International Journal of Computing and Optimization
Vol. 9, 2022, no. 1, 13 - 32
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ijco.2022.9829

Tesseract OCR Evaluation on Greek Food Menus Datasets

Alkiviadis Tsimpiris

Department of Computer, Informatics and Telecommunications Engineering
International Hellenic University, 62124, Serres, Greece

Dimitrios Varsamis

Department of Computer, Informatics and Telecommunications Engineering
International Hellenic University, 62124, Serres, Greece

Georgios Pavlidis

ATHENA - Research and Innovation Centre in Information,
Communication and Knowledge Technologies
University Campus at Kimmeria, 67100, Xanthi, Greece

Abstract

This article presents a procedure for optical character recognition (OCR) improvement, after image preprocessing of Greek food menus images. To achieve this goal, many well-known and other more sophisticated techniques for image preprocessing have been used. The performance of the Tesseract OCR engine has been studied for selected binarization, thresholding, noise and morphological filtering methods that applied to menu images before OCR feeding. The output text is compared to the reference text of each image (ground text) and the values of evaluation indices indicate the appropriate preprocessing method. Datasets of Greek food menu images with their respective ground text files, were generated for first time in this study, due to the lack of alternative datasets in any language. OCR outputs and ground texts were evaluated using error rate and accuracy on character and word levels. The results of OCR application on Greek menu images showed high accuracy values in high scanning resolution photos and in cases of menus with distinct and visible fonts.

Keywords: ocr, image preprocessing, food menu images

1 Introduction

The optical character recognition (OCR) problem relates with the identification of text in images. There is a rich bibliography on this challenge and various text detection approaches have been developed that successfully identify texts of interest in complex images [14]. Recently, this research was expanded to accommodate the needs of food and meal applications by recognising text in menu images [27], [16], [28].

This work deals with text detection in food menu photos that scanned or captured with mobile phone cameras. This can include images in different lighting conditions and possibly with arbitrary orientation and in many cases involves combinations of graphics, photographs, printed and handwritten texts. The project team conducted a series of experiments to design an image preprocessing strategy that will power an OCR system like the Tesseract engine [26].

A number of filtering techniques have been compared, including several thresholding approaches, morphological operations, global and adaptive gradient approaches, Gaussian techniques and Wiener filtering [19], [11], [9], [18], [21], [29].

The objective of our work is to organize, apply, evaluate and improve an OCR in image datasets with Greek texts, focused in images from Greek food menus, because the literature and scientific works in recognising Greek texts is very poor. We have studied the performance of different image filters and their ability to improve the performance of the OCR engine, using OpenCV functions in Python. Two datasets with Greek menu images were also generated, one with scanned food menus and a second one with menus captured by smart phones camera.

Rest of the paper includes related work, image preprocessing with filters and methods, ocr configuration and ocr evaluation indices, the datasets that were used, the results and discussion and finally the conclusions and future work.

2 Related Work

In a recent article of [28], a mix of old and modern innovative techniques was proposed for exporting and recognizing food names as well as their prices from grocery receipts in a easy way. Interesting applications have been also developed to extract information from a receipt [1], helping people to earn points for every receipt and to organize their daily purchases [4], [5].

Summaries of statistics and graphs, by groups of food consumed in the week or month and those that remain to be consumed, are provided via a food monitoring application by scanning the grocery receipts [24]. Image preprocessing, deskewing, binarization techniques etc for better results usually are applied on images with text and graphics before input to ocr systems [17], [2].

Training of a variety of OCR models with deep neural networks is presented in the work of [7], drives to an optimal deep neural networks for their data and, with additional Finnish and Swedish training sets, they achieved successfully train high-quality mixed-language models.

The most relevant research to our work is a successful menu recognition effort in Indonesian restaurants [27] where an Android-based application that presents gastronomic information from Indonesian restaurant menu images,was developed, trying to make more accessible and popular with foreigners the Indonesian gastronomy. Their application captures the restaurant menu images through a camera sensor, then pre-processes and recognizes the actual text using Convolutional Neural Networks (CNN).

The recognized text corresponds to the terms recorded in a predefined gastronomic database to present information about the selected food. This application was able to recognize 100% of the menus that used the Sans Serif font. The accuracy rate was reduced to 56% when the menus used the Times New Roman font in Indonesian. In their research, an Android application use raw images from restaurant menus taken by the camera of smart mobile devices. Each image was subjected to noise removal by grayscaling and binarization. The image preprocessing steps that used in their work, were also studied in our work and enriched with more methods and techniques.

