Image Dewarping and Text Extraction from Mobile Captured Distinct Documents

H.K.Chethan[a],*, G.Hemantha Kumar[b]

[a]*Research Scholar,DoS in Computer Science,University of Mysore,Mysore-57006*
[b]*Professor,DoS in Computer Science,University of Mysore,Mysore-57006*

Abstract

Camera Based Document Analysis (CBDA) is an emerging field in computer vision and pattern recognition. In recent days, cameras are moulded with several items of additional equipment. Thus, they play a vital role in the replacement of scanners with hand-held imaging devices (HIDs) like digital cameras, mobile phones and gaming devices. Warping is a common appearance in camera captured document images. It is the primary factor that makes such kind of document images hard to be recognized. Therefore it is necessary to restore warped document images before recognition. This paper presents a novel methodology to dewarp text from camera-captured warped images by co-ordinate transform model . Further, the extracted text from the warped and other document images will be recognized by maintaining a suitable database of all letters and numbers and converted into an editable form such as Notepad or as a MS Word document. The experimental results will be evaluated using a novel method to authenticate the methodology designed. Extensive experiments have been carried out on distinct documents and results are tabulated. Experimental results show the effectiveness of the proposed method.

ⓒ 2010 Published by Elsevier Ltd Open access under CC BY-NC-ND license.
Keywords: CBDA; Computer Visio;, Warping, Pattern recognition;HIDs; Template*;*

1. Introduction

Character recognition has been continuous importance for several years and recently new challenges have been devised in the field of document analysis. During the past decades, document analysis systems were mainly developed for scanner devices. Work in the field covers many different areas including pre-processing, Physical and logical layout analysis OCR/ICR, graphic analysis, form processing, signature verification and writer identification has been applied in numerous domain including office automation, forensic and digital libraries [1, 2, 3, and 4].Digital camera is now opening a new vista in Document analysis due to their capability to capture documents anywhere in the 3D environment [5] which gave rise to CBDA (Camera Based Document Analysis) [6]. Recent studies have shown that there is a shortfall of sales in scanners compared to HIDs [Kim, 2005].CBDA is characterized by a wider variety of input methods including digital still and Video cameras due to their capability to capture documents anywhere in the 3D environment [5].Recent invention like Portable digital cameras were developed for taking picture of joy and sad moments but the recent price -performance improvement, low weight ,portability ,low cost ,small dimension has given birth to several new application[1][7].Recent study reveals that there is a great amount of interest in content retrieval from images and videos[8].With the help of digital camera we can capture characters and documents anywhere in the 3D environment like signs and billboards,color,texture,shape as well as the relationship between them. One of the main research directions in camera-captured document analysis is to deal with warped images 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.

* Corresponding author. Tel.:+91-821-251-0789
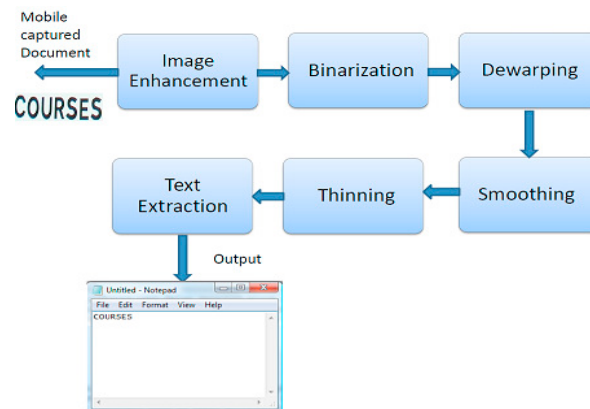*E-mail address*: chethan@compsci.uni-mysore.ac.in

Figure 1: Proposed Block Diagram

The goal of the work is to dewarp 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 [9]. 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 [10,11], structured light sources [8], or laser scanners [9].
2. Reconstruction of the page using a single camera in an uncontrolled environment [12, 13, 14].

Despite the existence of so many approaches for dewarping of documents, 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 In the following sections, We present dewarping technique in a easy way by creating the templates such that it is rotated and further experimental results on several distinct images classified manually are demonstrate to see the efficiency of the proposed technique.

2. Proposed Methodology

The proposed method has three important stages as described below.

