Dewarping of document image by global optimization

Hironori Ezaki†, Seiichi Uchida†, Akira Asano‡, and Hiroaki Sakoe†
† Graduate School of Information Science and Electrical Engineering, Kyushu University,
6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi, 812-8581 Japan
‡ Faculty of Integrated Arts and Sciences, Hiroshima University,
1-7-1 Kagamiyama, Higashi-Hiroshima-shi, 739-8521 Japan

Abstract

This paper proposes a novel dewarping technique for document images of bound volumes. This technique is a kind of model fitting techniques for estimating the warp of each text line by fitting some elastic curve model to the text line. Differing from conventional techniques, the proposed technique is applicable to document images including local irregularities such as formulae, short text lines, and figures, since the proposed technique dewarps whole document images by fitting splines while considering the global optimality that specifies the desirable relationship among the splines. The experimental results on several document images including the local irregularities indicated the effectiveness of the proposed technique. The experimental result also indicated the effectiveness of the vertical division of a document image into some partial document images for more accurate dewarping.

1. Introduction

Nonlinear warps are often observed in document images of bound volumes, if those images are captured by a flatbed scanner. Similar warps will be observed if document images are captured by a digital camera. In order to recognize such document images by commercial OCR, dewarping should be performed in advanced to text line extraction because most text line extraction techniques employed in OCR assume that text lines are straight.

Dewarping techniques can be classified into the following three types: (i) stereo or other 3D measurements[1, 2], (ii) shape-from-shading[3], (iii) model fitting[4, 5, 6]. The third type is the technique which estimates the warp of each text line by fitting some elastic curve (or surface) model to the text line. This model fitting-based dewarping technique is the most widely applicable one among the above three types because it needs neither calibration, known parameters, nor special devices.

The model fitting technique proposed by Zhang et al.[4] divides a document image into a shaded area and a non-shaded area, and then fits quadratic functions to the text lines in the shaded area and fits linear functions to the text lines in the non-shaded area. Wu et al. [5] have proposed a dewarping technique where a nonlinear curve is fitted to each text line by tracking the character boxes on the text line.

In those conventional techniques, the fitting of an elastic curve to a text line is optimized *locally*. Thus, for reasonable results, it should be carefully considered that there might be erroneously fitted curves due to local irregularities such as formulae, figures, short text lines, large gaps between words, font designs, etc. In fact, in [5], the erroneously fitted curves are detected and excluded by a post processing based on several heuristics.

In this paper a model fitting technique with the following properties is proposed:

- The warping model of a document image is defined as a set of cubic splines.
- Each cubic spline is nonlinearly fitted to a text line or a space between text lines.
- The splines are optimized *globally*. That is, the splines depend on each other at their optimization.

The essential difference of the proposed technique from the conventional techniques is the last property. Namely, the splines are optimized while considering their global goodness on the whole warping model. This property is quite useful to make the dewarping result robust to the local irregularities. The second property implies another promising property that the spaces between text lines are utilized for stable results.

The rest of this paper is organized as follows: In Section 2, our warping model is defined and then the dewarping problem is formulated as a global optimization problem. An optimization algorithm based on dynamic programming (DP) is also provided. In Section 3, several experimental results are provided. Unfortunately, those experimental results confess that the proposed simple warping model should be improved for perfect dewarping results. The experimental results, however, emphasize successfully that the proposed global optimization framework possesses enough ability to provide stable dewarping results even when document images include several local irregularities that disturb the dewarping results provided by the conventional techniques.

