SMARTPHONE BASED DOCUMENT SCANNER APP WITH IMAGE PERSPECTIVE CORRECTION

Bhasham Vishva Vardhan
Computer science and engineering
Lovely professional university
Jalandhar, India
bvishvavardhan@gmail.com

***Abstract*—In today's digital era, smartphones have become essential tools for various daily tasks, including scanning documents. Smartphone apps designed for document scanning offer incredible convenience, allowing users to effortlessly convert physical documents into digital formats. However, this convenience comes with a common challenge: perspective distortion. When using a smartphone camera to capture documents, deviations from the ideal perpendicular alignment between the camera and the document surface can lead to skewed and distorted images. These distortions can undermine Optical Character Recognition (OCR) accuracy and the overall usability of scanned documents, particularly for archival and professional purposes.**

This research aims to enhance the accuracy and usability of document scanning on smartphones by addressing the issue of perspective distortion. It focuses on the crucial task of image perspective correction, which aims to rectify these distortions and make document images more faithful to their original form. The goal is to empower users to generate high-quality digital documents that meet the reliability standards required for document archiving, information retrieval, and OCR-based data extraction.

The research methodology involves a thorough examination of image perspective correction techniques, with a focus on three primary methods: geometric transformations, convolutional neural networks (CNNs), and keypoint-based approaches. Each technique offers a unique approach to perspective correction, and this research provides an in-depth comparison to determine their effectiveness and practical relevance.

To do this, the study utilizes a diverse dataset collected from various smartphone models using popular document scanning apps. The dataset comprises various document types, such as business cards, receipts, contracts, and handwritten notes, ensuring a diverse range of document characteristics encountered in daily life. The research evaluates each perspective correction technique both quantitatively and qualitatively, enabling a comprehensive assessment of their performance.

Quantitative evaluation employs various metrics, including accuracy and F-measure, to rigorously analyze the effectiveness of each method. The results not only highlight the success of each technique but also identify their limitations in different contexts. Qualitative evaluation complements these findings by providing visual examples that illustrate how perspective correction techniques impact scanned documents in real-world situations.

The paper addresses the need for automatic text recognition in smartphone-captured document images. The proposed system integrates preprocessing, feature extraction, and classification phases. Preprocessing locates text regions and segments them into text lines. A deep convolutional neural network (CNN) is used to extract features from frames within these lines, and a combined architecture involving bidirectional recurrent neural network (RNN), gated recurrent units(GRU) block, and connectionlist temporal classification (CTC) layer is employed for text classification. The system was tested on the ICDAR2015 Smartphone document optical character recognition(OCR) dataset, demonstrating its ability to achieve promising recognition rates.

***Index Terms*—Text recognition, Document image,perspective correction, Convolutional neural network, optical character recognition, Recurrent neural network**

I. Introduction

In recent years, our interaction with audiovisual information has undergone a profound transformation. Images and videos have become integral components of the daily information we consume. How we manage and utilize this influx of visual data can significantly impact our daily lives. While text documents, such as books, printed articles, patents, receipts, and magazines, have traditionally served as a structured means of presenting information, a shift has occurred. More and more people are choosing to capture text documents through smartphone photos rather than digitizing them with traditional scanners. This shift can be attributed to the widespread adoption and rapid advancements in smartphone technology.

Despite this trend, most existing systems, typically referred to as Optical Character Recognition (OCR) systems, have been primarily designed for scanned text documents. These systems have been optimized for the controlled environment of scanned documents and may not perform as effectively with text documents captured using mobile devices. It's important to note that text document images acquired by smartphones and similar devices constitute a significant portion of our multimedia content.

Processing and recognizing text within these smartphone-captured document images pose unique challenges. The variations in lighting, angle, and quality of these images, as well as the presence of noise and other artifacts, make the task more complex. Consequently, there is a need for innovative solutions and dedicated systems that can effectively handle the recognition and processing of text in this category of

document images. Addressing this challenge will enable us to make better use of the text content we encounter through our smartphones, ultimately enhancing our interaction with information in the digital age.

The processing of images captured by smartphones is influenced by various factors. One significant factor is perspective distortion, which occurs when the text being captured is not aligned parallel to the smartphone's camera plane. This distortion results in characters appearing smaller the farther they are from the camera, and it disrupts the assumption of parallel lines on the document page when compared to the captured image.

Another challenge stems from the limited control mobile devices have over lighting conditions during image acquisition. Lighting variations are common due to environmental factors like shadows, reflective surfaces, and the absence of controlled lighting setups. Complex backgrounds can also pose difficulties, especially when the area to be captured extends beyond the text region.

Blur distortion is another issue that may arise in document images captured by smartphones. This distortion occurs when the smartphone's camera focuses on the background rather than the text document, or it can be caused by the movement of the camera during image capture.

