A Filtering and Image Preparation Approach to Enhance OCR for Fiscal Receipts

Manoela Auad†, Sarah Alves†, Gabriel Kakizaki, Julio C. S. Reis, and Michel M. Silva
Department of Informatics, Universidade Federal de Viçosa - UFV, Viçosa, Brazil.
{manoela.auad*, sarah.c.alves*, gabriel.kakizaki, jreis, michel.m.silva}@ufv.br

***Abstract*—Photographing fiscal receipts has become increasingly common with the rise of online storage and accounting services. However, capturing images in uncontrolled environments often leads to distortions that can compromise Optical Character Recognition (OCR) techniques, turning the output text unreadable. To address this problem, we propose an expert open-source filtering approach based on low-level features to identify and discard poor-quality fiscal images, select high-quality ones, and flag images that need preparation before OCR. The flagged images undergo a series of enhancement techniques, including homography transformation, super-resolution, noise reduction, sharpness adjustment, morphological operations, and binarization. Our extensive experimental evaluation, executed in a new proposed labeled dataset of fiscal receipt, shows that the proposed method lowers the average Character Error Rate metric by up to 11 points compared to baseline methods. Additionally, an ablation study reveals the impact on the accuracy of each image preparation step.**

I. Introduction

In recent years, the demand for document digitization has increased significantly, driven by the ease of misplacing physical documents, the time-consuming process of locating specific information, and the growing digitization of services, including accounting. While humans can read and interpret documents, computers see a scanned document as merely a collection of pixels, each representing a color value at a specific point in the image [1]. Thus, a computer cannot directly interpret textual information within the document. In this context, Optical Character Recognition (OCR) techniques, which recognize characters from documents in digital formats [2], allow these documents to be converted into searchable and editable text. With the rising demand for document data extraction, interest in OCR techniques and their implementation using increasingly advanced technologies is growing. However, OCR still struggles with challenges such as character variations and image quality issues, *e.g.*, noise, low sharpness, and rotated images. These factors affect recognition rates and can lead to incorrect text identification, as shown by Yago *et al*. [3]. Since the system lacks prior knowledge of image quality, it is crucial to provide high-quality images to the OCR system. Proper image processing should normalize and reduce variations in perspective, shapes, and character sizes [4]. Therefore, selecting and processing low-quality images is essential to improve OCR accuracy. Additionally, these images can assume distinct patterns across different domains [3], making the problem more complex and highlighting the need for tailored solutions in applied contexts [5], such as accounting.

Thus, the main goal of this work is to create an expert system to increase the overall accuracy and reliability of data recognition through the OCR system, specifically for fiscal documents. For this, we propose a pipeline divided into two main steps: i) we perform the selection of document images that need to be prepared before feeding them into an OCR system, and ii) we prepare the digitalized fiscal images aiming for better accuracy on OCR systems. The selection task flags the documents into three classes: (1) documents that OCR would have a poor performance; (2) documents that OCR would perform well; and (3) documents that can be prepared by the proposed pipeline for better OCR performance. Documents in group 1 will be discarded, 2 will be fed directly to OCR, and 3 will be prepared before feeding them to the OCR. Our hypothesis is that the OCR performance can be improved by filtering out poor-quality documents and correctly preparing medium-quality images before OCR using the proposed pipeline.

To assess the effectiveness of the proposed method, we extended an existing dataset of fiscal receipts by manually labeling it, followed by an extensive experimental evaluation. The results show that our approach is effective, achieving gains of up to 11 points on the average Character Error Rate (CER) compared to baselines. An ablation study further highlighted the impact to OCR accuracy of each step in the image preparation pipeline, underscoring their potential.

II. Related Work

We compile a review of some efforts concerning the application, development, and enhancement of the OCR system.

Feijó *et al*. [6], for instance, extracted text from hand-photographed images of receipts, based on the idea that even though OCR is a general-purpose algorithm, its accuracy can be improved if it is focused on a specific type of images, such as receipt images. The authors analyzed the impact of a clustering classification model followed by image processing and OCR techniques. The texts generated by OCR were evaluated demonstrating increased recognition accuracy.

Harraj *et al*. [2] point out that the process of image acquisition of a document by digital camera causes several distortions and produces poorly digitized text, leading to unreliable OCR. In order to be able to retrieve information from the document, the authors proposed a pipeline of image enhancement operations, that was tested on datasets of documents in English,

† Authors contributed equally. * Corresponding authors.

