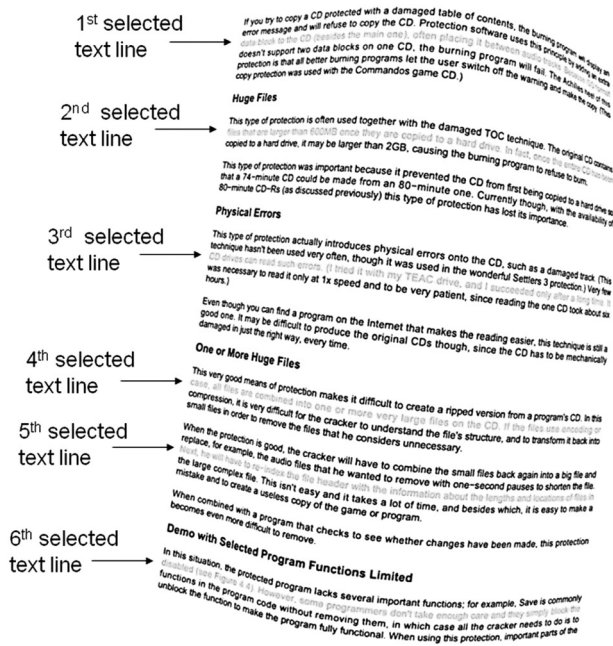
Fig. 3 *Example of a warped document image with six (N = 6) selected text lines that correspond to representative deformation*

order to examine the influence of the human intervention and the number of the selected text lines (N) in the proposed evaluation methodology, we present experimental results (Section 3) by using (i) different users to mark the same warped images and (ii) marked images with alternative values for the number of selected text lines.

2.2 Point-to-point matching

At this stage, the marked points of the warped image are to be matched with the corresponding points of the dewarped image using the SIFT [13].

The SIFT is used in computer vision to detect and describe local features in images. It has been used in a lot of applications as object recognition, robotic mapping and navigation, image stitching, 3D modelling, gesture recognition, video tracking and match moving. It transforms an image into a large collection of local feature vectors, each of which is

invariant to image translation, scaling and rotation, and partially invariant to illumination changes and affine or 3D projection. The scale-invariant features are efficiently identified by using a staged filtering approach. The first stage identifies key points by looking in scale space for maxima or minima at the difference-of-Gaussian image. Each point is used to generate a feature vector that describes the local image region. The resulting features vectors are called SIFT keys.

In our evaluation methodology, the SIFT keys from the warped and dewarped images are first extracted. Then, the matching SIFT key points between them are identified. Using this information, the marked points of the warped image are matched with the corresponding points of the dewarped image (see Fig. 5).This process is addressed as follows. Let $M(x_m, y_m)$ denote a marked point in the warped image (see Fig. 5). We find the two nearest SIFT key points $K_1(x_{k1}, y_{k1})$ and $K_2(x_{k2}, y_{k2})$ in the warped image using Euclidean distance. Then, using the matching SIFT key points $K'_1(x'_{k1}, y'_{k1})$ and $K'_2(x'_{k2}, y'_{k2})$, respectively, we define the corresponding point $M'(x'_m, y'_m)$ in the dewarped image with linear interpolation as follows

$$\left.\begin{aligned} x'_m = x_m * a_x + b_x \\ y'_m = y_m * a_y + b_y \end{aligned}\right\} \qquad (1)$$

where the coefficients a_x, b_x, a_y and b_y are calculated as follows

$$a_x = \begin{cases} \dfrac{x'_{k2} - x'_{k1}}{x_{k2} - x_{k1}} & \text{if } x_{k1} \neq x_{k2} \\ 1, & \text{otherwise} \end{cases}, \quad b_x = x'_{k1} - x_{k1} * a_x \quad (2)$$

$$a_y = \begin{cases} \dfrac{y'_{k2} - y'_{k1}}{y_{k2} - y_{k1}}, & \text{if } y_{k1} \neq y_{k2} \\ 1, & \text{otherwise} \end{cases}, \quad b_y = y'_{k1} - y_{k1} * a_y$$

