Pattern Recognition 48 (2015) 3600–3614

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Document dewarping via text-line based optimization

Beom Su Kim [a], Hyung Il Koo [b,*], Nam Ik Cho [a]

[a] Department of Electrical and Computing Engineering, INMC, Seoul National University, South Korea
[b] Department of Electrical and Computer Engineering, Ajou University, South Korea

ARTICLE INFO

Article history:
Received 1 November 2014
Received in revised form
9 April 2015
Accepted 27 April 2015
Available online 9 May 2015

Keywords:
Document image
Document rectification
Document dewarping
Generalized cylindrical surface
Optical character recognition
Text-line based dewarping

ABSTRACT

This paper presents a new document image dewarping method that removes geometric distortions in camera-captured document images. The proposed method does not directly use the text-line which has been the most widely used feature for the document dewarping. Instead, we use the discrete representation of text-lines and text-blocks which are the sets of connected components. Also, we model the geometric distortions caused by page curl and perspective view as the generalized cylindrical surfaces and camera rotation respectively. With these distortion models and the discrete representation of the features, we design a cost function whose minimization yields the parameters of the distortion model. In the cost function, we encode the properties of the pages such as text-block alignment, line-spacing, and the straightness of text-lines. By describing the text features using the sets of discrete points, the cost function can be easily defined and efficiently solved by Levenberg–Marquadt algorithm. Experiments show that the proposed method works well for the various layouts and curved surfaces, and compares favorably with the conventional methods on the standard dataset.

1. Introduction

The optical recognition of printed material has been an important problem in image processing and computer vision communities. Conventional approaches to this problem are to convert the printed documents into digital images with flatbed scanners, and then perform layout analysis and optical character recognition (OCR) [1,2]. However, as mobile (or wearable) devices equipped with high-performance digital cameras are becoming widely available, many researchers tried to replace the flatbed scanners with digital cameras [3–5]. The digital cameras have a number of advantages over the flatbed scanners, such as portability, fast response, non-contact property (no physical damages on documents), and unconstrained view points. However, the camera-captured document images usually suffer from geometric distortions caused by capturing processes (perspective distortions) and curved surfaces. In order to reduce these geometric distortions, a number of approaches have been proposed [5–16]. Among various approaches, the text-line based approach has received a lot of attention, because text-lines are the most salient properties of documents and this approach allows us to develop an algorithm that needs a single image as the input.

In this paper, we propose a new document dewarping algorithm that also exploits the text-lines. Although there are many methods based on the text-lines as mentioned above and our method is also based on some conventional techniques (i.e., the text-line and text-block extraction method [3] and the surface model of conventional methods [17–19]), the proposed method is different from them in several aspects. First, we formulate the dewarping problem as an optimization problem based on the given text-line and text-block information. Note that many algorithms consist of a series of steps, such as vanishing point estimation and aspect ratio estimation, where each step assumes that the previous steps are successful. On the other hand, we address the same problem by optimizing a single cost function, whose minimization yields a set of parameters enabling the reduction of geometric distortions. The cost function is designed to exploit various text cues (e.g., layouts, line spacing), so that the proposed method works for a variety of inputs (e.g., center-aligned text blocks and multiple columns). Second, our method does not require baseline fitting with polynomials or splines, so that we just use an existing text-block identification and text-line extraction method in [3] for the feature extraction. Finally, our method deals with not only planar documents but also curved surfaces in the same framework, which has been treated in a different manner [5,15]. Moreover, our method can be extended to the unfolded book surfaces by changing the surface models. To the best of our knowledge, the proposed method is the first approach that can deal with unfolded book surfaces with text-lines. We also make

* Corresponding author.
E-mail addresses: bskim@ispl.snu.ac.kr (B.S. Kim), hikoo@ajou.ac.kr (H.I. Koo), nicho@snu.ac.kr (N.I. Cho).

http://dx.doi.org/10.1016/j.patcog.2015.04.026
0031-3203/

our algorithm publicly available at our website (http://ispl.snu.ac.kr/bskim/DocumentDewarping).

The rest of this paper is organized as follows. In Section 2, we briefly review the conventional dewarping methods. Since our method is based on the text-lines, we also discuss the challenges in the text-line extraction problem. In Section 3, we present models and constraints used in our optimization-based framework. The details of the proposed cost function and its minimization are explained in Section 4. Experimental results on synthetic and real data are shown in Section 5 and we conclude this paper in Section 6.

2. Related work

Although there are some model-free approaches to the document dewarping problem [6,7], most of the existing works assume certain geometric models of document surfaces and try to find the global transformations that remove geometrical distortions in the camera-captured images. In this section, we review the conventional dewarping methods and then discuss the challenges in the text-line based approach.