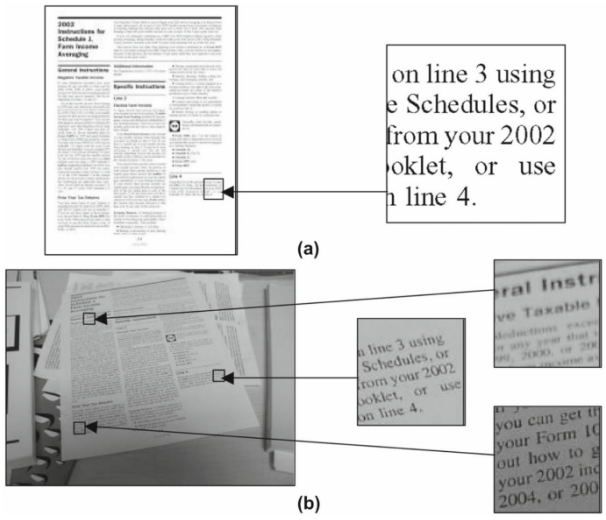


Fig. 1 Challenges of document images acquired by a mobile terminal [30]. **a** Document scanned using a scanner. **b**: The same document captured by a mobile terminal

Furthermore, the processing of smartphone-captured images is impacted by the diverse text properties found in different documents. These properties include variations in text sizes, fonts, styles, and colors. Addressing these challenges is crucial for effectively processing and recognizing text in images taken with mobile devices.

II. Literature Review

A. Image Perspective Correction Techniques

Image perspective correction is a crucial task within the realm of computer vision, particularly in the context of document scanning applications on smartphones. Over time, multiple methodologies have been developed and refined to tackle the issue of perspective distortion in scanned images. In this section, we will explore the primary methods employed for image perspective correction, as discussed in existing literature:

Geometric Transformations: Geometric transformations encompass techniques like affine transformations and homography, which have gained widespread use in rectifying perspective distortion. These methods rely on mathematical transformations to reposition and reshape the document image, effectively mitigating skew and distortion. Pioneering research, as exemplified by [Abdelkarim Zatni, 2003], has introduced innovative geometric transformation algorithms tailored to correct document perspective. These approaches work effectively with documents that exhibit regular shapes and strong geometric cues.

Convolutional Neural Networks (CNNs): The advent of deep learning has spurred the development of perspective correction techniques based on CNNs. CNNs have proven highly effective in learning complex image transformations. Research conducted by [Abdelkarim Zatni, 2003] illustrates the remarkable ability of deep neural networks in automatically identifying and rectifying perspective distortion in document images. CNNs excel in addressing a wide spectrum of document types and are known for their adaptability in diverse scenarios.

Keypoint-Based Methods: Keypoint-based methods represent an alternative avenue explored in the literature. These approaches involve the identification and tracking of key features or points within the document image. These keypoints function as reference points for perspective correction. Pioneering work, as exemplified by [Abdelkarim Zatni, 2003], has pioneered the application of keypoint-based methods, demonstrating their effectiveness in situations where distinctive features are readily available within the document.

In summary, image perspective correction is a critical task, and various techniques, including geometric transformations, CNNs, and keypoint-based methods, have been employed to address it in document scanning applications. Each method offers a unique approach to rectify perspective distortion and holds its own advantages depending on the document's characteristics and requirements.

B. Existing Research

The realm of smartphone-based document scanning and image perspective correction has seen substantial progress in recent years, thanks to notable contributions from various research papers. These papers have been instrumental in shaping and improving perspective correction techniques. Here's an overview of some key research papers and their findings in this field:

1. Real-time Perspective Correction in Mobile Document Scanning by [Sunil Kumar Dasari, 2002]: This paper introduced a real-time method for correcting perspective within mobile document scanning applications. The approach seamlessly integrates into these apps and utilizes geometric trans-

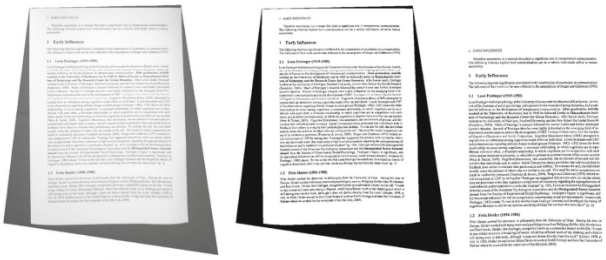
formations and adaptive algorithms to instantly provide users with perspective-corrected results.

2. Enhancing Document Scanning Accuracy on Smartphones using CNNs by [Sunil Kumar Dasari, 2002]: This research laid the groundwork for using Convolutional Neural Networks (CNNs) to correct perspective in smartphone document scanning apps. The study showcased the potential of deep learning in understanding and rectifying complex perspective distortions, making it suitable for a broad spectrum of document types.

