2009 10th International Conference on Document Analysis and Recognition

A Methodology for Document Image Dewarping Techniques Performance Evaluation

N. Stamatopoulos, B. Gatos, I. Pratikakis
Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos", GR-153 10 Athens, Greece
{nstam, bgat, ipratika}@iit.demokritos.gr

Abstract

One of the major challenges in camera document analysis is to deal with the page curl and perspective distortions. In spite of the prevalence of dewarping techniques, no standard for their performance evaluation method exists with most of the evaluation done to concentrate in visual pleasing impressions. This paper presents an objective evaluation methodology for document image dewarping techniques. First, manually selected sets of points of the initial warped image are matched with the corresponding points of the dewarping result using the Scale Invariant Feature Transform (SIFT). Each set corresponds to a representative text line of the image. Then, based on cubic polynomial curves that fit to the selected text lines, a comprehensive measure which reflects the entire performance of a dewarping technique in a concise quantitative manner is calculated. Experiments applying the proposed performance evaluation methodology on two state of the art dewarping techniques as well as a commercial package are presented.

1. Introduction

All modern OCR systems are based on the assumption that the text lines in a document are straight and horizontal while this description does not actually hold. Document image acquisition by a digital camera often results into several image distortions. Non-linear warping is a major distortion that occurs especially when the scanned documents are bounded volumes. Text in such cases is strongly distorted and influences the performance of further processing.

Over the last decade, many different approaches have been proposed for document image dewarping [1]. These approaches can be classified into two broad categories based on (i) 3-D document shape reconstruction [2] and (ii) 2-D document image processing [3-7]. Approaches of the first category require specialized hardware or prior metric knowledge, so they limit the flexibility of capturing document with camera. On the other hand, approaches in the second category, which have caught more attention recently, use only 2-D information from camera document images.

In spite of the prevalence of dewarping techniques, no standard performance evaluation methodology exists with most of the evaluation done to concentrate in visual pleasing impressions [3-5]. As a result, the performance of these methods is based on perceptual, subjective and qualitative human vision evaluation; hence objective evaluations or quantitative comparisons among the different dewarping techniques can not be obtained. Furthermore, the use of OCR as a means for indirect evaluation is widely used in the evaluation of dewarping techniques [6-8]. However, in many cases, such as in handwritten or historical documents, OCR isn't always available or it doesn't produce satisfactory results.

In this paper, a novel methodology for performance evaluation of document image dewarping techniques is presented, that avoids the dependence on an OCR engine or human interference. It is based on a point-to-point matching procedure using the Scale Invariant Features Transform (SIFT) [9] as well as the use of cubic polynomial curves for the calculation of a comprehensive measure which reflects the entire performance of a dewarping technique in a concise quantitative manner. The remainder of the paper is organized as follows. In Section 2, the proposed evaluation methodology is detailed. Experiments applying this performance evaluation methodology on two state of the art dewarping techniques as well as a commercial package are presented in Section 3. Finally, conclusions are drawn in Section 4.

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.160

2. Evaluation Methodology

The performance evaluation of a dewarping technique considers that the expected result should be constituted only from horizontal straight text lines, without suffering from any distortions due to perspective and page warping. The proposed evaluation methodology is described in the flowchart of Fig. 1. First, we manually mark specific points on the warped image which correspond to N representative text lines of the original image. Then, using SIFT transform, the marked points of the warped image are matched to the corresponding points of the dewarped image. Finally, we estimate the cubic polynomial curves which fit to these points and based on the estimated curves we proceed to the extraction of the dewarping evaluation measure ($\boldsymbol{DW}$). The distinct stages of the proposed methodology are analyzed in the following of this section.

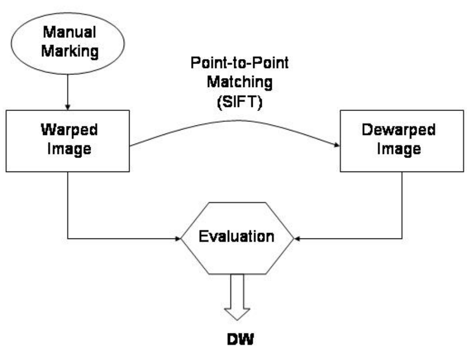


Figure 1. Flowchart of the proposed evaluation methodology.