Text recognition from Greek restaurants menu lists is an area that needs further research and application development. One of the reasons that the selection of local Greek delicacies by foreign visitors is less popular stems from the fact that the information about the food or the local delicacies is unknown. Reading the name of the food in the restaurant menus is usually difficult even with the use of popular applications such as Google Lens and does not provide additional information about the food, the ingredients, how it is prepared or a picture of the food itself. According to the above reasons, there is an urgent need for text recognition from restaurant menus and development of corresponding applications. To the best of our knowledge there are few works of Greek text recognition. GRPOLY-DB [8] is the first publicly available old Greek polytonic database, for the evaluation of several document image processing tasks. It contains both machine-printed and handwritten documents as well as annotation with ground-truth information. A complete database system for storing images of Greek unconstrained handwritten characters is GCDB [15].

3 Process Description

The methodology followed in this work is described in Figure1. According to this diagram, collections of Greek food menus images with the corresponding ground text file were used. Image preprocessing techniques and filters were initially applied to these images before forwarding them to the OCR machine in order to investigate an appropriate filter or combination that would improve the performance of the OCR machine. Sixteen popular and efficient filters and thresholding techniques were examined. The OCR output text of every preprocessed image compared with the according ground text using indicators like word error rate, character error rate, cosine and Jaccard.

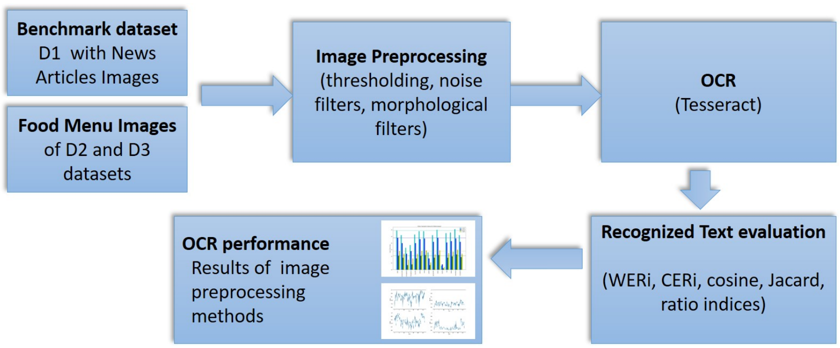

Figure 1: The workflow of OCR evaluation after food menu images preprocessing

3.1 Image Processing

To improve the text output quality from an OCR, there is the need of image pre-processing so that the text can easily distinguished from the background before it is fed to the OCR engine. The OCR systems are using by default binarization processes to convert coloured images to black and white images. Noisy background, blurred images or images captured by mobile devices need threshold methods like binarization to give assistance at OCR systems [10]. Traditional threshold and binarization methods were applied at this study, on restoraunt menu images, before OCR and the results were compared in order to select an appropriate method that improves OCR efficiency. The OpenCV library functions were used for all image preprocessing steps that adopted at this work [3].

3.1.1 Threshold and binarization

The methods adopted in this study are briefly described below:

Simple thresholding: The main idea of thresholding is simple and effective, finding an appropriate threshold and then convert to white (0) all pixels of the image with values above this threshold and rest of the pixels to black(255) [25].

Adaptive thresholding: This method calculates the threshold for small or local areas of the image giving better results for images with different type of lighting in different areas of the same image [12]. Two alternatives of this method were used at this study the a) ADAPTIVE THRESH MEAN when the threshold value is the average pixel intensity in the neighbourhood of an area and b) ADAPTIVE THRESH GAUSSIAN. According this method the neighbourhood pixel values are multiplied by the weights obtained from a Gaussian distribution and the definition of the block size (neighbourhood area) is also needed.

Binarization Otsu: This method is used to perform automatic image thresholding. In the simplest form, the algorithm returns a single intensity threshold that separate pixels into two classes, foreground and background calculating the threshold value from images histogram [19].

3.1.2 Image Filters

The meaning of image noise is the random variation of color or brightness between the pixels of an image. Undesirable image distortions can occur during the process of image creation, storage, transmission or editing a digital image. The result of image noise, drives to an image of bad quality and low readability. Observed distortions may be due to various types of noise from different sources such as sensors, devices or processing. The following well-known and efficient noise removal techniques used additionally at this study before OCR application [20]:

Open and Close filters: The impulse noise that usually called as "salt" and "pepper" is the type of noise that is observed with randomly distributed white and black pixels on an image, looking like salt and pepper dropped on a gray paper. The goal of the "open" filter function is the removal of "false positive" pixels and for the "closing" filter is the elimination of "false negatives" pixels.