2. Dewarping by global optimization

2.1. Warping model

The proposed warping model of a binarized document image (H pixel height and W pixel width) is defined as a set of H cubic splines. As shown in Fig. 1, the ith spline ($1 \leq i \leq H$) is controlled by the positions of three nodes, $P = (\alpha_i, 1)$, $Q = (i, \beta_i)$, and $R = (\gamma_i, W)$. The node P is the left-most point of the spline and moves vertically by controlling α_i. The node R is the right-most point of the spline and also moves vertically by γ_i. The remaining node Q is the center point of the spline and moves horizontally by β_i. The cubic spline which passes those three points can be uniquely determined by α_i, β_i, and γ_i. Our technique theoretically can compensate an arbitrary warp represented by the cubic spline. [1]

One of the novel features of the proposed technique is the use of constraints between the warpings of adjacent lines. Specifically, the following constraints are imposed in order to avoid large gaps between adjacent splines and intersections of the splines and improve the robustness against the local irregularities:

$$\left\{ \begin{array}{rcccl} 0 & \leq & \alpha_i - \alpha_{i-1} & \leq & 2, \\ -1 & \leq & \beta_i - \beta_{i-1} & \leq & 1, \\ 0 & \leq & \gamma_i - \gamma_{i-1} & \leq & 2. \end{array} \right. \quad (1)$$

With these constraints, we cannot optimize the fitting of the ith spline (i.e., we cannot optimize the parameters α_i, β_i and γ_i) independent of the $(i-1)$th and the $(i+1)$th splines. In other words, the splines are optimized while considering their global goodness on the whole warping model (i.e., all the H splines).

Note that the proposed warping model is "dense" in the sense that the above splines are prepared at all vertical positions i. Thus, the splines are fitted to not only text lines but also spaces between text lines. While the splines fitted to the spaces may seem somewhat redundant, they play an important role in the proposed technique as shown in the followings.

[1] From a practical viewpoint, the cubic spline model is slightly modified when the spline has "over-shoot" defined as the part outside of the circumscribing rectangle of P, Q, and R. The over-shoot part is clipped and connected by a straight line. This modification is very naive but effective for improving the accuracy of the spline model on approximating actual warped text lines.

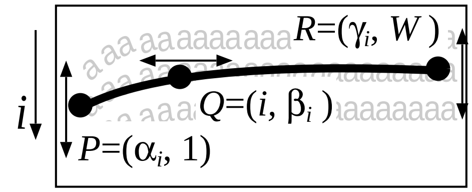


Figure 1. The ith spline, controlled by three nodes.

/ DP recursion */*
for $i := 1$ **to** H **do begin**
  for all possible $\theta_i = (\alpha_i, \beta_i, \gamma_i)$ **do begin**
    $g(\theta_i|i) := \{f(\theta_i|i) + \omega\lambda(\theta_i, \theta_{i-1})\} + \min_{\theta' \in \mathrm{Prev}(\theta_i)} g(\theta'|i-1)$
    $b(\theta_i|i) := \arg\min_{\theta' \in \mathrm{Prev}(\theta_i)} g(\theta'|i-1)$
  end
end
/ Back-tracking */*
$\theta_H^o := \arg\min_{\theta_H} g(\theta_H|H)$
for $i := H-1$ **down to** 1 **do**
  $\theta_i^o := b(\theta_{i+1}^o|i+1)$

Figure 2. DP algorithm for globally optimal parameters $\{\theta_i^o \mid i = 1, \ldots, H\}$, where $\theta_i^o = (\alpha_i^o, \beta_i^o, \gamma_i^o)$.