$$(3)$$

Fig. 6 shows an example of this stage. In Section 3, experimental results on performance evaluation of point-to-point matching using SIFT transform are presented.
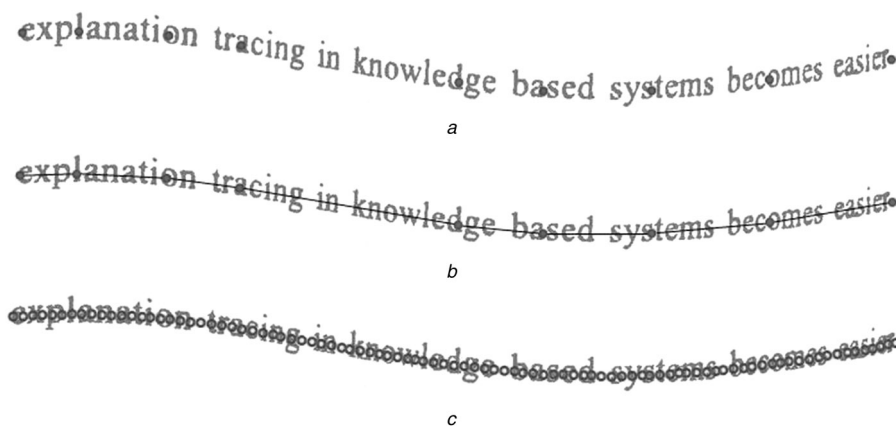
Fig. 4 *Example of a marked text line*

a User-defined points
b Straight line segments between consecutive points
c Final set of points of the text line
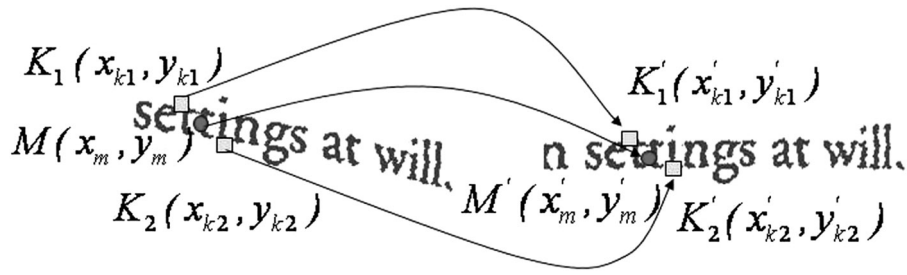
Fig. 5 *Point-to-point matching using SIFT transform*

$M(x_m, y_m)$ denotes a marked point in the warped image and $M'(x'_m, y'_m)$ its corresponding point in the dewarped image; $K_1(x_{k1}, y_{k1})$ and $K_2(x_{k2}, y_{k2})$ denote the two nearest SIFT key points in the warped image and $K'_1(x'_{k1}, y'_{k1})$, $K'_2(x'_{k2}, y'_{k2})$ denote the corresponding matching SIFT key points in the dewarped image
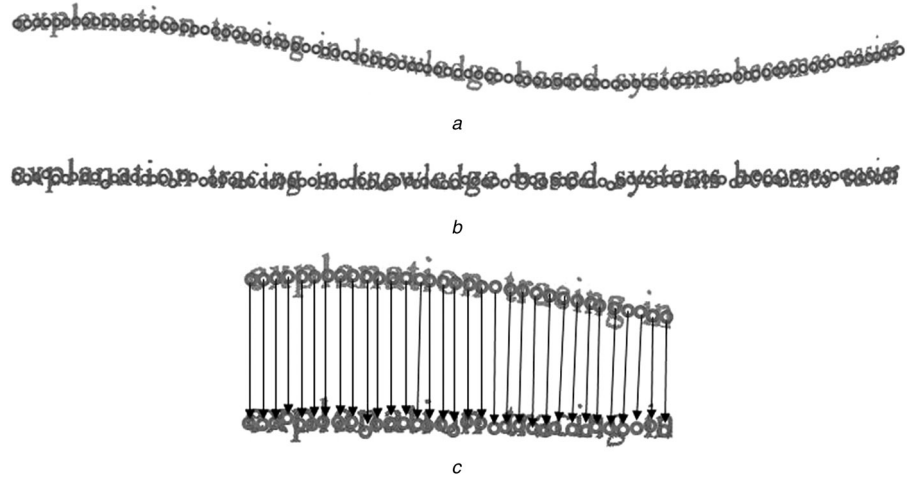
Fig. 6 *Example of point-to-point matching*

a Marked points in warped image
b Matching points in dewarped image using SIFT transform
c Enlarged image portion of *a* and *b* where matching points are indicated with arrows

2.3 Evaluation measures

Once the point-to-point matching has been applied, we proceed to the calculation of the DM that calculates the entire performance of a dewarping technique in a concise quantitative manner. It takes into account the deviation of the dewarped text lines from a horizontal straight reference which is considered to be the optimal result. This measure is expressed by the integral over the dewarped text line curves.