Median filter: One of the simplest and popular linear filters is the median filter that is used to smooth images and reduce the observed image noise. The median filter replace the brightness of a pixel with the median brightness of the pixel's neighborhood. This results in reduced variability locally on each pixel and therefore a blurry image. The median filter works efficiently if the impulse noise density is low ($< 20\%$).

Gaussian filter: The cumulative noise that observed in the images is usually described as Gaussian noise. Gaussian filters are low-pass filters modifying the input signal by convolution with the sampled Gaussian kernel that is produced by sampling points from the continuous Gaussian.

Erosion filter: The basic goal of erosion filter is to erode away the boundaries of regions of foreground pixels leading the areas of foreground pixels to shrink in size, and holes within those areas become larger. It is useful to erode white noises of image foreground for example to disconnect two connected objects that represent two connected letters.

Dilation filter: The basic effect of dilation filter is exactly the opposite of erosion. The boundaries of regions of foreground pixels gradually enlarge leading the areas of foreground pixels to increase in size, and holes within those areas become smaller. The application of this filter to image pre-processing it is useful for joining broken parts of an object.

Gardient filter: The morphological gradient filter is the difference between dilation and erosion of an image and explains how the newly created white pixels are distributed.

Histogram Equalization filter: Increasing the color contrast of an image makes easier the discrimination of neighboring areas. Histogram equalization is a method to process images in order to adjust the contrast of an image by modifying the intensity distribution of the histogram. The objective of this technique is to give a linear trend to the cumulative probability function associated to the image.

CLAHE filter: The CLAHE (Contrast Limited Adaptive Histogram Equalization) technique was originally developed by [Zuiderveld, 94] and [Pisano, 98] for medical data imaging and has been shown to be successful in enhancing low contrast images. The CLAHE algorithm separates images into individual regions and applies histogram balancing to each of them. This equates the distribution of gray values and thus makes the hidden features of the image more visible. The full spectrum gray is used to express the image. There is a variety of customized histogram balanced contrast (CLAHE) techniques available. Selective amplification is achieved by first detecting the edge of a field in an image and then processing only those regions of the image located within the edge of the field.

conv2dkernel filter: This filter performs Fourier-based convolution of an image file using the provided 2D kernel. At this study the function *conv2dkernel* computes the gradient of an image by 2D convolution with a [5x5] kernel giving as output an image with the same size [18].

Wiener-Hunt filters: Wiener filtering is a widely used technique in both signal processing and image processing. In both cases, the Wiener filter is a filter used to produce an estimation of a desired or stochastic random process, with linear time-invariant filtering of an observed noisy processing. The statistical parameters used are known and constant, and are as follows: the spectral power density of the signal or image, the spectral power density of the noise

Table 1: Table with the filters and the thresholding methods that applied on images before ocr feeding

notation	description
f_1	no filter
f_2	clahe filter after Otsu thresholding
f_3	adaptive thresholding with mean pixel intensity
f_4	adaptive thresholding with Gaussian distribution
f_5	binarization with Otsu thresholding
f_6	opening filter
f_7	closing filter
f_8	gradient morph filter
f_9	gauss blur filter
f_{10}	median blur filter
f_{11}	histogram equalization filter
f_{12}	clahe filter
f_{13}	dilation morph filter
f_{14}	erosion morph filter
f_{15}	conv2dkernel filter

and the characteristics of the additive noise. The Wiener filter minimizes the mean square error between the estimated random and the desired process [18].

The notation of the filters and the thresholding methods is presented at Table 1, where the Wiener filter is not included due to the poor performance in this application.

3.1.3 Image Resizing

It is observed that low resolution images (lower than 200DPI) will give poor OCR results while images with resolution above 600DPI will unnecessarily increase the processing time without improving the OCR output. The OCR engine that used at this study works best on images which have at least 300 dpi, so the images were scaled to the right size in cases of high resolution.

3.2 TESSERACT-OCR

Tesseract is an open source Optical Character Recognition (OCR) Engine or API, available under the Apache 2.0 license appropriate for extraction text from images of different formats [26]. It is available for Linux or Windows operating systems and is supported by many programming languages. In this study Windows 10, Python 3.7.3, openCV 4.1 and Tesseract 4.0, on an i7 Intel processor with 16GB Ram. Tesseract OCR by default use Otsus binarization as an initial step for image preprocessing, supports optical character recognition for Greek and other languages and also use a line finding algorithm so that a skewed page can be recognized without having to de-skew, thus saving loss of image quality. Moreover Tesseract handle pages with curved baselines, measuring gaps in a limited vertical range between the baseline and mean line. Tesseract 4.0 added a new OCR engine based on LSTM neural network. Although considered the best open source OCR engine, a well designed image preprocessing is able to increase significantly the accuracy of this engine.