2.1. Manual marking on the warped image

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

The user marks N set of points on the original warped image where each set corresponds to a representative text line of the image. The user should mark the points in the middle of the main body of the words. Also, the selected text lines should not be too small, should not be titles or subtitles so that they will be representative of the image. Since document deformation due to perspective and page warping is generally not uniform, we select N appropriate text lines of the document with representative deformation which must be corrected by a dewarping technique. Fig. 2 depicts an example with three selected text lines ($N = 3$).

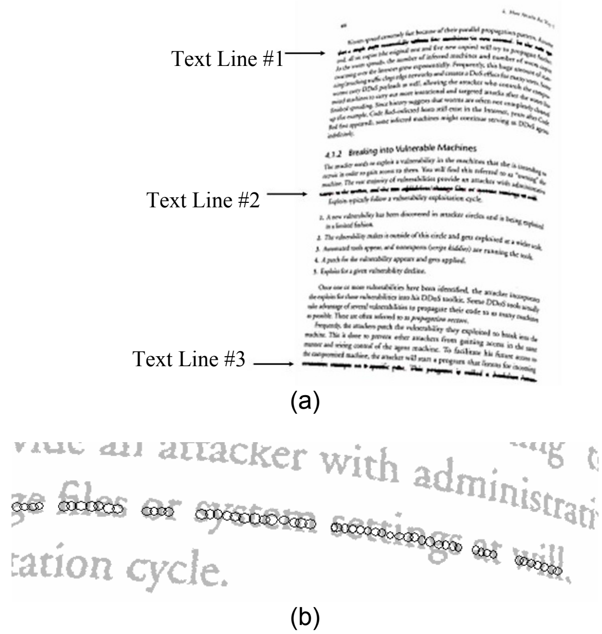


Figure 2. (a) An example of a marked warped image with three selected text lines (N=3); (b) Enlarged image portion of (a).

2.2. Point-to-Point Matching

At this stage, using the SIFT transform [9] the manually marked points of the original warped image are matched with the corresponding points of the dewarped image. SIFT transform 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 feature vectors are called SIFT keys.

In our methodology, we extract the SIFT keys from the warped and the corresponding dewarped image and then we identify matching key points between them. In this way, with the help of matching key points, the manually marked points of the warped image are matched with the corresponding points of the dewarped image.

For each marked point $M(x_m, y_m)$ in the warped image we find the two nearest key points $K_1(x_{k1}, y_{k1})$ and $K_2(x_{k2}, y_{k2})$ using Euclidean distance. Then, using the matching 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 (see Fig 3):

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

where a_x, b_x, a_y and b_y are denoted as follows:

$$a_x = \begin{cases} (x'_{k2} - x'_{k1})/(x_{k2} - x_{k1}) & if\ x_{k1} \neq x_{k2} \\ 1 & otherwise \end{cases} \qquad (2)$$

$$b_x = x'_{k1} - x_{k1} * a_x \qquad (3)$$

$$a_y = \begin{cases} (y'_{k2} - y'_{k1})/(y_{k2} - y_{k1}) & if\ y_{k1} \neq y_{k2} \\ 1 & otherwise \end{cases} \qquad (4)$$

$$b_y = y'_{k1} - y_{k1} * a_y \qquad (5)$$

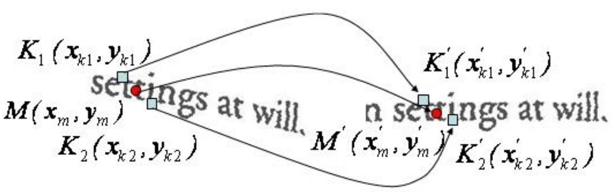
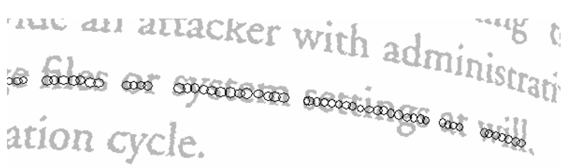
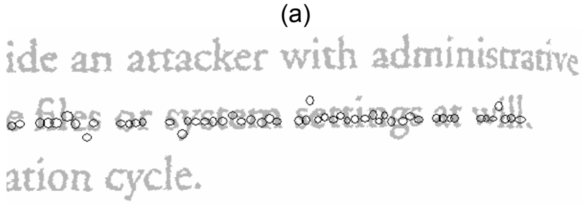


Figure 3. Point-to-Point matching using SIFT transform and linear interpolation.