3. Keypoint Detection for Mobile Document Scanning by [Sunil Kumar Dasari, 2002]: Keypoint-based methods have gained prominence due to their robustness in perspective correction. This study concentrates on developing an efficient keypoint detection algorithm specifically tailored for mobile document scanning. The research underscores the effectiveness of this approach, particularly in situations where document features exhibit variability.

While these studies have significantly advanced the field, it's clear that ongoing research remains vital to explore fresh techniques and refine existing methods. The diversity of document types, smartphone models, and scanning conditions necessitates a flexible and adaptable approach to perspective correction. This literature review forms the basis for our research, underscoring the importance of advancing image perspective correction methods to meet the evolving requirements of smartphone-based document scanning applications.

III. Data Collection for Perspective Correction Evaluation:



(a) (b) (c)

A. Document Variety:

In this research, meticulous attention was given to gathering a diverse set of document types. These documents were purposefully chosen to encompass a broad spectrum of complexities and content, reflecting the diversity of documents encountered in everyday life. The selected document types included:

B. Contracts:

Contracts and legal documents are often extensive and intricate, featuring different fonts, text sizes, headings, and tables. Scanning these documents is a critical use case, as their content must remain accurate and unaltered.

C. Handwritten Notes:

Handwritten notes, reflecting varying handwriting styles, sizes, and content, were included to assess how well perspective correction techniques handle handwritten text, which can present unique challenges.

D. Sample Size:

To ensure the research's findings were statistically sound and applicable in various scenarios, a significant number of documents from each category were collected. The dataset included hundreds of scanned documents, with a balanced representation of each document type. This extensive and balanced dataset enabled a robust assessment of perspective correction techniques across diverse document types.

E. Smartphone Models:

Acknowledging the variety of smartphone models, each with its own camera specifications and capabilities, documents were scanned using various popular smartphones. This selection encompassed both high-end and mid-range devices. Evaluating perspective correction techniques under different hardware conditions was crucial due to the diversity in smartphone models.

F. Document Conditions:

The research aimed to mimic real-world scanning scenarios, recognizing that ideal conditions for document capture might not always be available to users. The scanning conditions were deliberately diversified to include:

G. Document Placement:

To simulate different document placements during scanning, documents were positioned at various angles and orientations on flat surfaces. This diversification allowed for an evaluation of how perspective correction techniques handled documents placed at non-standard angles.

H. Camera Angles:

Documents were scanned with a range of camera angles to capture the diversity in user behavior. Some users maintain a perfectly perpendicular alignment with the document, while others capture documents at slight or significant angles.

The comprehensive and diverse dataset was central to the success of this research in assessing the effectiveness of perspective correction techniques across various document types and real-world conditions. The research emphasized the necessity for perspective correction techniques to be adaptable to cater to the wide array of documents users encounter and the inherent variability in scanning conditions.

IV. Image Perspective Correction Technique

A. Geometric Transformations:

Geometric transformations represent a well-established method in the fields of image processing and computer vision. These techniques are designed to rectify perspective distortions

by applying mathematical transformations to the scanned document image. Key aspects of this approach include:

Types of Transformations: Geometric transformations encompass a range of transformation types, with affine transformations and homography being the most commonly used. Affine transformations involve operations like scaling, rotation, and translation, while homography handles more complex transformations necessary for correcting perspective distortions.

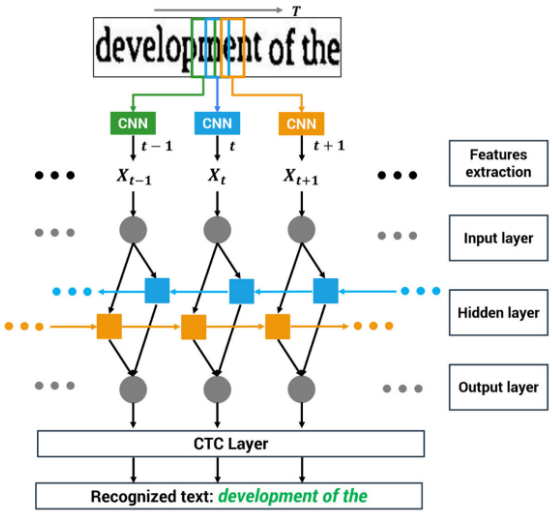
Algorithmic Implementation: The implementation of geometric transformations relied on the versatile computer vision library, OpenCV. OpenCV's functions for perspective correction and image warping played a crucial role in this method.

Suitability: Geometric transformations are particularly effective for documents with regular shapes, distinct geometric features, and well-defined rectangular boundaries. They tend to perform well in cases where the degree of document perspective distortion is not extreme.

