International Journal on Document Analysis and Recognition (IJDAR) (2023) 26:175–186
https://doi.org/10.1007/s10032-023-00434-x

SPECIAL ISSUE PAPER

Inv3D: a high-resolution 3D invoice dataset for template-guided single-image document unwarping

Felix Hertlein[1,2] · **Alexander Naumann**[1,2] · **Patrick Philipp**[1]

Received: 15 November 2022 / Revised: 13 February 2023 / Accepted: 12 April 2023 / Published online: 29 April 2023

Abstract
Numerous business workflows involve printed forms, such as invoices or receipts, which are often manually digitalized to persistently search or store the data. As hardware scanners are costly and inflexible, smartphones are increasingly used for digitalization. Here, processing algorithms need to deal with prevailing environmental factors, such as shadows or crumples. Current state-of-the-art approaches learn supervised image dewarping models based on pairs of raw images and rectification meshes. The available results show promising predictive accuracies for dewarping, but generated errors still lead to sub-optimal information retrieval. In this paper, we explore the potential of improving dewarping models using additional, structured information in the form of invoice templates. We provide two core contributions: (1) a novel dataset, referred to as Inv3D, comprising synthetic and real-world high-resolution invoice images with structural templates, rectification meshes, and a multiplicity of per-pixel supervision signals and (2) a novel image dewarping algorithm, which extends the state-of-the-art approach GeoTr to leverage structural templates using attention. Our extensive evaluation includes an implementation of DewarpNet and shows that exploiting structured templates can improve the performance for image dewarping. We report superior performance for the proposed algorithm on our new benchmark for all metrics, including an improved local distortion of 26.1 %. We made our new dataset and all code publicly available at https://felixhertlein.github.io/inv3d.

Keywords Document Unwarping · Dataset · Template · OCR · Transformer

1 Introduction

Numerous business workflows in enterprises involve printed forms, such as invoices, bills, or receipts. The receiving party has to manually digitalize the document in order to persistently access, search, or store the provided data, causing significant personnel costs. Here, existing solutions make use of scanners to create flatbed digital copies of the paper document and apply optical character recognition (OCR) to automatically extract information. This, however, creates additional costs and reduces the flexibility of the given solution.

In order to overcome the hardware restriction, current state-of-the-art approaches attempt to analyze document images taken with smartphones. Prominent examples are DewarpNet [6] and GeoTr [9] which learn to dewarp images using supervised learning, having the available dewarping meshes as ground truth. While these approaches generate promising results, they still are not satisfactorily robust when dealing with environmental factors, such as light incidence, shadows, occlusions, crumpled or folded paper, and perspective transformations.

One potential remedy is the use of additional structured information in the form of templates, which confine the general structure of the documents in order to improve unwarping. While it might be tedious to define initial templates, the added value is significant due to the considerable increase in dewarping precision and robustness.

In this paper, we follow exactly this path and propose a novel labeled invoice dataset with additional structural

✉ Felix Hertlein
hertlein@fzi.de

Alexander Naumann
anaumann@fzi.de

Patrick Philipp
philipp@fzi.de

[1] FZI Research Center for Information Technology, Haid-und-Neu-Str. 10-14, Karlsruhe 76131, Baden-Württemberg, Germany

[2] KIT Karlsruhe Institute of Technology, Kaiserstraße 12, Karlsruhe 76131, Baden-Württemberg, Germany