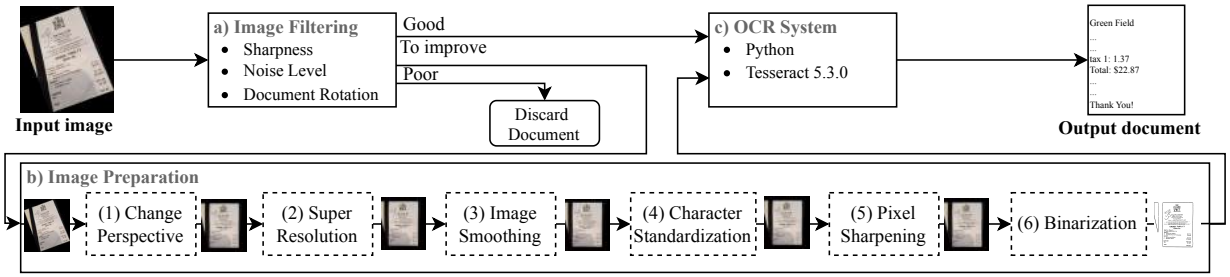

Fig. 1. Overview of the proposed method to improve the OCR result in unconstrained manually captured photos of fiscal documents.

leading to a higher recognition accuracy compared to the original images, in all tested datasets.

Last, Sporici *et al*. [7] proposed a system to identify the best convolution filters that maximize OCR accuracy through a reinforcement learning process. After defining the convolution filters, $10{,}000$ images were pre-processed with the created filters and fed to the OCR software. Results demonstrated that all error metrics decrease when images are pre-processed before being fed to OCR, and accuracy increased $\sim 360\%$. Moreover, even though remarkable results were obtained through OCR engines based on different strategies (*e.g*., neural networks), the application of refined pre-processing methods for noise reduction and image normalization is still necessary [8], [9].

Clustering classification, as performed by the cited works, can be subjective as it is not known what similar characteristics will define each cluster. Thus, this work differs from the others because it applies a filtering step followed by an image preparation pipeline composed of an image super-resolution technique. We also adjust the perspective of the image, filter out pixel noise, normalize characters, and increase the image's sharpness. In our proposed expert approach, the analysis of the images is done in a more objective way, since it is known that imperfections are expected in the fiscal images.

III. Methodology

While scanned images are easier to handle, manually captured images can have characteristics that hinder OCR performance. With the convenience of photographing receipts instead of scanning them, this method has become more common. However, factors like lighting, resolution, noise, blur, and perspective can negatively affect OCR accuracy and reliability. Therefore, our work aims to develop an expert approach to enhance OCR performance specifically for receipt documents captured by unconstrained devices. Specifically, we propose a three-step approach composed of *a)* image filtering, *b)* image preparation, and *c)* OCR application steps to accurately extract information from fiscal documents, as shown in Fig. 1. Each of the steps is discussed in more detail in the following sections.

A. Image Filtering

In order to filter the image, we analyze low-level features, such as sharpness, noise level, and the rotation angle of the document in the image (Fig. 1-a). These features were selected based on the documentation of existing OCR systems [10] and through an experimental evaluation.

The purpose of sharpening analysis is to filter the images based on the focus quality. Since the image was captured by an unconstrained device, *i.e*., a smartphone camera, due to a countless number of factors, the document itself can be out of focus on the image. Thus, we use the *Tenengrad* method to estimate the sharpness of an image. In sum, it calculates the magnitude of the gradient at each point in the image and performs the sum of these magnitudes above a threshold [11]. Regarding the noise analysis, the purpose is to filter out images with high levels of noise. Unconstrained capturing devices can produce noise images due to the quality of the capturing system optics, dirt on the lens, or poor illumination conditions. In this work, we apply the *Estimate Sigma* method[1], to estimate the noise standard deviation based on the image provided as input.

To calculate the rotation angle of the document in the image, we develop a strategy that estimates the document corners present in the image based on the premise that the document will be in the center and occupy the majority of the image. The first step is to identify the document in the image, and for this task, we employ the *Segment Anything Model* (SAM) [12] using the image center point as the key point query for the model. Between the three SAM's returned masks, we select the one which has the highest predicted Intersection Over Union and is larger than 35% of the image area. Since the mask is not guaranteed to form a convex polygon, we find the contours of the mask and calculate the convex hull, for the sake of simplicity, using only the four longest identified contours. Last, we simplify the convex hull output to a quadrilateral using the OpenCV *approxPolyDP* function[2], which returns the document corners. The output of each step can be visualized in Fig. 2. From the inferred corners, we calculate the rotation angle of the document as the angle formed by the line passing through the bottom document corners and the line parallel to the bottom image border.