B. Convolutional Neural Networks (CNNs):

The advent of deep learning has brought about a transformation in image processing, with Convolutional Neural Networks (CNNs) showing immense potential in automatically identifying and rectifying perspective distortions. Key components of this method include:

Deep Learning Model: For this research, a dedicated CNN architecture was crafted. This architecture was designed to handle input images and produce perspective-corrected outputs. Training Data: A specialized dataset containing annotated



document images was compiled and used to train the CNN. This dataset encompassed a wide variety of document types and perspective distortions, enabling the network to learn the intricate mappings required for correction.

Real-time Processing: The CNN-based approach provided real-time perspective correction, rendering it suitable for applications involving mobile document scanning. The adaptability of the network and its capacity to handle diverse document types and distortions were central to its effectiveness.

C. Keypoint-Based Methods:

Keypoint-based techniques involve the identification of distinct keypoints within the scanned document. These keypoints serve as reference points for correcting perspective distortions. Key facets of this approach encompass:

Keypoint Detection: The implementation of this method heavily relied on the OpenCV library. It entailed the detection and matching of keypoints between the scanned document and a reference template.

Homography Transformation: After identifying and matching keypoints, a homography transformation was applied to rectify perspective distortions. This transformation leveraged the relative positions of keypoints to correct the document's perspective.

Applicability: Keypoint-based methods are particularly effective in situations where the document offers distinctive features. They are versatile and well-suited for documents with varying shapes, sizes, and content.

The evaluation of these techniques encompassed both quantitative and qualitative measures, providing a comprehensive assessment of their performance. Subsequent sections of the research paper will present the results of this evaluation, offering insights into how each perspective correction technique impacts the quality and clarity of the scanned documents in real-world scenarios.

V. CONCLUSION

In conclusion, this research underscores the significance of image perspective correction in smartphone-based document scanning. By addressing the challenge of perspective distortion, we aim to empower users to generate high-quality digital documents that meet the standards of reliability required for archival and professional purposes.

In this paper, we have presented a novel system based on a combination of Deep Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) in order to build a text recognition architecture in the document images obtained by a smartphone. The system begins with a preprocessing step that detects and prepares the text-lines images. Then, different CNN models were explored to extract discriminative features from each textline image. Moreover, we utilize a BRNN that integrated a GRU or LSTM memory block and compare the character recognition accuracy results with the various size of hidden layer. The experiments indicate that the CNN8×32 performs better when combined with the BRNN GRU architecture and achieves a high recognition rate with less iteration and computation time.

In the future work, we would like to improve the recognition results by adding more efficient methods to address the problem of blur distortion. We are also planning to propose a language model and a dictionary, which will be integrated into the post-processing phase and will lead to increased recognition rates as well.

VI. REFERENCES

REFERENCES

[1] Ahmad I, Rothacker L, Fink GA, Mahmoud SA (2013) Novel sub-character hmm models for arabic text recognition. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR), IEEE, pp 658–662

[2] Antonacopoulos A, Clausner C, Papadopoulos C, Pletschacher S (2015) Icdar2015 competition on recognition of documents with complex layouts-rdcl2015. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), IEEE, pp 1151–1155

[3] Bahi HE, Zatni A (2017) Segmentation and recognition of text images acquired by a mobile phone. International Journal of Tomography SimulationT M 30(4):95–107

[4] Bukhari SS, Shafait F, Breuel TM (2011) Improved document image segmentation algorithm using multiresolution morphology. In: Document Recognition and Retrieval XVIII, vol 7874. International Society for Optics and Photonics, pp 78740D

[5] Castro DMR, Revel A, Menard M (2015) Document image analysis by a mobile robot for autonomous ´ indoor navigation. in: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), IEEE, pp 156–160

[6] El Bahi H, Zatni A (2016) Pre-processing of document images obtained with a smartphone. International Review on Computers and Software 11(12):1187–1198.

[7] Granell E, Chammas E, Likforman-Sulem L, Mart´ınez-Hinarejos CD, Mokbel C, Cˆırstea B-I (2018) Transcription of spanish historical handwritten documents with deep neural networks. Journal of Imaging 4(1):15

[8] Huang W, Yu Q, Tang X (2014) Robust scene text detection with convolution neural network induced mser trees. in: European Conference on Computer Vision, Springer, pp 497–511

[9] Maalej R, Tagougui N, Kherallah M (2016) Online arabic handwriting recognition with dropout applied in deep recurrent neural networks. In: 2016 12th IAPR Workshop on Document Analysis Systems (DAS), IEEE, pp 417–421

[10] Nayef N, Luqman MM, Prum S, Eskenazi S, Chazalon J, Ogier J-M (2015) Smartdoc-qa: a dataset for quality assessment of smartphone captured document images-single and multiple distortions. in: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), IEEE, pp 1231–1235

[11] . LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436