2.2. Criterion of model fitting

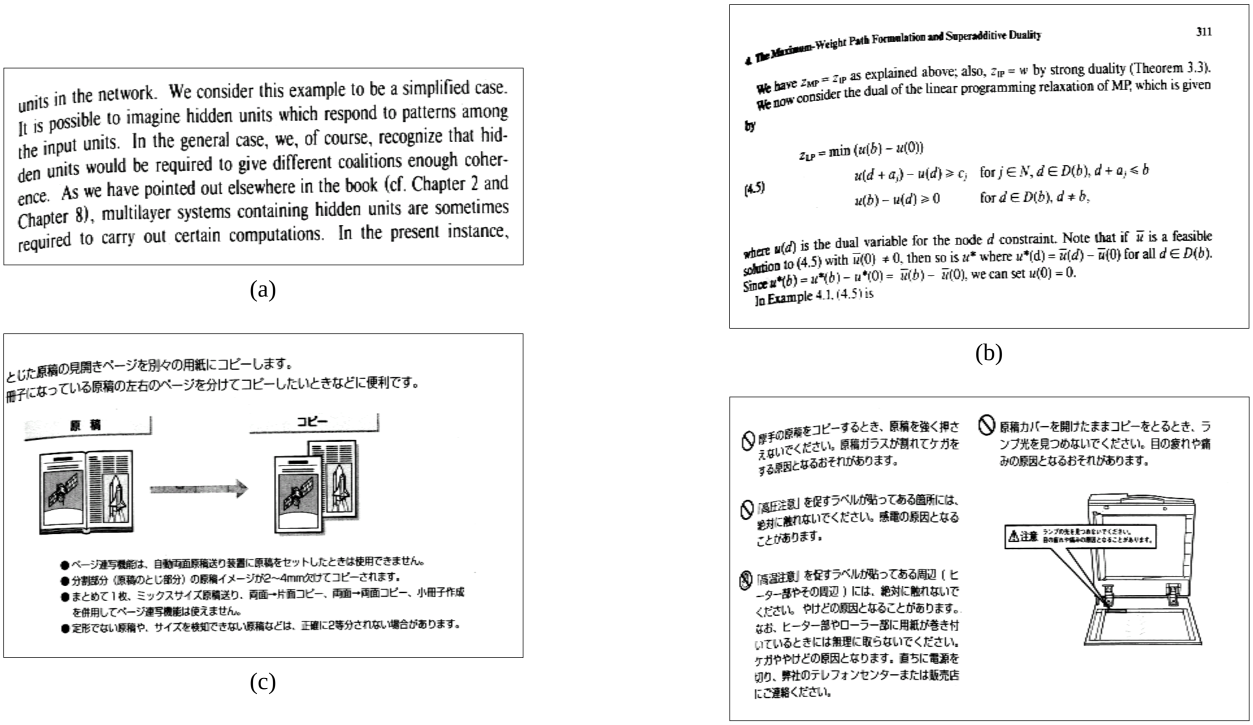
In the proposed technique, spaces between text lines are emphasized. The spaces are useful because they often express the warping of a document image more clearly and stably than text lines. That is, while text lines are often discontinuous by the gaps between words and sometimes very short, spaces between text lines are generally continuous and long. Thus, the splines fitted to the spaces will be more reliable than the ones fitted to the text lines.

According to this strategy, the optimal fitting problem can be formulated as the minimization problem of the following criterion function which evaluates the fitting of H splines,

$$J = \sum_{i=1}^{H} \{f(\theta_i|i) + \omega\lambda(\theta_i, \theta_{i-1})\}, \quad (2)$$

where $\theta_i = (\alpha_i, \beta_i, \gamma_i)$ and ω is a positive constant.

The function f evaluates the difference between the spline model and actual warp on the document image, especially emphasizing spaces between text lines. The function



units in the network. We consider this example to be a simplified case. It is possible to imagine hidden units which respond to patterns among the input units. In the general case, we, of course, recognize that hidden units would be required to give different coalitions enough coherence. As we have pointed out elsewhere in the book (cf. Chapter 2 and Chapter 8), multilayer systems containing hidden units are sometimes required to carry out certain computations. In the present instance,

(a)

4. The Maximum-Weight Path Formulation and Superadditive Duality 311

We have $z_{MP} = z_{IP}$ as explained above; also, $z_{IP} = w$ by strong duality (Theorem 3.3). We now consider the dual of the linear programming relaxation of MP, which is given by