In addition to these methods, the image quality of the input can be measured by evaluating the brightness or contrast. However, due to the inherent pre-processing of OCR systems in general, it was noted that these characteristics can have a low impact on OCR accuracy (see Sec. V). For this reason, the analysis of these features is not included in the proposed filtering pipeline. To measure the OCR accuracy, we use a method proposed by Rassouni and Harraj [2], which calculates the number of errors present in the output text related to the

[1]https://scikit-image.org/docs/stable/api/skimage.restoration.html#skimage.restoration.estimate_sigma

[2]https://docs.opencv.org/4.x/d3/dc0/group__imgproc__shape.html

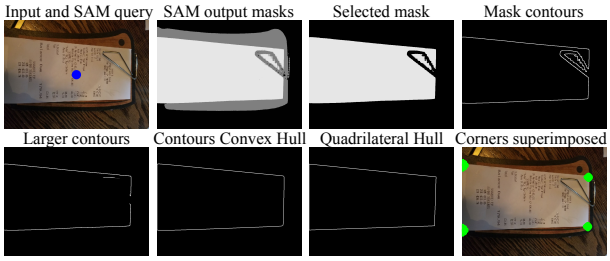

Fig. 2. Output of each step to infer the corners of the document.

total number of characters n in the input text. The number of errors e is measured using the Levenshtein distance concept through the normalized Character Error Rate (CER) between the output text and the correct text: $accuracy = \frac{n-e}{n}$.

After selecting features and defining the accuracy metric, the next step is to define threshold values for each feature to filter the data. For this, we run an analysis of the OCR output in a validation set (20% of the images on the dataset). For each image in the validation set, we fed it to the OCR and calculated the accuracy metric between the output and the ground-truth annotation. We also visually analyze the OCR output and relate its accuracy with the returned text readability, setting the T_{max}. We then calculate the average accuracy of the images in which it was possible to obtain the essential information from the document and defined that a readable OCR output should have accuracy at least 0.85, *i.e.*, $T_{max} \geq 0.85$. Since the scope of the documents analyzed are receipts, the information to be considered essential in the document are the items, the unit prices, and the total price.

In order to set a minimum accuracy T_{min}, we fed all images with an accuracy smaller than T_{max} in a preparation pipeline (described in Sec. III-B) and fed it to the OCR again, calculating the metric. We then relate the accuracy of the OCR with returned texts that were empty, with few characters, or totally illegible. Afterward, we calculated the average accuracy of documents where the output was empty or with only a few random characters, and consider an image as bad in case its accuracy after the preparation image pipeline is less than 0.15, *i.e.*, $T_{min} \leq 0.15$.

From that, we explore the images to extract the thresholds for each group of them. First, we select all the images in the validation set that had at least accuracy greater than T_{max}, and for each of them we calculate the values for sharpness, noise, and text rotation angle. Then, we calculate the average of the values obtained for each feature, obtaining the minimum sharpness value, the maximum noise value, and the maximum rotation value used as parameters to filter the images where OCR would perform well. In the same way, all images that had accuracy up to T_{min} after the image preparation pipeline were selected, and for each of them the values for sharpness, noise, and text rotation angle were calculated. Then, we also calculate the average of the values obtained for each feature within this group of images, obtaining the maximum sharpness value, the minimum noise value, and the minimum rotation value used as parameters to filter the images in which OCR would perform poorly, and therefore, will be discarded.

Last, for each new image, the filtering pipeline calculates the noise level, image sharpness, and rotation angle of the document within the image. From the calculated values, the image is filtered as: (1) "good", which means images has a high probability of resulting in good OCR accuracy; (2) "to be improved", meaning that the image has a chance of resulting in a good OCR accuracy after a pre-processing step; and (3) "poor", which means images with these characteristics at values that result in low OCR accuracy and therefore will be discarded.