2.1. Pre-processing

The key function of the pre-processing is to improve the image in ways that increases the chances for success of other process. It ultimately deals with enhancement of image. Camera captured images usually captured in uncontrolled environment suffer from noise due to low brightness contrast and various illuminated environment, low resolution and broken characters. In this step camera captured image is converted into a gray level image and image enhancement algorithm is applied to the image the enhanced image is binarized and then the noises are removed by salt and pepper algorithm[15].Camera captured image is converted to gray scale image as below:

$$I_s(x, y) = (0, 1, 2, ..255) \tag{1}$$

$1 \le x \le I_x$

$1 \le y \le I_y$

Where 0 corresponds to black and 255 to white.

2.2. Image enhancement

In image enhancement, a variety of methods exist for removing image degradations and emphasizing important image information, and in computer graphics, digital images can be generated, modified, and combined for a wide variety of visual effects. A gray level image has been binarized for classifying foreground pixels from background ones. Partly very low contrast of intensity on image exists because of illumination variation and photographing angle. It causes misclassification of foreground characters from background. To overcome this problem, image enhancement by gray level normalization is processed before binarization. The formula for enhancement of an image is described as follows: The reduce operation is carried out by convolving the image with a Gaussian low pass filter.

$$f_1(x, y) = (L - 1)\frac{f(x, y) - min}{max - min} \tag{2}$$

$$max = max[f(x, y)] \tag{3}$$

$$min = min[f(x, y)] \tag{4}$$

Where $f(x, y)$, $f_1(x, y)$ denotes the gray level value at pixel (x, y) and the pixel level after image enhancement, respectively. L denotes gray level range of image to be converted and M denotes the height and width of image. Max and min are maximum value and minimum value among pixels on image, respectively.

3. Dewarping

Digital Image De-warping is the technique that deals with the geometric transformations on an image [13]. One of the major challenges in camera document analysis is to deal with the wrap and perspective distortions. In spite of the prevalence of dewarping techniques, there is no standard algorithm for the performance evaluation method [9], with most of the evaluation done are to concentrate in visual pleasing impressions. Our main aim is to recover document images that are mainly bounded volumes captured by a digital camera and suffer from non-linear warp. The proposed technique is applied on gray scale document images and is based on several distinct steps using an adaptive document image binarization[14], finally, a complete restoration of the original grayscale warped image guided by the binary dewarping result using co-ordinate transform model where 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 .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.

Figure 2: Sample Wrapped Image

Figure 3: Dewarpped Image

4. Binarization

Binarization is the starting step of most document image analysis systems and refers to the conversion of the grayscale image to a binary image. Image binarization converts an image of up to 256 gray levels to a black and white image. Survey[16] have showed that global thresholding is not ideal for Camera-captured images due to lightning variations so We proposed locally adaptive thresholding method that is robust to variation of illumination. The simplest

4

Figure 4: Image after Binarization

way to use image binarization is to choose a threshold value, and classify all pixels with values above this threshold as white, and all other pixels as black.

$$I_S(x, y) = (0, 1) \qquad (5)$$

$$T_{rxr} = T * g_{high} - g_{low} \qquad (6)$$

Where g_{high} and g_{low} are maximum and minimum intensity value of pixels in rxr sub-window respectively.

5. Thinning

It is well known that one of the key issues in pattern recognition area is how to extract distinctive features from the input pattern. A pattern itself consists of lots of data bits, many of which can be deleted from the image without distorting the information. In other words, through the process called "thinning" the number of data of an input pattern can be significantly reduced without losing any stroke features. Thinning algorithms are commonly classified into (1) sequential and (2) parallel algorithms. Main difference between these two types is that sequential algorithm operates on one pixel at a time and the operation depends on preceding processed results, while parallel algorithm operates on all the pixels simultaneously. In the thinning process, a 3x3 window is generally used as shown in figure 6.

1. Here pixel 'P0' is the central pixel under consideration.
2. N (P0) is the number of non-zero neighbors of 'P0'
3. T (P0) is counted as one when a neighbor is valued 'zero' and its immediate neighbor under consideration is valued '1'

P8	P1	P2
P7	Po	P3
P6	P5	P4

Figure 5: 3 x 3 Window

Figure 6: Thinning applied

Here in our work we employ Aradhya's Thinning algorithm which resolves the problems of is connectivity and end-point reduction that take place in one-pass thinning algorithm. Furthermore, this algorithm is effective in high speed operation. A processor for this algorithm, capable of handling variable image-sizes (from 25 to 40 bits in width), is implemented and tested. This processor's capability of high speed operation and flexibility should find excellent applicability in various areas.