2.1. Document dewarping without text-lines

Since the 3D structure of an object can be estimated by using structured lights or laser scanners, many document dewarping methods were developed based on shape measuring hardwares [8,9,20,21]. Although these approaches allow us to handle a range of surface distortions and yield accurate results, the need of special equipments is their major drawback. We can also reconstruct 3D structures from multiple images, and this approach is applied to the document surfaces in [17,22]. Although they can rectify pages without specialized hardwares, high complexities and the need of multiple images are the problems. The 3D shape information of documents can also be estimated from a single image with special assumptions. For example, under controlled lighting conditions, we can reconstruct document surfaces by the shape-from-shading approach [10,11]. While they provide a theoretically sound rectification framework, the strict assumptions on illuminations are not practical in many situations. There is another single-view based method that removes geometric distortions in camera-captured images by exploiting the regularity of textures [23]. However, there are limitations in dealing with the documents that have complex layout and non-text regions, where textures are no longer regular.

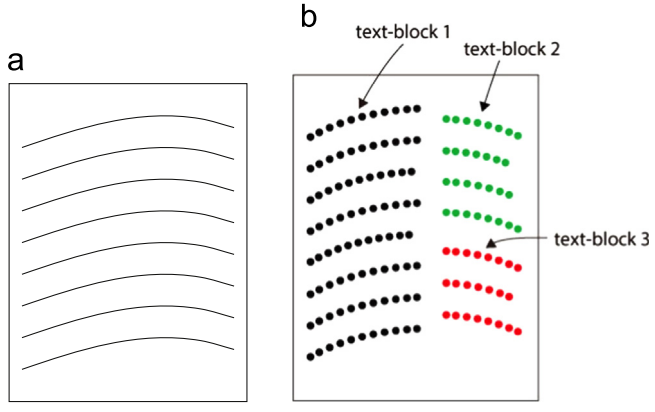

Fig. 1. Text-line representation. (a) Continuous representation of text-lines (polynomials or B-splines), which is used by the conventional methods and (b) text-line and text-block representation [3] adopted in this paper.

2.2. Document dewarping with text-lines

Among various approaches to document image dewarping, the most popular approach is the text-line based method. It is probably because the text-line is the most salient feature of documents and this approach enables a single-view algorithm. The authors in [12,24] considered the text-lines as top and bottom boundaries and applied boundary interpolations, which were also used for the rectification of pictures under the generalized cylindrical surface (GCS) assumptions [17–19]. This approach is efficient, but it lacks theoretical justification and works only for the single-column and justified layout. Some methods estimated vertical lines [14,15,25] and vertical stroke boundaries [26–28] as well as text-lines (horizontal directions). Since 2D text-field is available in these approaches, they less suffer from the ambiguities and are able to handle complex document surfaces [14]. However, building the 2D text-field from document images (possibly having complex layouts, non-textual objects) is a more challenging problem than the text-line extraction problem. Recently, the authors in [5] showed that GCS surfaces can be reconstructed from the horizontal text-lines only (with mild approximations). Even though its theoretical background is sound, its sensitivity to the errors in baseline fitting is not addressed. Also, this approach works only for curved surfaces having the single-column layout.

2.3. Challenges in text-block identification and text-line extraction

Text-block identification (text-region segmentation) is a very important topic in document image processing, and many methods have been developed based on a variety of techniques, such as filter banks, learned classifiers, and the tensor representation of image patches [29–31]. However, they mainly focussed on scanned documents, and they were not incorporated into document dewarping algorithms. Rather, conventional dewarping algorithms assumed user interactions for border noise removal, or started from given text regions [5,15]. Therefore it is not clear whether the concatenation of (i) a conventional text-block identification algorithm and (ii) a conventional dewarping algorithm based on noise-free assumptions works well for practical cases.

Also, the conventional dewarping methods required continuous baselines as illustrated in Fig. 1(a), and they tried to estimate these lines by fitting the fiducial points of characters with polynomials [5]. The accuracy of this fitting procedure is one of the key factors for the success of existing text-line based methods [12,14,5], however, the sensitivity to the fitting error is not clearly understood.

3. Our approach

The text-block identification and the text-line extraction in camera-captured document images are challenging problems due to perspective distortions, geometric distortions, and non-textual objects (including background clutters). In order to deal with these challenges, we adopt a method in [3]. The method first estimates the scale and orientation of each connected component (CC) with the optimization framework, and builds text-lines and text-blocks by partitioning CCs with the estimated states, as shown in Fig. 1(b). Unlike the conventional methods that tried to represent the baselines as continuous curves [12,14,5], we work directly with the set of extracted CCs [3], which is represented by

$$\mathcal{P} = \{p_i^{k,j} \in \Re^2\} \tag{1}$$

where $p_i^{k,j}$ is the center of the i-th CC in the k-th text-line in the j-th text-block. Our discrete representation (as shown Fig. 1(b)) is a novel approach and it provides two major advantages over the conventional models. First, we do not need to perform the baseline