$z_{LP} = \min\,(u(b) - u(0))$

(4.5) $u(d + a_j) - u(d) \geq c_j$ for $j \in N$, $d \in D(b)$, $d + a_j \leq b$

$u(b) - u(d) \geq 0$ for $d \in D(b)$, $d \neq b$,

where $u(d)$ is the dual variable for the node d constraint. Note that if $\bar{u}$ is a feasible solution to (4.5) with $\bar{u}(0) \neq 0$, then so is u^* where $u^*(d) = \bar{u}(d) - \bar{u}(0)$ for all $d \in D(b)$. Since $u^*(b) = u^*(b) - u^*(0) = \bar{u}(b) - \bar{u}(0)$, we can set $u(0) = 0$.

In Example 4.1, (4.5) is

(b)

とじた原稿の見開きページを別々の用紙にコピーします。
冊子になっている原稿の左右のページを分けてコピーしたいときなどに便利です。

- ページ連写機能は、自動両面原稿送り装置に原稿をセットしたときは使用できません。
- 分割部分（原稿のとじ部分）の原稿イメージが2～4mm欠けてコピーされます。
- まとめて1枚、ミックスサイズ原稿送り、両面→片面コピー、両面→両面コピー、小冊子作成を併用してページ連写機能は使えません。
- 定形でない原稿や、サイズを検知できない原稿などは、正確に2等分されない場合があります。

(c)

厚手の原稿をコピーするとき、原稿を強く押さえないでください。原稿ガラスが割れてケガをする原因となるおそれがあります。

「高圧注意」を促すラベルが貼ってある箇所には、絶対に触れないでください。感電の原因となることがあります。

「高温注意」を促すラベルが貼ってある周辺（ヒーター部やその周辺）には、絶対に触れないでください。やけどの原因となることがあります。なお、ヒーター部やローラー部に用紙が巻き付いているときには無理に取らないでください。ケガややけどの原因となります。直ちに電源を切り、弊社のテレフォンセンターまたは販売店にご連絡ください。

原稿カバーを開けたままコピーをとるとき、ランプ光を見つめないでください。目の疲れや痛みの原因となるおそれがあります。

(d)

Figure 3. Warped document images.

f is simply defined as

$$f(\theta_i|i) = \begin{cases} 0 & \text{if the spline specified by } \theta_i \\ & \text{does not pass any black pixel,} \\ 1 & \text{otherwise.} \end{cases} \quad (3)$$

Thus, the function f attains 0 only if the spline is fitted to a space between text lines. If the spline passes across a text line or lies on a text line, the value of f becomes worse, i.e., $f = 1$.

The function λ is a regularization function to relieve the following two properties of the function f;

- The evaluation by f is rough and ambiguous. Thus, for example, the splines with irregular intervals and those with regular intervals may have the same value of J. The splines placed with regular intervals are better in most cases.

- If only f is used in J, the splines tend to avoid being fitted to text lines.

The function λ is introduced to penalize irregular intervals and defined as

$$\lambda(\theta_i, \theta_{i-1}) = |\alpha_i - \alpha_{i-1} - 1| + |\beta_i - \beta_{i-1}| + |\gamma_i - \gamma_{i-1} - 1|. \quad (4)$$

Note that if only λ is used in J, the optimized splines are the straight horizontal lines placed with perfectly regular intervals.

The above criterion function f assumes that a document image does not include an elongated near-vertical black area due to shading. One remedy to relax this assumption is the use of a sophisticated binarization technique to remove the shading. Another remedy is slight modification of the criterion function f. For example, we can modify f to disregard the pixels whose vertical black run-length exceeds a threshold.

2.3. Dynamic programming-based algorithm for globally optimal dewarping

The globally optimal dewarping problem is now formulated as the constrained minimization problem of J with respect to $\{\alpha_i, \beta_i, \gamma_i \mid i = 1, 2, \ldots, H\}$. Since the constraints of (1) and the regularization function λ are defined between adjacent lines, i and $i-1$, the problem can be solved by DP.