B. Image Preparation

To improve the quality of the images filtered as "to improve", we propose a sequence of image processing operations before feeding them to the OCR system, as depicted in Fig. 1-b. The perspective of a document in an image is an important factor for correct recognition considering most of the available OCR systems. Photos manually captured using unconstrained devices often place the document in a virtual plane not coincident with the image plane of the camera. This perspective view can alter the document angles and distorts the characters, negatively interfering with the recognition process. Aiming to address this problem, our first step in image preparation is a change of document perspective to make the virtual plane of the document parallel to the virtual image plane of the camera.

In order to correct the document perspective (Fig. 1-b-1), we need first to identify the document in the image by applying the proposed corner estimator defined in Sec. III-A. Once we identify the corners, we change the document perspective by applying a homography transformation, in which the source points are the corner points and the destination points are calculated so that the proportion between the sides of the document is maintained. To do this, we define X as the difference between the largest horizontal side (of the identified document) and the width of the original image and Y as the difference between the largest vertical side and the height of the original image. For $X' = X/2$ and $Y' = Y/2$, the destination points of the transformation are (X', Y') incremented by the largest horizontal side and the largest vertical side of the document. In sum, this means that the four corners of the document lie in the same plane and form a rectangle and that the plane of the document is parallel to the image plane. In case any of the four corners of the document are undefined on the image, to avoid miss transformations, we do not perform the homography transformation.

Since we are dealing with unconstrained capturing devices, the resolution of the document can negatively impact the recognition accuracy. To address this problem, we propose to use a super-resolution technique on the original image (Fig. 1-b-2). In this step, we apply a *Residual Dense Network* [13] at the beginning of processing in order to improve the effectiveness of image treatments that are applied later, as well as increase the overall recognition accuracy.

Another feature negatively highly correlated to the OCR accuracy is noise. In our image preparation pipeline, we address the noise problem by convoluting the image with the

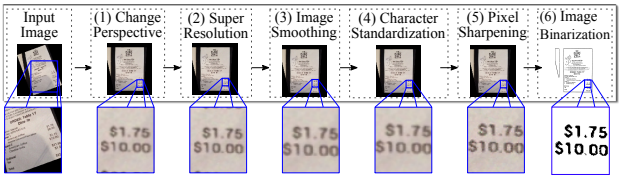

Fig. 3. Detailed view of the output for each step of the Image Preparation.

median filter, of size 3×3 (Fig. 1-b-3). In comparison with linear smoothing filters of similar size, median filters offer random noise reduction with significantly lower blurring [1].

Thin characters or serif characters are undesired features since they may negatively interfere with the recognition of text details [10]. Since these undesired features for OCR are commonly seen in printed receipts, we performed the Character Standardization by applying morphological operation Dilation, aiming to enhance the characters in the image to improve recognition (Fig. 1-b-4). Applying dilation can also bridge the gaps in characters [1], which enhances its structure. We use a square-3-pixel structuring element of zeros for the dilatation operation. In general, documents have a light background and dark text. Therefore, the application of this structuring element with dark pixels causes the text to be a dilated object and generates an expanding effect on the characters.

Regarding character recognition, simple details can lead to undesired results. There are characters that differ from each other by only one stroke. Therefore, if the input image is out of focus or somehow blurred, this stroke could not be identified, leading the OCR to a wrong result. In order to visually enhance image details and textures [2], we apply a sharpening adjustment (Fig. 1-b-5) by performing a convolution operation of the image with a ω filter, of dimension 3×3, whose coefficients should emphasize the difference between neighboring pixels [14], defined as: $\omega = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & 5 & -1 \end{bmatrix}$. Since the image sharpening process enhances the gradients in the image, we apply this step after the Image Smoothing step. Otherwise it can worsen the effects of noise already present in the image, as noise creates gradients on the image.

Finally, we apply document binarization since document images may have information that is unnecessary for processing and harmful to the recognition, *e.g.*, watermarks (Fig. 1-b-6). Although the applied OCR system has built-in methods for this task, we enforce a different adaptive binarization algorithm, as suggested in [10], to separate the relevant information of the image, *i.e.*, the text, from the dispensable or redundant information. The operation order applied on the proposed Image Preparation was defined based on the priors presented here and on a deep experimental evaluation. The output of each step up to here can be visualized in Fig. 3.

C. OCR System

As afore-presented in Fig. 1-c, for each input image, we fed it to the OCR case the image is filtered as "Good" or "To improve". In the first case, the image is forwarded directly to the OCR system. Otherwise, the input image is prepared for the OCR system, and the output of this step is fed to the OCR. Images classified as "Poor" are not fed to the OCR.