3.3 OCR Evaluation

The comparison between ground text A related to a food menu image and the OCR output text B is the procedure for OCR's efficiency evaluation. There are many text similarity indices at character level or word level. At this work the Word Error Rate(WER), the Character Error Rate(CER), the $cosine$, the $Jaccard$ and the $ratio$ indicators were used.

The Word Error Rate (WER) is based on the Levehnstein distance between B and A [13] according the following equation:

$$WER = \frac{(iw + sw + w)}{nw} \tag{1}$$

where iw, sw and dw refer to the number of words that need to be inserted, replaced, and deleted to transform the exported text B into the reference text A. The total count of these words is normalized by the number of words in the reference text A (nw). In this study the number of new words in the extracted text was ignored ($iw = 0$), transforming this index to the classical error rate index.

$$WERi = H/nw \tag{2}$$

where $H = sw + dw$ The according accuracy index is formed as follows:

$$WAcci = Accuracy = 1 - WERi \tag{3}$$

In this work the $WAcci$ index is used (ranged between 0 and 1), without taking into account the new words of the extracted text B ($iw = 0$) and is independent of words position in the text.

At character level, the character error rate (CER) is formed as follows:

$$CER = \frac{(ic + sc + dc)}{nc} \tag{4}$$

where ic, sc and dc refer to the number of characters that need to be inserted, replaced, and deleted to transform the exported text into the ground text and nc refers to the total characters of ground text A.

The same assumption of $WERi$ index, in relation with the number of ignored inserted words, adopted at $CERi$ index as well for the inserted characters that also ignored, setting the parameter $ic = 0$. The equivalent index of characters accuracy $CAcci$ is also calculated.

$$CAcci = 1 - CERi \tag{5}$$

The word-level error rate is usually higher than the character-level error rate, as failures in recognizing individual characters significantly affect a word recognition. In many OCR application the correct words identification is much more important than correctly identifying single characters, numbers, or punctuation. It is defined that a word is any sequence of one or more letters. If mw of the nw words in the reference text A are correctly recognized, then the word accuracy is $WAcci = mw/nw$. It is considered that a word is not correctly recognized if one or more letters of the OCR extracted word are incorrect.

A very popular similarity index that also used at this study is the *cosine* similarity index at the word level. The texts A and B are transformed into word vectors and the *cosine* index is calculated as follows:

$$cosine = \frac{(\mathbf{A} \times \mathbf{B})}{\|\mathbf{A}\|.\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n}(A_i B_i)}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{6}$$

where $\mathbf{A} \times \mathbf{B}$ is the inner product of two word vectors, the ground text A and the OCR text B respectively, the term $\| \ldots \|$ represents the corresponding magnitude of the vectors and n is the total number of both texts words. The resulting similarity ranges from -1 to 1 with 0 indicating the complete difference between two texts.

The *Jaccard* similarity index was additionally used, defining the size of the word intersection (w_{common}) between two texts, divided the total number of both texts words n.

$$Jaccard = \frac{w_{common}}{n} \tag{7}$$

The *ratio* index based on the Gestalt similarity of two texts [22] also used, comparing pairs of texts. The idea is to find the longest continuous sequence of A relative to B that does not contain ”trashes”. This index is calculated according to the following equation:

$$ratio = \frac{2K_m}{|A| + |B|} \tag{8}$$

where K_m is the number of commonly found sequence characters between the two texts A and B, divided by the total characters number of two texts ($|A|+|B|$). The returned number is the score that reflect the percentage match. It is expected that this index returns lower values than the other indexes.

4 Datasets

Three different datasets were used at this study. The first dataset is a reference dataset comprised by images of English newspapers articles. The OCR engine configured and tested on this dataset in order to achieve the same or better performance of other OCR engines that used on this benchmark dataset. The second dataset consists of scanned Greek menus images and the third dataset consists also of Greek menus images, captured by smart devices with different type of cameras, resolution and lighting. These datasets provide also, the reference texts of each image (ground true texts), for the evaluation of OCR results.