In each selected text line of the warped and dewarped image, multiple connected cubic polynomial curves are fitted using the N set of points. The fitted text line exhibits great accuracy, even if it suffers from various distortions, as a number of cubic polynomial curves are used. Next, the integral of each cubic polynomial curve will be used for the extraction of the DM. The detailed procedure in order to calculate the DM is as follows:

Step 1: Each set of the jth text line points ($1 \leq j \leq N$) in the warped and dewarped image is divided into k groups of points. Parameter k is equal with the number of the points that has been defined by the user (see Section 2.1). The groups are connected, so the last point of each group is the first point of the immediate neighbouring group.

Step 2: A least squares estimation method is used to find the coefficients of third degree polynomial curve that fit all points of each group. After this process, the cubic polynomial curves that fit the mth group of the warped and dewarped image, respectively, are defined as follows

$$^m y = {}^m a_3 x^3 + {}^m a_2 x^2 + {}^m a_1 x + {}^m a_0, \quad 1 \leq m \leq k \quad (4)$$

and

$$^m y' = {}^m a'_3 x^3 + {}^m a'_2 x^2 + {}^m a'_1 x + {}^m a'_0, \quad 1 \leq m \leq k \quad (5)$$

Step 3: Calculate the integral of the jth text line in the warped and dewarped image by adding the integrals of each cubic polynomial curve (see Fig. 7)

$$S_j = \sum_{m=1}^{k} \int_{{}^m x_s}^{{}^m x_e} ({}^m a_3 x^3 + {}^m a_2 x^2 + {}^m a_1 x)\, \mathrm{d}x \quad (6)$$

and

$$S'_j = \sum_{m=1}^{k} \int_{{}^m x'_s}^{{}^m x'_e} ({}^m a'_3 x^3 + {}^m a'_2 x^2 + {}^m a'_1 x)\, \mathrm{d}x \quad (7)$$

where ${}^m x_s$, ${}^m x_e$ and ${}^m x'_s$, ${}^m x'_e$ represent the start and end points of the mth group of the warped and dewarped image, respectively. It is obvious that the S'_j equals zero if the jth
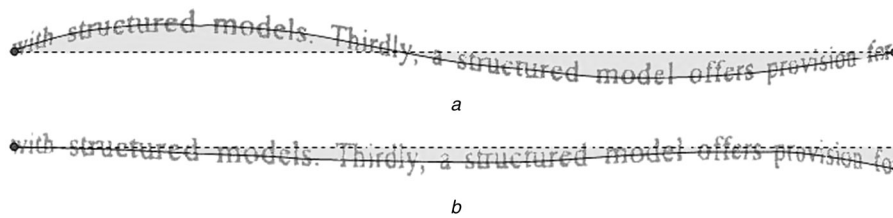
Fig. 7 *Integral regions of a text line in*

a Warped image
b Dewarped image

text line in the dewarped image is a horizontal straight text line that is considered to be the optimal result.

Step 4: Based on the estimated S_j and S_j', we define the DM_j dewarping measure that reflects the performance of the dewarping technique with respect to the jth text line as follows

$$DM_j = \begin{cases} 1 - \dfrac{S_j'}{S_j}, & \text{if } \dfrac{S_j'}{S_j} < 1 \\ 0, & \text{otherwise} \end{cases} \qquad (8)$$

As it can be observed, DM_j dewarping measure ranges in the interval $[0, \ldots, 1]$. It equals one ($DM_j = 1$) when the jth text line in the dewarped image is a horizontal straight text line that is the expected optimal result, so the integral of it equals zero ($S_j' = 0$). It indicates that the dewarping technique produces the best result. On the other hand, dewarping measure equals zero ($DM_j = 0$) when the dewarped image is the same or worse than the original warped image ($S_j' \geq S_j$), so the integral S_j is used as an upper limit for the jth text line.