IV. Experiments

In this section, we describe the data, metrics, and tools used in this work, as well as the experimental evaluation setup.

A. Datasets

The dataset used in this work is an extension of the *Express Expense SRD* dataset[3], which comprises 200 hand-photographed or scanned images of restaurant receipts. The free version of the original dataset has no OCR task labels. Since, this information is required to calculate the accuracy of OCR and analyze the effects of the proposed approach, we created a new version of the existent dataset with manual annotations (*i.e.*, texts in English) for the receipts and also the four corners of the documents. The corners are relevant to check the accuracy of the proposed corners estimator. As annotations in this context are not subjective, each image was labeled by a single annotator and checked by another one. This dataset is public here: https://zenodo.org/records/13688441.

Alongside the created annotations, we also propose image perturbations for the original images aiming to simulate or amplify controversial effects related to unconstrained capturing devices. From the 200 images, we randomly split it in 20-80% set respectively for validation and test. Then, for each image, we create five new images applying a combination of image perturbations following described, resulting in 1,200 images with annotations related to the text in the images and the document's corners. The applied image perturbations are:
a) Rotation: with an angle randomly chosen from $0°$, $5°$, $10°$, $15°$, and $20°$, either clockwise or counterclockwise.
b) Noise: Gaussian-distributed white additive noise with variance randomly selected from the values 0.001, 0.003, or 0.005.
c) Sharpness: to decrease the sharpness of the image, we apply a square-2D Gaussian blur filter of size randomly chosen between 3, 5 or 7.
d) Resolution: To downscale the image, we use the image pyramid concept[4]. We randomly chose to downscale the image by going up on the pyramid 0, 1, or 2 levels, which means the resolution goes down up to two halves of the original image resolution. All values for the image perturbations were empirically set in a range to simulate a real-world scenario.

B. Metrics

For recognition accuracy, we explore the following metrics:
a) Character Error Rate (CER) is applied on non-empty OCR results to measure characters recognized incorrectly based on the number of character editions and corrected inferred characters C. The editions are measured by counting the: insertions I, substitutions S, and deletions D in the input text to match the ground-truth label, as follows: $CER = \frac{S+D+I}{S+D+C}$.
b) Word Error Rate (WER) has the same input, interpretation, and equation of CER, however, the editions are calculated in the level of words and not characters.

[3]https://expressexpense.com/blog/free-receipt-images-ocr-machine-learning-dataset/.

[4]https://docs.opencv.org/3.4/d4/d1f/tutorial_pyramids.html

c) Binary Result (BR) is applied to identify if the OCR result is empty (returning 1), meaning that the OCR system was not able to recognize a single character, or not (returning 0).
d) Binary Accuracy is applied to identify if the OCR result is different from the ground truth considering a margin. It returns 1 case CER > 0.05, or 0 otherwise.

For all metrics, smaller values are better.

C. Experimental Setup

We evaluated the impact of each image feature cited in Sec. III-A on the OCR result. Using the test split, for each image, we apply distortions varying the parameters as depicted in Tab. I. The analysis of the results of this experiment can lead to the definition of the features used in the filtering process.

Regarding the proposed strategy to infer the corners of the document, we applied it to all validation images and calculate the Euclidean distance from the inferred points to the annotated corners proportionally to the image size.

Aiming to evaluate the effect of Image Filtering and Image Preparation steps on the OCR overall accuracy, we compared the OCR metrics by applying the proposed methods on various arrangements. First, in the "Naïve" method, we do not filter or prepare the images before feeding them to the OCR. For "Only Filtering", we only apply the filtering process. In the "Adjust all" method, we do not apply the filtering process and prepare all images before feeding them to the OCR System. At last, we apply "Ours". We performed this experiment in both original and perturbed test split. Results of this experiment are presented in Tab. II.

Finally, we propose an ablation study to explore the contribution of each step of the Image Preparation pipeline, executing the Image Filtering process on the perturbed test split and selected only the images filtered as "to be improved". Then, we executed the Image Preparation activating different steps of the process, fed the prepared images to the OCR, and calculated the metrics (Tab. III). We performed the experimental evaluation using Tesseract v.5.3.0[5] as an OCR system [15], as it is a multilingual and open-source tool and has been widely explored in the literature through work in similar contexts [16][6]. The codes implemented is this study are available at the following link: https://github.com/MaVILab-UFV/Filtering-Preparation-for-OCR_SIBGRAPI-2024.