4.1 Benchmark dataset ($D1$)

The Information Science Research Institute at the University of Nevada, Las Vegas tested optical character recognition systems on an annual basis from 1992 through 1996. The dataset from Fourth Annual Test of OCR Accuracy [23] used at this study, consisting of 200 news articles with English text. These articles collected random from 50 American Newspapers. All these articles were scanned in resolution 300dpi and the correspondent ground-truth texts are also available for these images. The $(1-WERi)$ accuracy of different OCR systems in the specific competition ranged between 83.35% and 98.59%. The Tesseract OCR engine in this study, configured and applied on this benchmark dataset for verification of its performance.

4.2 Scanned images dataset ($D2$)

The $D2$ dataset is consists of 128 images of 40 different Greek menus with different background, fonts and brightness. This fact makes text recognition a difficult and complicate procedure. All menu pages of this dataset were scanned in high resolution (600dpi) with fixed scan angle, giving high quality images for text recognition in order to achieve maximum OCR performance. This dataset of high resolution scanned Greek menus and the corresponded ground texts is a unique dataset for OCR evaluation in Greek language texts.

4.3 Smart phone images dataset ($D3$)

$D3$ is a unique dataset like D_2, consists of Greek restaurant menus images. It has generated for this study, with menu photos captured by smart mobile devices, such as mobile phones or tablets. Each photo of this dataset is paired with the appropriate ground-truth text. This dataset contains 81 images, that belong to 8 different Greek menus. The difference between datasets $D2$ and $D3$ is focused on image quality. $D3$ photos were captured in real time conditions

(a) (b)

Figure 2: Benchmark dataset $D1$. (a) A sample of an article with embedded photo and (b) a simple text article image

so the quality is worst than $D2$ scanned images, for the following reasons a) the lighting between day and night b) the different angle that refers to the degree at which the camera points towards the menu paper and c) different devices and camera's resolution. Some problems that observed at $D3$ and $D2$ datasets are the different fonts, the drawings behind the important text, different shapes, colors and size of letters and also the brightness scaling. Text extraction of the above difficult cases drives to low performance of the OCR engine and our study, with the application of pre-processing methods, aids to improve OCR performance.

5 Results

The experimental setup that described at Section 3 to quantify the improvement of Tesseract OCR using image preprocessing before OCR feeding, applied on datasets $D1$ to $D3$ and the results are presented at the following figures and tables.

Figure 5 shows the cosine, Jaccard, 1-CER, 1-WER and ratio average values of Tesseract OCR, empowered with the pre-processing methods, on 200 news article images of D1 dataset. It is observed that the results of average *cosine* and $1-CERi$ values are very close to 1, indicating complete similarity between

(a)

(b)

Figure 3: Sccaned images dataset $D2$. (a) A good scanned menu image and (b) a complex scanned menu image

ground text and OCR text for all images of $D1$ dataset and for almost all preprocessing filters. The worst performance, observed for all indices of the gradient filter. The $Jacard$ and $1-WERi$ indices present lower values against the others but this fact is expected because the positions of the recognized words are taking into account during the computation of $ratio$ and $1-WERi$ indices.

It is also observed that feeding the OCR with the original images without any pre-processing, gives the same good results as the results of majority of the filters. The f_{10} (median blur) is one of the filters that present hight quality results and the profile of this filter with the detailed values is displayed at Figure6. The values of $cosine$, $1 - CERi$ indices ranged between 0.92 and 1.00. $Jaccard$ and $1 - WERi$ indices, despite that they are more sensitive in word errors, give values over 0.85 for most of the articles images, which is in line with the results of the competition that provided the images and thus confirms the high performance of tesseract ocr engine in English characters recognition.

The performance of tesseract ocr engine is not as high for $D2$ and $D3$ datasets as for $D1$. The language of the articles at $D1$ dataset is English and the training of ocr dictionary is much better than the corresponding Greek dictionary for the images of the $D2$ and $D3$ datasets.

(a) (b)

Figure 4: Menu images of dataset $D3$. (a)A good menu image shot and (b) a bad menu image shot

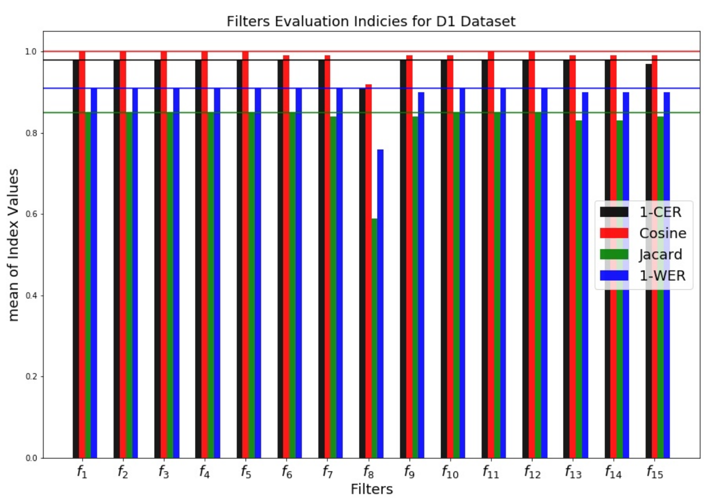

Figure 5: Average performance results for 15 pre-filtering methods of $1-CERi$, $1-WERi$, *cosine* and *Jaccard* indices using Tesseract engine on scanned images of dataset $D1$.