6. Text Extraction

Text data present in images and video contain useful information for automatic annotation, indexing, and structuring of images [1 and 2]. Extraction of this information involves detection, localization, tracking, extraction, enhancement, and recognition of the text from a given image. However, variations of text due to differences in size, style, orientation, and alignment, as well as low image contrast and complex background make the problem of automatic text extraction extremely challenging. A Text Information Extraction (TIE) system receives an input in the form of a still image or a sequence of images [18]. The images can be in gray scale or color, compressed or un-compressed

Figure 7: Bounding Box

Figure 8: Text Extraction

7. Template Matching

Template matching is a technique in digital image processing for finding small parts of an image which match a template image. The next stage after extracting an object from a document image is template matching for recognition of characters. This task is done by maintaining a database of characters (A-Z) and digits (0-9) in one folder called template folder. The elements inside the template folder are called templates. The matching process moves the template image to all possible positions in a larger source image and computes a numerical index that indicates how well the template matches the image in that position. Template matching is carried out through Correlation, which is a measure of the degree to which two variables agree, not necessary in actual value but in general behavior. The two variables are the corresponding pixel values in two images, template and source.
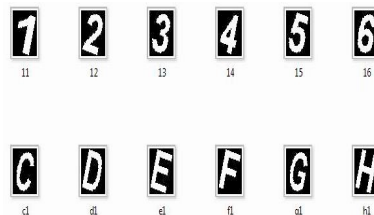

Figure 9: Sample Templates

Figure 9 describes the template database which will be used for matching the data received or extracted in the input image. The object extracted from a document image is correlated and every element which is present inside the template folder .When there is a match the extracted object will be stored in one array. For example character 'M' as shown in figure 10 which is extracted from a document image is correlated with each and every template which is present inside the template folder. When the extracted object is correlated with the template the object is recognized

as a particular text and it will be stored in an array. The process will be continued for each and every extracted object since each character is considered as an object in the present context of the problem.
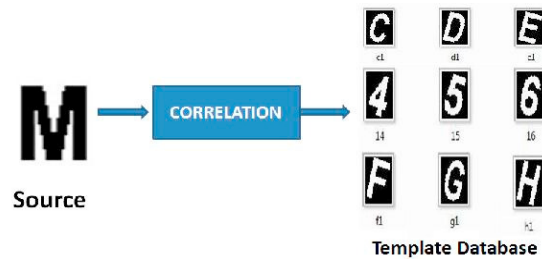

Figure 10: Templates Matching

8. Database

In order to capture images of any material containing the text, we use a mobile mounting stand. This apparatus firmly holds the mobile camera and thus avoids the distortion of the images. The images are captured under uncontrolled environment such as distance between the object of interest and the mobile camera, intensity of light, alignment, varying surfaces of the object like cylindrical, spherical or any curved surface [17].The experimental setup for capturing an image containing text requires a 2 Mp mobile camera, a mobile camera mounting stand .The same image is captured at different distances from the mobile camera. Further, the intensity of light is also varied for the images at different distances. The images captured are manually classified/ categorized into five types, where type 1 is of clean image, type 2 is of blur image, type 3 is of image captured with different variance and orientation, type 4 is of kind under various illumination condition and type 5 is of documents pertained to Dewarp.

Figure 11: Sample Database

The resolution of the image captured from the mobile camera is 960 X 1280. The size of each image captured varies from 60 - 80Kb. The variations of text present in the images captured will be due to variations in size, style, orientation, and alignment, as well as low image contrast and complex background. Some of the samples are presented in figure 11 for better understanding

Table I: Results of the proposed Method

Sl. No.	Group	Type of Image	Recognition Rate before Restoration (%)	Recognition Rate after Restoration (%)
1	Type 1	Clea	88.5	98.0
2	Type 2	Blu	70.0	83.3
3	Type 3	Different Orientation	72.0	86.9
4	Type 4	Lighting condition	75.8	87.
5	Type 5	Dewarped Documents	58.0	92.