Figure 4 depicts an example of point-to-point matching. As it can be observed, the marked points of the warped image are matched with the corresponding points of the dewarped image with great success. In Section 3, we present experimental results on performance evaluation of point-to-point matching using SIFT transform.



(a)



(b)

Figure 4. An example of point-to-point matching; (a) manually selected points in warped image; (b) corresponding points in dewarped image using SIFT transform.

2.3. Evaluation

This is the final stage of the proposed evaluation methodology in which we proceed to the computation of the dewarping evaluation measure ($\boldsymbol{DW}$) that reflects the entire performance of a dewarping technique in a concise quantitative manner.

Using the selected N set of points of the warped image (each set corresponds to points that belong to a representative text line of the image) and the matching points of the dewarped image, we approximate each text line of the document with cubic polynomial curves. The integral of each cubic polynomial curve, over an interval delimited by the curve endpoints, indicates the performance of the dewarping technique, as the expected result should be constituted only from horizontal straight text lines.

Let $({}^j x_i, {}^j y_i)$ represent the marked points of the j^{th} text line in the warped image (see Fig. 2), where ${}^j x_{min} \leq {}^j x_i \leq {}^j x_{max}$, $0 < i \leq L$ and $1 \leq j \leq N$. Similarly, $({}^j x'_i, {}^j y'_i)$ represent the matching points of the corresponding text line in the dewarped image, where ${}^j x'_{min} \leq {}^j x'_i \leq {}^j x'_{max}$. Also, let $LettH$ represent the average character height of the dewarped image which is calculated as in [10]. We calculate the average value ${}^j X'$ of all ${}^j x'_i$ and every point $({}^j x'_i, {}^j y'_i)$ is excluded if $|{}^j x'_i - {}^j X'| > LettH$ in order to eliminate possible errors of the point-to-point matching and therefore we will have a better estimation of the cubic polynomial curve. A Least Square Estimation (LSE) method is used to find the cubic polynomial curves that fit all points of both images. After this process the cubic polynomial curves that fit the j^{th} text lines are defined as follows:

$${}^j y = {}^j a_3 x^3 + {}^j a_2 x^2 + {}^j a_1 x + {}^j a_0 \qquad (6)$$

and

$${}^j y' = {}^j a'_3 x^3 + {}^j a'_2 x^2 + {}^j a'_1 x + {}^j a'_0 \qquad (7)$$

In this way, based on the estimated cubic polynomial curves, we define the DW_j measure which reflects the performance of the dewarping technique with respect to the j^{th} text line as follows:

$$DW_j = \begin{cases} 1 - \dfrac{Ar_j}{Ar'_j}, & if\ \dfrac{Ar_j}{Ar'_j} < 1 \\ 0, & otherwise \end{cases} \qquad (8)$$

where

$$Ar_j = \frac{\int_{^j x_{min}}^{^j x_{max}} {^j a_3} x^3 + {^j a_2} x^2 + {^j a_1} x}{^j x_{max} - {^j x_{min}}} \tag{9}$$

and

$$Ar_j' = \frac{\int_{^j x_{min}'}^{^j x_{max}'} {^j a_3'} x^3 + {^j a_2'} x^2 + {^j a_1'} x}{^j x_{max}' - {^j x_{min}'}} \tag{10}$$

As it can be observed $0 \le DW_j \le 1$. DW_j measure is equal to one when the j^{th} text line in the dewarped image is a horizontal straight text line that is the expected optimal result. It shows that the dewarping technique produces the best result. On the other hand, DW_j measure is equal to zero when the dewarped image is the same or worse than the original warped image. For an overall quantitative measure that considers the complete document image we define the $\boldsymbol{DW}$ as the mean value of all DW_j measures:

$$\boldsymbol{DW} = \frac{\sum_{j=1}^{N} DW_j}{N} \times 100\% \tag{11}$$

where $0\% \le \boldsymbol{DW} \le 100\%$. The higher the value of the $\boldsymbol{DW}$, the better is the performance of the dewarping technique.