V. Results

Results presented in all tables contain the average $\pm$ standard deviation across all images on the dataset for each metric. Tab. I values confirm the statements in Sec. III-A that Brightness and Contrast changes have less impact on the overall OCR results when compared with Original Data results (see line 1). Therefore, we did not consider those features in the filtering process. All other image perturbations led to an expressive negative variation on the OCR results, justifying the selected features for the filtering process. We highlight the impact of the rotation, even with an angle small as $5°$ we observe an increase of 25 points, while an angle of $20°$ completely vanished the recognition. This result motivates the usage of homography transformation to correct the document perspective and angle. The presence of noise in the image also impacts the final accuracy, motivating the use of the median filter in the Image Preparation process. A low definition and low OCR accuracy are correlated, thus, we include a sharpening process. Finally, the Image Pyramid Scale demonstrated a substantial burden on the metrics, sustaining the application of a Super-Resolution step in the Image Preparation step. As the results were consistent across all metrics, in this analysis we did not delve into any specifics in detail.

TABLE I
IMPACT OF THE IMAGE PERTURBATION ON THE OCR ACCURACY. PYR. SC. STANDS FOR PYRAMID SCALE DOWN.

Data		Metrics WER	CER	BR	BA
Original		54.3 ± 26.5	40.8 ± 28.1	2.6 ± 16.0	94.8 ± 22.3
Rotate	+5	77.6 ± 21	65.6 ± 28	14.1 ± 35	100 ± 0
	+10	97.5 ± 6	93.4 ± 11	52.4 ± 50	100 ± 0
	+15	99.5 ± 4	98.2 ± 6	78.5 ± 41	100 ± 0
	+20	99.8 ± 3	99.2 ± 4	85.9 ± 35	100 ± 0
	−5	75.6 ± 24	64.3 ± 30	16.2 ± 37	98.4 ± 12
	−10	96.3 ± 7	91.9 ± 14	52.9 ± 50	100 ± 0
	−15	99.5 ± 2	98.3 ± 5	82.2 ± 38	100 ± 0
	−20	99.9 ± 1	99.6 ± 2	93.2 ± 25	100 ± 0
Brightness	+10	54.2 ± 26	40.7 ± 28	2.6 ± 16	94.2 ± 23
	+30	54.6 ± 27	41.2 ± 28	2.6 ± 16	94.8 ± 22
	+50	55.0 ± 27.0	41.5 ± 29	3.1 ± 17	95.3 ± 21
	−10	54.0 ± 26	40.5 ± 28	2.1 ± 14	94.8 ± 22
	−30	53.1 ± 27	39.6 ± 28	2.6 ± 16	93.7 ± 24
	−50	$\mathbf{52.6 \pm 27}$	$\mathbf{39.2 \pm 27}$	3.1 ± 17	$\mathbf{93.2 \pm 25}$
Contrast	1.25	54.8 ± 27	41.7 ± 29	3.1 ± 17	95.3 ± 21
	1.5	58.5 ± 27	45.1 ± 29	6.8 ± 25	96.3 ± 19
	1.75	63.3 ± 27	49.1 ± 31	8.4 ± 28	97.4 ± 16
	0.75	54.2 ± 27	41.0 ± 28	2.6 ± 16	95.3 ± 21
	0.5	54.5 ± 26	40.8 ± 28	$\mathbf{2.1 \pm 14}$	94.2 ± 23
	0.25	55.0 ± 26	41.4 ± 28	2.6 ± 16	95.8 ± 20
Noise	0.5	68.7 ± 27	57.4 ± 31	11.5 ± 32	97.4 ± 16
	1.0	80.0 ± 25	72.2 ± 30	27.2 ± 45	98.4 ± 12
	1.5	87.4 ± 20	81.8 ± 25	45.5 ± 50	99.5 ± 7
Blur	3	62.4 ± 29	52.5 ± 33	10.5 ± 31	95.3 ± 21
	5	71.0 ± 28	62.5 ± 33	18.8 ± 39	97.4 ± 14
	7	80.0 ± 25	72.8 ± 31	30.9 ± 46	99.5 ± 7
Pyr. Sc.	lv.1	81.1 ± 24	73.5 ± 30	30.9 ± 46	99.0 ± 10
	lv.2	98.7 ± 8	97.8 ± 10	84.8 ± 36	100 ± 0

The average and standard deviation regarding the displacement of the inferred document's corner related to the manually annotated points on the validation test was 4.2 ± 11.6 of the larger side of the image. It means that the inferred points are expected to be located no further from the correct point. Such a small displacement lead to a senseless impact on the homography transformation.