Step 5: Repeat all the previous steps for the N selected text lines in order to extract an overall quantitative measure that considers the complete document image. We define the DM as the mean over all DM_j measures

$$DM = \frac{\sum_{j=1}^{N} DM_j}{N} \times 100\% \qquad (9)$$

DM ranges in the interval $[0, \ldots, 100]$ and the higher the value of the DM, the better is the performance of the dewarping technique.

DM dewarping measure depends on the N selected text lines of the document image without taking into account the degree of distortions for every particular text line. For this purpose, an additional evaluation measure (wDM) is proposed that gives more weight to text lines which suffer from larger distortions. The degree of distortions of each text line is expressed by its integral S_j (6). So, the wDM evaluation measure is defined as follows

$$wDM = \sum_{j=1}^{N} w_j DM_j \times 100\% \qquad (10)$$

where

$$w_j = \frac{S_j}{\sum_{l=1}^{N} S_l}, \quad 1 \leq j \leq N \qquad (11)$$

As it can be observed, the higher the value of S_j, the higher is

the weight w_j. In Section 3, we present experimental results using both DM and wDM.

3 Experimental results

The proposed evaluation methodology for document image dewarping techniques was examined on two different datasets. The first set (SET-1) consists of 100 warped document images in which OCR results are available [7]. The second set (SET-2) consists of 50 warped historical document images in which OCR results are not available. For both datasets, we applied three state-of-the-art dewarping techniques [6, 7, 9] as well as the dewarping method used in the commercial package BookRestorer [14]. The first method [6] (TSD-v.1) uses a two-step approach. At the first stage, a coarse dewarping is accomplished with the help of a transformation model. At second step, fine dewarping is obtained on the word level. The second method [7] (TSD-v.2) is an extension of the method TSD-v.1 [6] which incorporates a new method for the curved surface projection, the word baseline fitting as well as the restoration of horizontal alignment. It also rectifies the distortion of individual words using baseline estimation. Finally, the third method [9] segmentation based dewarping (SBD) uses a novel segmentation technique appropriate for warped documents and then all words are pose normalised guided by the lower and upper word baselines.

Fig. 8 shows the performance of all dewarping techniques using SET-1 in terms of the average DM measure, wDM measure and character accuracy. In order to calculate the DM and wDM measure we have manually marked six text lines ($N = 6$) in each document image. OCR is achieved using ABBYY FineReader Engine 8.1 [15]. It is worth mentioning that the overall comparative ranking is the same for all evaluation measures. According to these results, TSD-v.2 method [7] outperforms all the state-of-the-art
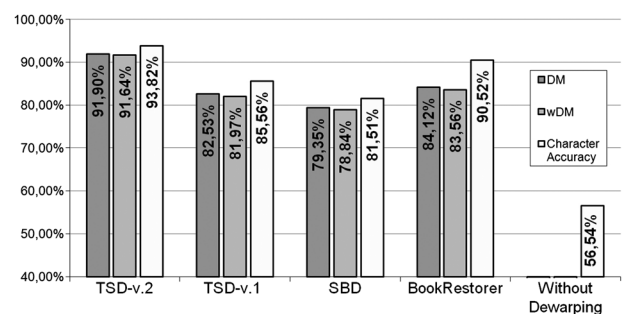
Fig. 8 *SET-1: comparative experimental results of all dewarping techniques using character accuracy, DM measure and wDM measure with six selected text lines (N = 6)*

Char. Acc.=44.38%

a

DM=83.11%
wDM=83.04%
Char.
Acc.=87.81%

b

DM=95.86%
wDM=95.22%
Char.
Acc.=96.02%

c

DM=79.47%
wDM=78.31%
Char.
Acc.=82.39%

d

DM=90.91%
wDM=89.74%
Char.
Acc.=93.82%

e

Fig. 9 *Recovery of a warped document image*

a Original document image; dewarped document image using
b TSD-v.1 method [6]
c TSD-v.2 method [7]
d SBD method [9]
e BookRestorer [14]
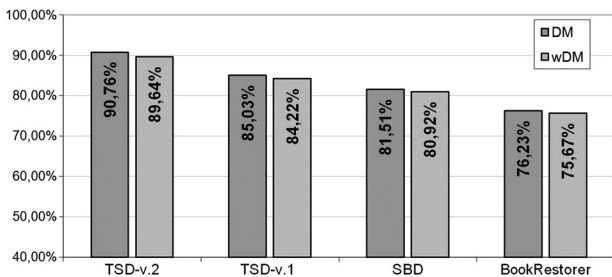
Fig. 10 *SET-2: comparative experimental results of all dewarping techniques using DM and wDM evaluation measures with six selected text lines (N = 6)*