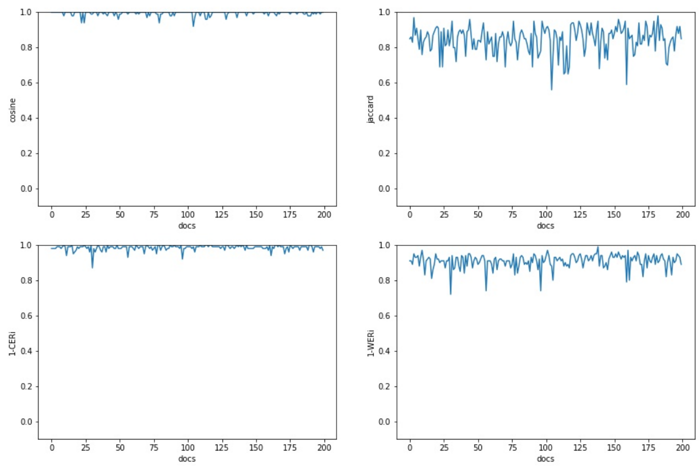

Figure 6: Values of $cosine$, $Jaccard$, $1 - CERi$, and $1 - WERi$ indices for $medblur$ filter applied on 200 news articles images of $D1$ dataset.

In particular the results of $cosine$, $Jaccard$, $1 - CERi$, and $1 - WERi$ average values, for the data set $D2$, are presented in Figure7 concerning all filters of Table 1. The horizontal lines denote the corresponding index value for f_1 case, where no preprocessing was applied on the images. It is observed that for $1 - CERi$ index there is a slight improvement in ocr efficiency when f_7, f_9, f_{10}, f_{13} and f_{15} filters were applied on the scanned food menu images, compared to f_1 (**no filter**) results. For $Jaccard$ index the f_5 filter belong to same group of the above filters and for indices of $cosine$, $1 - WERi$, two more filters are added at the same group, f_2 and f_6, indicating that image preprocessing with more than the half of the filters that applied, can improve the efficiency of tesseract ocr engine on Greek food menus images of D_2 dataset.

The profiles of $cosine$, $1 - CERi$, $Jaccard$ and $1 - WERi$ indices for f_{10} (median blur) filter, applied on 128 scanned food menu images of $D2$ dataset, are presented at Figure8. The values of $cosine$ and $1 - CERi$ range between 0 and 1. The values close to zero are indicating cases of menu images where ocr is incapable to recognise the characters of the filtered images due to complex background (see Figure 3(b)). There are also cases of scanned images that ocr recognized all characters of the food menu and the values of $1 - CERi$ index are close to 1.

Figure9 shows the average values of $cosine$, $1 - CERi$, $Jaccard$ and $1 - WERi$ indices for the filters that applied on 80 menu images of $D3$ dataset. Horizontal lines denote the corresponding value of f_1, where no preprocessing was applied on the images. It is observed that there is a slight improvement

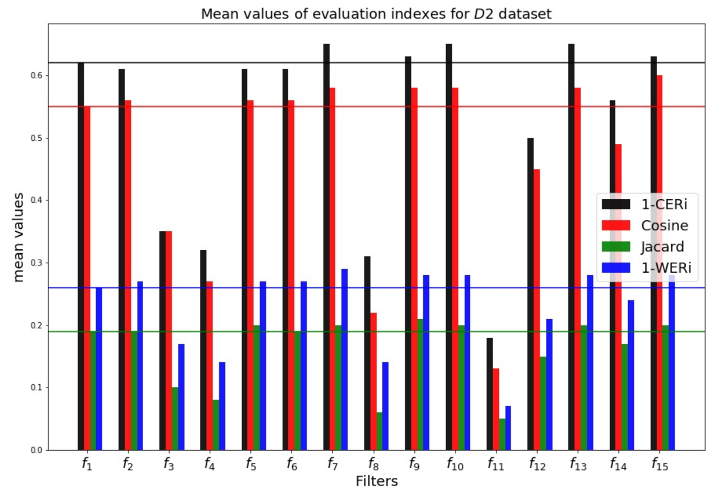

Figure 7: Average performance results of 15 pre-filtering methods for $1-CERi$, $1-WERi$, *cosine* and *Jaccard* indices using Tesseract engine on scanned images of dataset $D2$.