Figure 2 shows a DP-based algorithm for the problem. Step 3 is the so-called DP recursion. The set $\mathrm{Prev}(\theta_i)$ is the set of $\theta_{i-1} = (\alpha_{i-1}, \beta_{i-1}, \gamma_{i-1})$ which satisfy the constraints of (1) with $\theta_i = (\alpha_i, \beta_i, \gamma_i)$. By repeating Step 3 from $i = 1$ to H for all possible splines θ_i, the minimum value of J is obtained as $\min_{\theta_H}\ g(\theta_H \mid H)$. Then the back-tracking operation using the back-pointer stored by Step 4 provides optimal fitting parameters $\{\theta_i^o = (\alpha_i^o, \beta_i^o, \gamma_i^o) \mid i =$

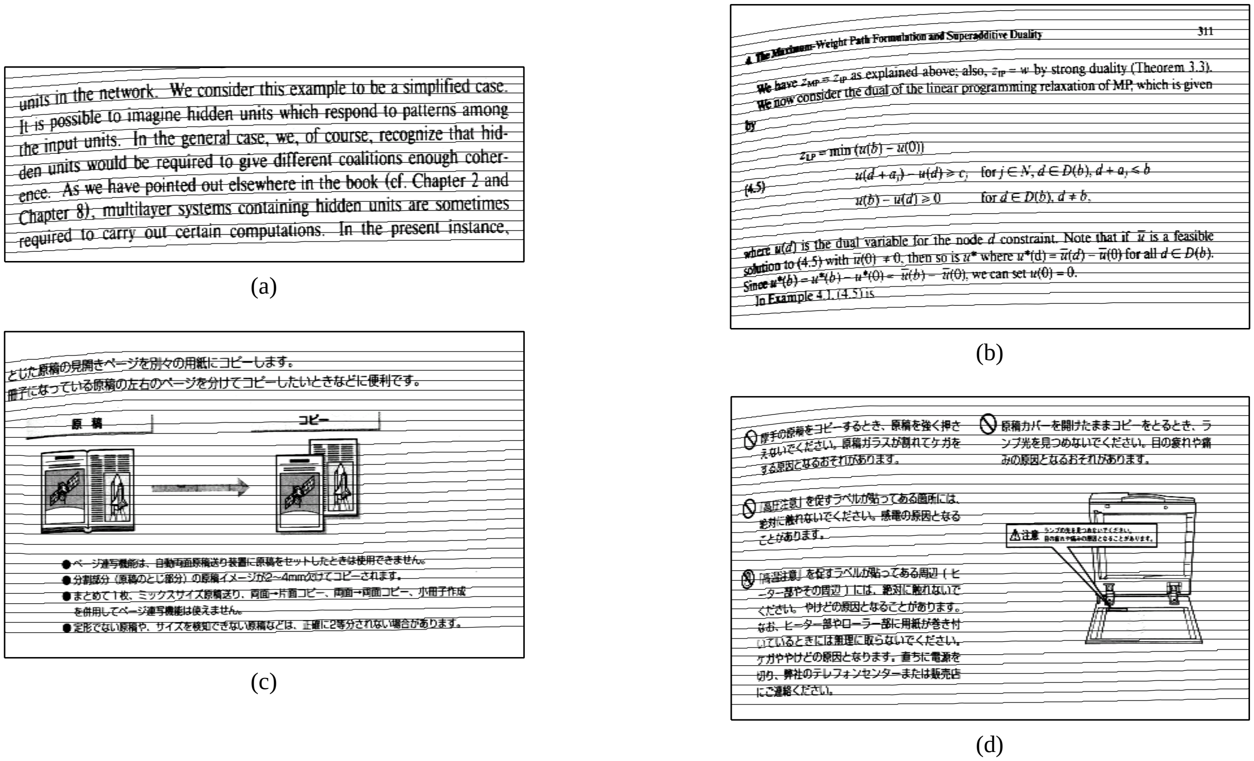


Figure 4. Optimized splines. Plotted on every 15 lines for visibility.