We observe by Tab. II values that, as expected, "Naïve" method presented high values for CER and WER, and a high rate of empty results represented by BR. The "Only Filtering" method scored the best CER and WER since only the images filtered as good will be processed on the OCR system. However, it presents the highest value for BR and also a high number of incorrectly recognized text (BA value), since all images, but the good ones, will have no recognized text. The "Adjust All" method presents the worst CER and WER

[5]https://github.com/tesseract-ocr/tesseract
[6]https://nanonets.com/blog/ocr-with-tesseract/

TABLE II
EVALUATION OF THE PROPOSED METHOD COMPARED TO BASELINES.

Method	Metrics				
	WER	**CER**	**BR**	**BA**	**Time**
Naïve	50.6 ± 24	36.9 ± 25	4.0 ± 20	95.8 ± 20	$\mathbf{0.1 \pm 0}$
Only F.	$\mathbf{41.2 \pm 23}$	$\mathbf{26.7 \pm 20}$	68.7 ± 46	97.9 ± 15	$\mathbf{0.1 \pm 0}$
Adj. all	62.1 ± 38	34.7 ± 25	$\mathbf{1.3 \pm 11}$	$\mathbf{93.9 \pm 24}$	80.3 ± 44
Ours	56.1 ± 37	33.6 ± 25	18.7 ± 39	95.9 ± 20	39.3 ± 49
Naïve	67.6 ± 28	56.5 ± 32	67.1 ± 47	97.6 ± 15	$\mathbf{0.1 \pm 0}$
Only F.	$\mathbf{42.9 \pm 23}$	$\mathbf{28.5 \pm 22}$	94.3 ± 23	98.0 ± 14.0	$\mathbf{0.1 \pm 0}$
Adj. all	88.3 ± 39	60.5 ± 30	$\mathbf{25.3 \pm 43}$	97.9 ± 14	56.0 ± 59
Ours	67.4 ± 39	45.5 ± 31	77.7 ± 42	$\mathbf{97.0 \pm 17}$	11.3 ± 32

Above dashed line: original dataset || Below dashed line: distorted dataset

values, but it is noteworthy that it also achieves the lowest BR and BA values. The decrease in the overall accuracy is explained due to, many documents that did not produce any results before the image preparation, produced some output. Compared to the "Naïve" method, the number of images that present a result is almost double. It is noteworthy the Super Resolution step is time-consuming (470ms per image), and with previous results, we conclude that this step is not necessary for images filtered as good. Ours achieved the best CER and WER compared to "Only Filtering" and "Adjust all" methods, with the benefit of not processing all images, which considerably reduces the processing time (less than half of the "Adjust All"). We also highlight the reduction of 11 points in CER when compared to "Naïve", and also the BR value indicates that the filter is removing images that would produce some result. Since we crave the recognition to be as good as possible, a higher BR does not implicate a worse result.

Ablation Study. In Tab. III, the acronyms stand for the image preparation steps depicted in Fig. 3, and the results suggest that each step guarantee that more images will generate results, as we see the decrease in the BR metric. We highlight that the complete method compared to images without processing led to an improvement of 22 points in BR while preserving the same CER. Last, we notice that Super Resolution (SR) increases the image size and, together with Homography Transformation (HM), which places the document in the correct perspective for the OCR system, have the highest influence on the number of documents with output.

VI. CONCLUSION

As demonstrated by prior works, the image quality and attributes such as rotation or perspective directly impacts on the accuracy of OCR system. In this work, we present an expert approach to enhancing the recognition accuracy of OCR systems when dealing with hand-photographed fiscal documents. Our proposal includes filtering and image preparation steps designed to eliminate low-quality images and enhance features beneficial to recognition systems. Experimental results using specific metrics reveal that our filtering step effectively eliminates poor-quality images, a significant outcome for on-demand services where OCR systems charge per request. This approach can automatically determine if an image is suitable for processing, prompting users to upload a new version if necessary. Additionally, the filtering step identifies images that need preparation before OCR processing. The results of our ablation study confirm that the proposed image preparation step leads to higher recognition accuracy and an increased number of successfully processed images. This step is crucial for on-demand services as it reduces the need to re-upload files by enhancing image quality. Overall, our approach ensures fewer unnecessary OCR requests and fewer discarded images while maintaining high recognition accuracy. As future work, we plan to evaluate the impact on the recognition rate of applying more complex methods, such as transformers or CNN models, to each step of the image preparation process, as well as investigating the proposed approach in different OCR tools (*e.g.*, Google Drive OCR and Easy OCR).