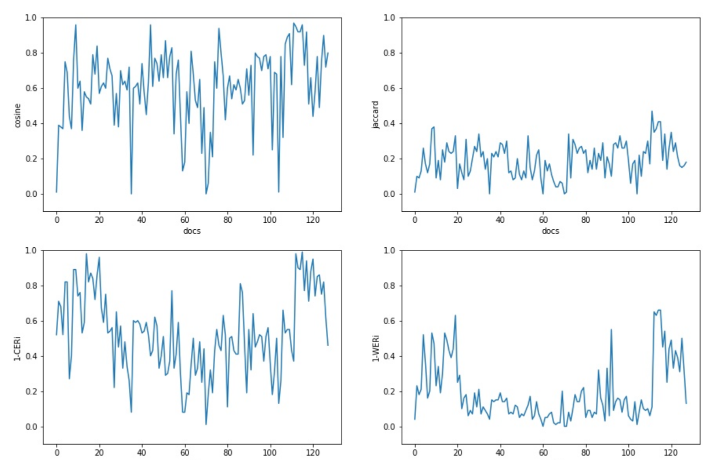

Figure 8: Values of *cosine*, *Jaccard*, $1-CERi$, and $1-WERi$ indices for f_{10} (median blur) filter applied on 128 scanned food menu images of $D2$ dataset

in ocr efficiency at character level ($1 - CERi$) when f_{14} (**erosion**) and f_{10} (**median blurring**) filters were applied on $D3$ images. The indices of *Jaccard* and $1-WERi$ are also better for f_{14} (**erosion**), f_{10} (**median blurring**) and f_7 (**closing**) filters compared to f_1.

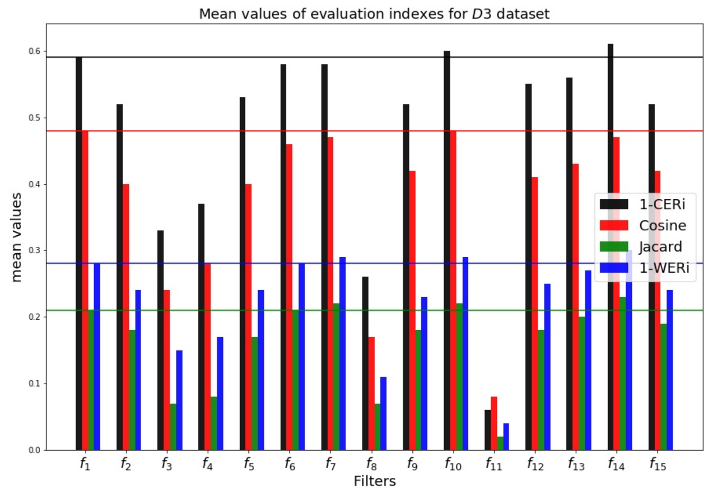

Figure 9: Average performance results for 15 pre-filtering methods in $1-CERi$, $1-WERi$, *cosine* and *Jaccard* indices using Tesseract engine on scanned images of dataset $D3$.