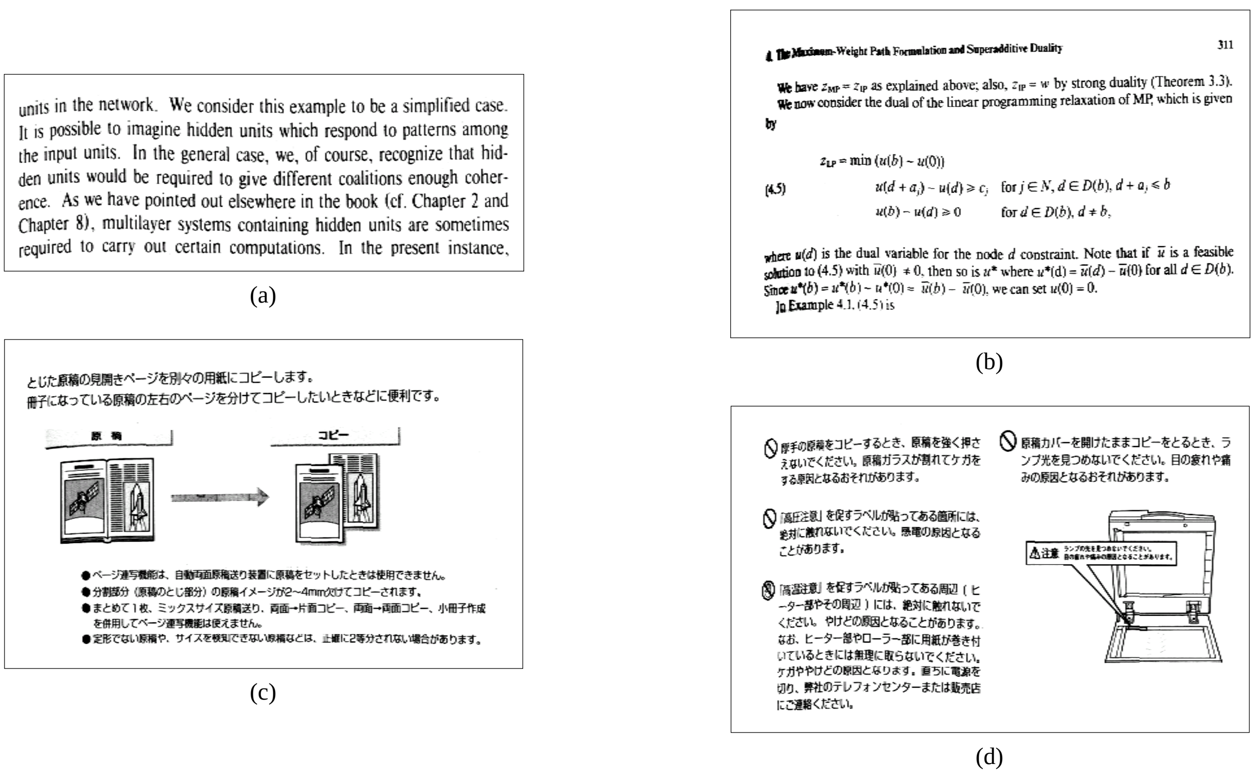


Figure 5. Dewarped images.

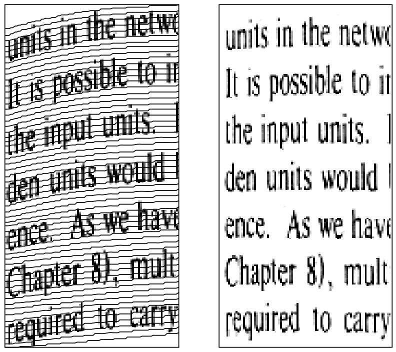


Figure 6. Dewarping limited to the region of curved lines.