TABLE III
ABLATION STUDY REGARDING THE IMAGE PREPARATION PROCESS.

Proposed Method Steps					Metrics			
HM	**SR**	**SM**	**MP**	**SH**	**WER**	**CER**	**BR**	**BA**
✗	✗	✗	✗	✗	62.3 ± 28	51.2 ± 31	50.7 ± 50	95.1 ± 22
✗	✗	✗	✗	✓	80.2 ± 21	55.5 ± 28	44.0 ± 50	99.1 ± 9
✗	✗	✗	✓	✓	79.0 ± 22	56.9 ± 29	43.5 ± 50	98.3 ± 13
✗	✗	✓	✓	✓	74.2 ± 28	55.3 ± 31	43.5 ± 50	97.5 ± 16
✗	✓	✓	✓	✓	79.7 ± 39	53.1 ± 31	35.9 ± 48	98.5 ± 12
✓	✓	✓	✓	✓	75.8 ± 40	51.3 ± 32	28.2 ± 45	96.7 ± 18

ACKNOWLEDGMENT

The authors are grateful to agencies CAPES, FAPEMIG, and CNPq for funding different parts of this work.

REFERENCES

[1] R. C. Gonzalez and R. E. Woods, *Digital image processing*. Pearson Education Ltd., 2018.
[2] A. E. Harraj and N. Raissouni, "Ocr accuracy improvement on document images through a novel pre-processing approach," *ArXiv*, 2015.
[3] Y. Santos, M. Silva, and J. C. Reis, "Evaluation of optical character recognition (ocr) systems dealing with misinformation in portuguese," in *SIBGRAPI*, 2023, pp. 223–228.
[4] N. Arica and F. T. Yarman-Vural, "An overview of character recognition focused on off-line handwriting," *Trans. Syst., Man, and Cyb.*, vol. 31, no. 2, pp. 216–233, 2001.
[5] B. G. Buchanan and R. G. Smith, "Fundamentals of expert systems," *Annual review of computer science*, vol. 3, no. 1, pp. 23–58, 1988.
[6] J. V. F. d. A. Feijó *et al.*, "Análise e classificação de imagens para aplicação de ocr em cupons fiscais (in portuguese)," Master's thesis, Universidade Federal de Santa Catarina, Florianópolis, SC., 2017.
[7] D. Sporici, E. Cușnir, and C.-A. Boiangiu, "Improving the accuracy of tesseract 4.0 ocr engine using convolution-based preprocessing," *Symmetry*, vol. 12, no. 5, p. 715, 2020.
[8] M. Ganis, C. L. Wilson, and J. L. Blue, "Neural network-based systems for handprint ocr applications," *TIP*, vol. 7, no. 8, pp. 1097–1112, 1998.
[9] M. Koistinen, K. Kettunen, and J. Kervinen, "How to improve optical character recognition of historical finnish newspapers using open source tesseract ocr engine," *Proc. of LTC*, pp. 279–283, 2017.
[10] "Tesseract documentation: Improving the quality of the output," https://tesseract-ocr.github.io/tessdoc/ImproveQuality.html, 2023.
[11] L. G. Barros *et al.*, "Impacto da analise da nitidez em metodos de classificacao de imagens de madeira (in portuguese)," Master's thesis, Universidade Estadual de Ponta Grossa, 2013.
[12] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, *et al.*, "Segment anything," *ArXiv*, 2023.
[13] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *CVPR*, 2018, pp. 2472–2481.
[14] A. Polesel, G. Ramponi, and V. J. Mathews, "Image enhancement via adaptive unsharp masking," *TIP*, vol. 9, no. 3, pp. 505–510, 2000.
[15] R. Smith, "An overview of the tesseract ocr engine," in *ICDAR*, 2007.
[16] B. A. Dangiwa and S. S. Kumar, "A business card reader application for ios devices based on tesseract," in *ICSPIS*, 2018, pp. 1–4.