dewarping methods, as well as the commercial package, which is verified also by Stamatopoulos *et al.* [7]. A representative result is shown in Fig. 9.

Fig. 10 illustrates the average DM and wDM measure using SET-2 in which OCR is not available. As in the previous experiment, we have manually marked six text lines ($N = 6$) in each document image. The experimental results are in agreement with the previous experiment in which SET-1 is used, as the TSD-v.2 method [7] outperforms all the other dewarping techniques. However, the commercial package BookRestorer does not perform well on this dataset, as it is not capable of dealing well with historical document images suffering from arbitrary warping.

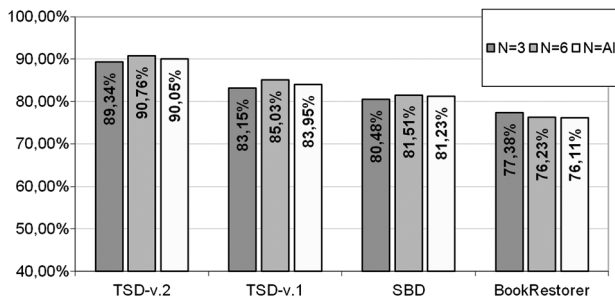**Fig. 11** *SET-2: comparative experimental results using DM dewarping measure by marking points in three text lines (N = 6), six text lines (N = 6) and all text lines of each document image (N = all_text_lines)*

In both previous experiments, we used six selected text lines ($N = 6$). In order to examine how the number of the selected text lines is able to influence the proposed evaluation methodology, an experiment has been carried out using three cases of interaction by marking points in three text lines ($N = 3$), six text lines ($N = 6$) as well as all text lines of each document image ($N = $ all_text_lines). For this purpose, we used the dataset SET-2. The evaluation results are shown in Fig. 11 using the DM dewarping measure. In all cases, the overall comparative ranking is the same with small variations in the absolute measurement. Thus, we can draw the conclusion that in order to obtain reliable results, we do not have to mark all the text lines of the document images, but we should select a few text lines that correspond to representative deformation.

The manual marking of the document image is the only stage of the whole evaluation process that requires user's intervention (see Section 2.1). We carried out two experiments using three different users in order to examine the influence of the human intervention in the proposed evaluation methodology. In the first experiment, each user selects and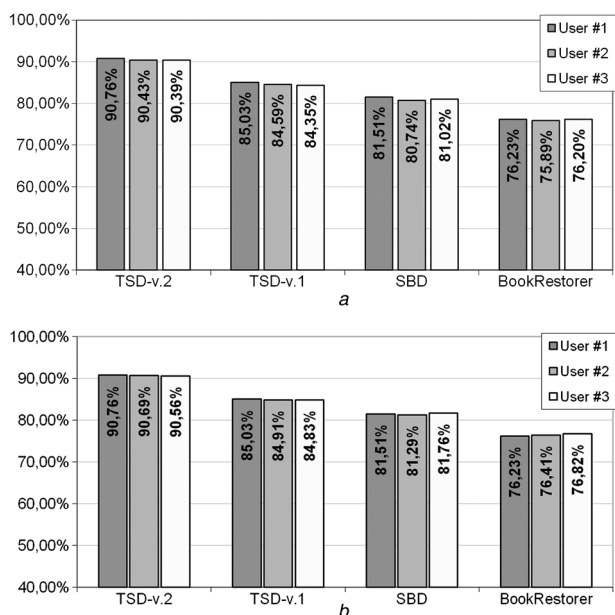 marks six text lines ($N = 6$) in all document images of the SET-2. Our aim is to examine how the selection of the text lines influences the evaluation process. In the second experiment, the users mark the same six representative text lines ($N = 6$) in order to examine how the marking process influences the dewarping measure. Fig. 12 illustrates the DM measure of both experiments. The results from both experiments indicate that the marking process does not influence the dewarping measure as we notice very small variations in the absolute measurement and the overall comparative ranking is the same.