9. Experimental Results and discussion

The proposed method was tested using large dataset which include clean, blur, illumination, and arbitrarily warped documents. The performance is calculated based on the Recognition rate and false recognition rate as mentioned below in the equations.

$$\text{Recognition Rate} = \frac{CDW}{CDW+FDW} *100\% \tag{7}$$

$$\text{False Recognition Rate} = \frac{FDW}{CDW} *100\% \tag{8}$$

Correctly detected words are the block that contains text. False detected block does not contain any text. Recognition rate is ratio of correctly detected word to the sum of correctly detected words (CDW) added with false detected text (FDW). False rate is detected by the ratio of false detected text to the correctly detected block. The results obtained on images are shown in figure 7 and 8. From table 1 it is clear that recognition rate obtained before rotation for all the type is worse than the recognition rate obtained after rotation. A satisfying result of 98.03% is obtained for clean images and very less recognition rate has been reported for the type blur images .Where as Dewarped documents show a promising result of 92.50% às mentioned in Table 1.Almost there is an average of 10 to 15% increase in Recognition rate after applying our technique.

10. Conclusion and Future Enhancement

In this paper we present a novel methodology for efficient restoration of wrapped document images and also extraction and recognition of text in camera captured distinct documents. The proposed scheme is applied on images of gray scale images and is based on an adaptation of document image binarization and some pre-processing step where labeling is done based on connected component and dewarping is done based on the simple logic of Co-ordinate transform method and text is recognized using correlation between the object extracted from the document and the template database .Experimental results show that the proposed method of Text extraction and recognition is fast, accurate and robust In the present work, we have worked with multicolored samples but the efficiency of the algorithm was better in non multicolored text images. In brief, the work provides an overview of template matching of alphanumeric text which further could be enhanced for multilingual and multitextural images.

References

[1] Jian Liang, David Doermann, Huiping Li (2005), Camera Based analysis of Text and Documents: A Survey. IJDAR
[2] Keechul Jung. Wang IN Kim, Nil K.Jain (2004), Text Information Extraction in Images and Videos.
[3] Plamondon R, Srihari S (2000) On-line and off-line handwriting recognition: a comprehensive survey. IEEE Trans Pattern Analysis and Machine Intelligence 22(1):63-84
[4] Vinciarelli A (2002) A Survey on off-line word recognition. Pattern Recognition 35:1433-1446
[5] Doermann D (1998) .The indexing and retrieval of document images: a survey. Compute Vis Image Understand 70(3):287-298.
[6] Cai M, Song J-Q, Lyu MR (2002) A new approach for video text detection. In: In: Proc. ICIP, pp 117-120.
[7] Majid Mirmehdi: Special issue on camera-based text and document recognition, Springer-Verlag .2005.
[8] Palaiahnakote Shivkumara, Weihua Huang and Chew Lim Tan ,Efficient Video Text Detection using Edge Features,IEEE 978-1-4244-2175,2008.

[9]   J. Liang, D. Dorermann, and H. Li. Camera-based analysis of text and documents: a survey. Int. Jour. of Document Analysis and Recognition, 7(2-3):84104, 2005.

[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 192196, 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 10011005, Aug. 2005.

[12] J. Liang, D.F. DeMenthon, and D. Doermann. Flattening curved documents in images. In Proc. Computer Vision and Pattern Recognition, pages 338345, 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 971974, Aug. 2006.

[15] P.Shivkumara, G.H.Kumar: New Filter Based Unsupervised Rules for Boolean Metric, ICCTA, Kolkata, India, Pp.611-617, 2007.

[16] Y.Zhong and H.Zhang and A.K.Jain'Automatic Caption Localization in Compressed Video IEEE Trans. On Pattern Analysis and Machine Intelligence 22(4), pp.385-392, April 2000.

[17] Stephen Pollard and Maurizio Pilu'Building camera for capturing documents,IJDAR(2005)7:123-137 10.1007/S10032-004-0129-0.

[18] Keechul Jung, Kwang Kim and Anil K. Jain: Text Information Extraction in Images and Video: a Survey, Pattern Recognition, 37 PP.977-997, 2004.