$1, 2, \ldots, H\}$. Once these optimal parameters are found, the dewarped image is obtained immediately by sifting every pixel on the ith spline vertically to the ith row of the dewarped image. A practical problem is a huge amount of computations. Specifically, $O(H^3W^2)$ computations are required. One of remedies to relax this problem is the use of prunning in the DP search algorithm.

3. Experimental results

Figures 3, 4, and 5 show warped document images obtained by a flat-bed scanner, optimized splines, and dewarped images, respectively. In these figures, the image (a) consists of full lines only, (b) contains short lines such as a title and formulae, (c) contains a figure, and (d) is in the double column format. Note that the weight ω was set at 1 by some preliminary experiments.

The conventional techniques that estimate local warps by using the connectivity of text line (e.g., [5]) may fail in the cases of (b) and (c), i.e., document images including short text lines, formulae, or figures, since the connectivity of lines are broken many times. On the contrary, the proposed technique achieves appropriate fittings in the whole document images for all the cases.

It is observed that warps still remain in the dewarped images of Fig. 5. The most dominant reason of such results is the limited expression ability of our warping model. These results indicate that the cubic spline with three nodes (Fig. 1) cannot simulate the actual warping of text lines completely. There are the following remedies for solving this problem;

- Improving expression ability of the cubic splines, such as assigning more control nodes around the curved part of the text lines, and vertically dividing a document image into some partial document images.
- Employing a more accurate warping model of bound volumes, such as [6, 7]. Especially, the employment of a cylinder model [6] is promising for our framework.

According to the first remedy, we have experimented with a local improvement of the expression ability of the warping model by limiting the region of fitting. Only the curved parts of the lines in Fig. 3(a) were extracted manually and then dewarped using three control nodes. It is found that almost no warps remained in the resultant image (Fig. 6), comparing with the result of Fig. 5(a). It indicates that the model was fitted to the lines with a higher accuracy. This result indicates that it is effective to dewarp after vertically dividing a document image into some partial document images.

4. Conclusion

We have proposed in this paper a dewarping technique of document images. Differing from the conventional techniques, this technique is applicable to document images including formulae, short text lines, or figures, since the proposed technique dewarps whole document images by fitting splines while considering the global optimality that specifies the desirable relationship among the splines. The experimental results on several practical document images have indicated the effectiveness of the proposed technique. Our experiment has also indicated the effectiveness of the vertical division of a document image into some partial document images for more accurate dewarping.

The construction of a more accurate line-warping model while reducing the computational complexity is one of our future works.

References

[1] M. S. Brown, W. B. Seales, "Image restoration of arbitrarily warped documents," IEEE Trans. PAMI, vol. 26, no. 10, pp. 1295–1306, 2004.

[2] A. Doncescu, A. Bouju, and V. Quillet, "Former books digital processing: image warping," Proc. Workshop on Doc. Image Anal., pp. 5–9, 1997.

[3] T. Wada, H. Ukida, T. Matsuyama, "Shape from shading with interreflections under a proximal light source: distortion-free copying of an unfolded book", Int. J. Comput. Vis., vol. 24, no. 2, pp. 125–135, 1997.

[4] Z. Zhang, C. L. Tan, "Correcting document image warping based on regression of curved text lines," Proc. ICDAR, pp. 589–593, 2003.

[5] C. Wu, G. Agam, "Document image de-warping for text/graphics recognition," SSPR&SPR 2002,LNCS 2396, pp. 348-357, 2002.

[6] H. Cao, X. Ding, and C. Liu, "Rectifying the bound document image captured by the camera: a model based approach," Proc. ICDAR, pp. 71–75, 2003.

[7] T. Kanungo, R. M. Haralick, and I. Phillips, "Global and local document degradation models," Proc. ICDAR, pp. 730–734, 1993.