Finally, we evaluate the point-to-point matching stage (see Section 2.2) by using ten warped images in which the correct matching points on their dewarped images have been manually marked. The two first stages of our methodology as described in Sections 2.1 and 2.2 are applied in which six representative text lines ($N = 6$) have been selected. Then, we calculate the Euclidean distance between the matching points produced by our methodology and correct marked points. The average error in terms of Euclidean distance in ten images is 1.41. This result demonstrates the success of point-to-point matching using SIFT transform as it denotes that the matching points suffer only from small variations which do not affect the proposed evaluation methodology.

4 Conclusion

A performance evaluation methodology that extracts a comprehensive evaluation measure which reflects the entire performance of a dewarping technique in a concise quantitative manner has been proposed. It avoids the dependence on an OCR engine which widely used as a means for indirect evaluation of dewarping techniques. Using the proposed methodology, we can obtain objective evaluations and quantitative comparisons among the different dewarping techniques. The manual marking of the warped document image is the only stage that requires user's intervention and according to the experimental results using different users it does not influence the evaluation measure. Furthermore, the proposed methodology was examined on two sets of warped document images using three state-of-the-art methodologies as well as a commercial package. The evaluation performed with the proposed evaluation measure is fully compatible with the evaluation achieved in terms of OCR accuracy.

5 References

1 Zhang, L., Zhang, Y., Tan, C.L.: 'An improved physically-based method for geometric restoration of distorted document images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, (4), pp. 728–734

2 Cao, H., Ding, X., Liu, C.: 'Rectifying the bound document image captured by the camera: a model based approach'. Seventh Int. Conf. on Document Analysis and Recognition, Scotland, 2003, pp. 71–75

3 Liang, J., DeMenthon, D., Doermann, D.: 'Geometric rectification of camera-captured document images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, (4), pp. 591–605

4 Ulges, A., Lampert, C.H., Breuel, T.M.: 'Document image dewarping using robust estimation of curled text lines'. Eighth Int. Conf. on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 1001–1005

5 Brown, M.S., Tsoi, Y.C.: 'Geometric and shading correction for images of printed materials using boundary', *IEEE Trans. Image Process.*, 2006, **15**, (6), pp. 1544–1554

6 Stamatopoulos, N., Gatos, B., Pratikakis, I., Perantonis, S.J.: 'A two-step dewarping of camera document images'. Eighth Int. Workshop on Document Analysis Systems, Nara, Japan, 2008, pp. 209–216

7 Stamatopoulos, N., Gatos, B., Pratikakis, I., Perantonis, S.J.: 'Goal-oriented rectification of camera-based document images', *IEEE Trans. Image Process.*, 2011, **20**, (4), pp. 910–920

Fig. 12 *SET-2: comparative experimental results using DM dewarping measure using three different users to*

a Select and mark six text lines
b Mark the same six text lines in each document image

8 Bukhari, S.S., Shafait, F., Breuel, T.M.: 'Dewarping of document images using coupled-snakes'. Int. Workshop on Camera-Based Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 34–41

9 Gatos, B., Pratikakis, I., Ntirogiannis, K.: 'Segmentation based recovery of arbitrarily warped document images'. Int. Conf. on Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 989–993

10 Ezaki, H., Uchida, S., Asano, A., Sakoe, H.: 'Dewarping of document image by global optimization'. Int. Conf. on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 500–506

11 Wu, C., Agam, G.: 'Document image de-warping for text/graphics recognition'. Joint IAPR Int. Workshop on Structural, Syntactic and Statistical Pattern Recognition, Windsor, Canada, 2002, pp. 348–357

12 Stamatopoulos, N., Gatos, B., Pratikakis, I.: 'A methodology for document image dewarping techniques performance evaluation'. Tenth Int. Conf. on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 956–960

13 Lowe, D.G.: 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, 2004, **60**, (2), pp. 91–110

14 BookRestorer: available at http://www.i2s-bookscanner.com/

15 ABBYY FineReader OCR: available at http://finereader.com/