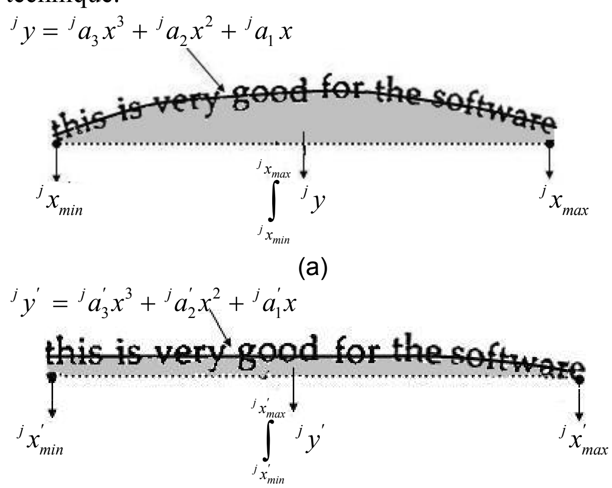


Figure 5. The integrals of cubic polynomial curves of (a) the warped image and (b) its corresponding dewarped image.

3. Experimental Results

Point-to-point matching is a crucial stage of the proposed evaluation methodology. For this reason we evaluate this stage using 10 warped images where we have manually marked the correct matching points on their dewarped images. Then, we apply the two first steps of our methodology as described in Sections 2.1 and 2.2. Because small variations don't affect the proposed evaluation methodology we consider a match as correct if the Euclidean distance between the matching point and correct marked point is smaller than a threshold. Table 1 illustrates the percentage of the marked points that match correct if Euclidean distance is 0 (perfect match), 1, 2, 3, and 4 respectively. The results of the evaluation demonstrate the success of point-to-point matching using SIFT transform.

Table 1. Evaluation of point-to-point matching

Euclidean Distance Threshold	Correct Match (%)
0	79,4
1	87,8
2	95,6
3	97,0
4	97,4

The proposed evaluation methodology for document image dewarping techniques was examined on a set of 20 warped images using two state of the art dewarping methods [3,6] as well as the commercial package BookRestorer [11]. The first method [3] uses a novel segmentation technique appropriate for warped documents and then all words are pose normalized guided by the lower and upper word baselines. The second method [6] 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 rectification is obtained on the word level.

In our experiments we used two cases of interaction by marking points in three text lines ($N = 3$) and six text lines ($N = 6$) with representative deformation instances. Table 2 illustrates the average $\boldsymbol{DW}$ of all dewarping techniques. For comparative reasons we have also included the dewarping results after applying only the coarse dewarping step of the method [6]. It is worth mentioning that the parameter N does not influence the comparative performance. According to these results method [6] had better results than method [3], which is verified also by work [6] where OCR accuracy was used. Moreover, as it was expected, method [6] had better performance than applying only the coarse dewarping step. Figure 6 shows a representative result.

Table 2. Comparative results using the proposed evaluation methodology

Dewarping Technique	***DW***	
	N = 3	*N* = 6
Using BookRestorer [11]	72%	73%
Using dewarping method [3]	78%	79%
Using dewarping method [6] applying only the coarse dewarping step	86%	86%
Using dewarping method [6] applying coarse and fine dewarping steps	**90%**	**89%**

4. Conclusions

This paper proposes an objective methodology for performance evaluation of document image dewarping techniques. Manually selected sets of points of the warped image are matched with the corresponding points of the dewarping result using the SIFT transform. Each set corresponds to a representative text line of the image. Then, based on cubic polynomial curves that fit to the selected text lines we extract a comprehensive measure which reflects the performance of a dewarping technique in a concise quantitative manner. The evaluation results of the proposed methodology are verified also by work [6] where OCR accuracy was used.

5. Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement n° 215064 (project IMPACT).

6. References

[1] J. Liang, D. Doermann, and H. Li. "Camera-based analysis of text and documents: a survey" *International Journal on Document Analysis and Recognition*, 7(2-3), 2005, pp. 84–104.

[2] C.L. Tan, L. Zhang, Z. Zhang and T. Xia, "Restoring Warped Document Images through 3D Shape Modeling", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(2), 2006, pp. 195-208.

[3] B. Gatos, I. Pratikakis and I. Ntirogiannis, "Segmentation Based Recovery of Arbitrarily Warped Document Images" *International Conference on Document Analysis and Recognition*, Curitiba, Brazil, 2007, pp. 989-993.

[4] H. Ezaki, S. Uchida, A. Asano & H. Sakoe, "Dewarping of document image by global optimization" *International Conference on Document Analysis and Recognition*, Seoul, Korea, 2005, pp. 500-506.

[5] Y. C. Tsoi and M. S. Brown, "Geometric and shading correction for images of printed materials–A unified approach using boundary" *Conference on Computer Vision and Pattern Recognition*, 2004, pp. 240-246

[6] N. Stamatopoulos, B. Gatos, I. Pratikakis and S. J. Perantonis, "A Two-Step Dewarping of Camera Document Images", 8th International Workshop on Document Analysis Systems, Nara, Japan, 2008, pp. 209-216.

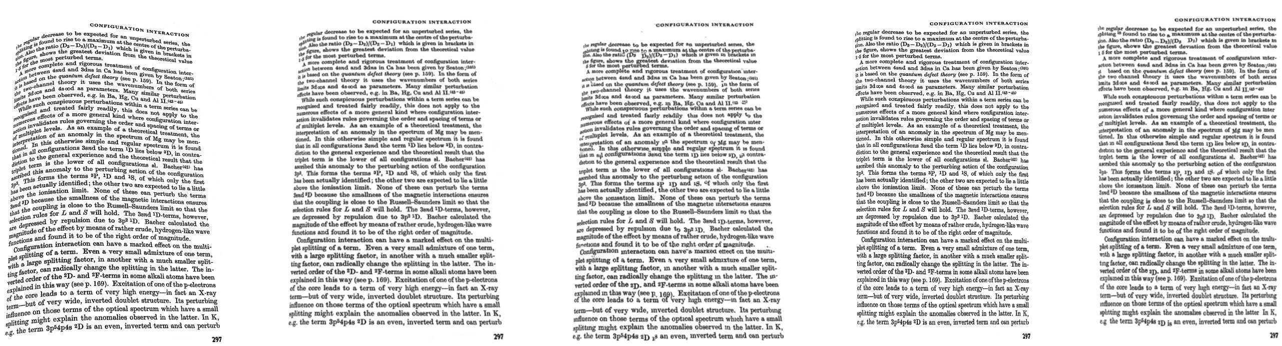
[7] J. Liang, D. DeMenthon, and D. Doermann, "Flattening curved documents in images" *Conference on Computer Vision and Pattern Recognition*, San Diego, USA, 2005, pp. 338–345.

[8] F. Shafait, T. M. Breuel, "Document Image Dewarping Contest", *In 2nd Int. Workshop on Camera-Based Document Analysis and Recognition*, Brazil, 2007, pp. 181-188.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints'" *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110.

[10] B. Gatos, T. Konidaris, K. Ntzios, I. Pratikakis and S. J. Perantonis, "A Segmentation-free Approach for Keyword Search in Historical Typewritten Documents", *International Conference on Document Analysis and Recognition*, Seoul, Korea, 2005, pp. 54-58.

[11] BookRestorer: http://www.i2s-bookscanner.com/



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

Figure 6. Recovery of a warped image: (a) original warped image; (b) using BookRestorer; (c) using method [3]; (d) using method [6] applying only the coarse dewarping step; (e) using method [6] applying coarse and fine dewarping steps.