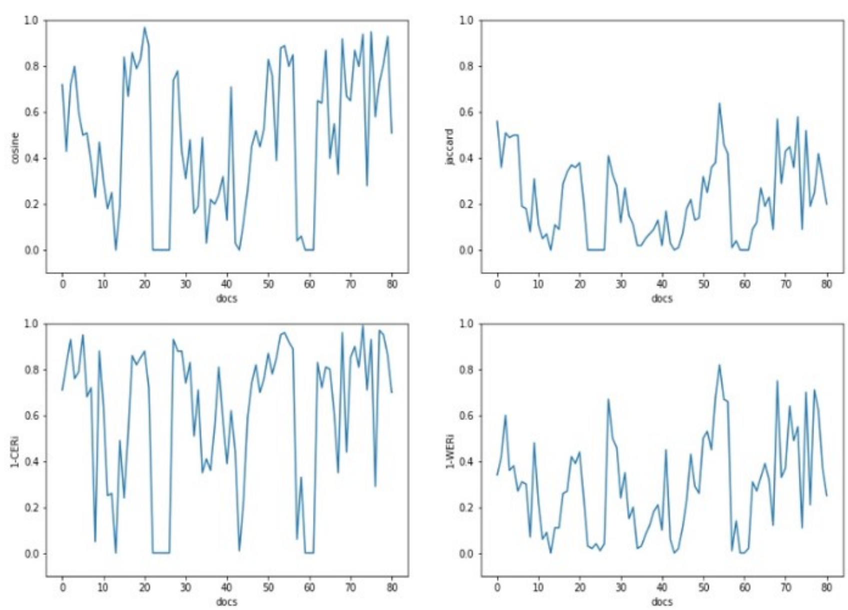

Figure 10: Values of *cosine*, *Jaccard*, $1-CERi$, and $1-WERi$ indices for median blurring (f_{10}) filter applied on 81 menu images of $D3$ dataset

The profiles of *cosine*, $1-CERi$, *Jaccard* and $1-WERi$ indices for f_{10} (median blur) filter, applied on 80 menu images of $D3$ dataset are also presented at Figure10. The values of *cosine* and $1-CERi$ range between 0 and 1 (similar to $D2$) where the values close to zero indicating a bad shot of the image (Figure 4(b)) or a very difficult background or an unknown font. The values of $Jacard$ or $1 - WERi$ indices are low for most of the cases but this is an expected result for word level accuracy.

6 Conclusion

An effective OCR evaluation of Greek menu images using multiple pre-processing filters and threshold methods in order to reveal filters that can improve ocr efficiency without retraining the ocr internal Greek dictionary, is presented at this work. The application of Tesseract ocr to benchmark images with English news articles ($D1$), yielded values of the evaluation indicators close to 1, proving the high performance of the ocr in images with English texts.

The performance of the ocr tested on two ($D1$ and $D2$) novel and unique datasets of Greek food menu images. These datasets, with the corresponding ground text files, were developed for this study. The results of ocr efficiency on $D2$ and $D3$ was worse than that of $D1$ indicating that OCR performance in English texts is superior to Greek.

A noteworthy observation is that some filters (like median blur and erosion)can improve ocr efficiency at $D2$ and $D3$ datasets, indicating that a pre-processing of images with the specific popular filters, before ocr feeding is an appropriate procedure to be followed.

In addition, the median blur filter was found to perform well on both $D2$ and $D3$ datasets, which is consistent with other literature results of this filter [6].

Future work of this study is the application and comparison of other OCR engines, like Azure, Google Vision, ABBYY Finder, on $D2$ and $D3$ Greek food menu datasets.

Acknowledgements. This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH CREATE INNOVATE (project code:(T1EDK-02015).

References

[1] Automate manual data entry using ai, https://nanonets.com/, urldate: 2021-06-13.

[2] Integrate ai-powered ocr features into your applications. urldate: 2021-06-13. https://www.abbyy.com/ocr-sdk/features/image-processing/,

[3] Intel corporation, open source computer vision library. https://github.com/itseez/opencv/, urldate: 2021-06-13.

[4] Snap any receipt, https://www.fetchrewards.com/, urldate: 2021-06-13.

[5] Turn receipts into rewards, https://www.receiptpalapp.com/, urldate: 2021-06-13.

[6] S. Dey, R. Bhattacharya, F. Schwenker and R. Sarkar, Median filter aided CNN based image denoising: An ensemble approach, *Algorithms*, **14** (2021), no. 4, 109. https://doi.org/10.3390/a14040109

[7] S. Drobac and K. Linden, Optical character recognition with neural networks and post-correction with finite state methods, *International Journal on Document Analysis and Recognition*, **23** (2020), no. 4, 279-295. https://doi.org/10.1007/s10032-020-00359-9

[8] B. Gatos, N. Stamatopoulos, G. Louloudis, G. Sfikas, G. Retsinas, V. Papavassiliou, F. Sunistira and V. Katsouros, Grpoly-db: An old greek polytonic document image database, *In 2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, (2015), 646-650. https://doi.org/10.1109/icdar.2015.7333841

[9] A. C. Glasbey and W. G. Horgan, *Image Analysis for the Biological Sciences*, Volume 1, Wiley Chichester, 1995.

[10] M.R. Gupta, Nathaniel P., N.P. Jacobson and E. K. Garcia, OCR binarization and image pre-processing for searching historical documents, *Pattern Recognition*, **40** (2007), no. 2, 389-397. https://doi.org/10.1016/j.patcog.2006.04.043

[11] B. Hunt, A matrix theory proof of the discrete convolution theorem, *IEEE Transactions on Audio and Electroacoustics*, **19** (1971), no. 4, 285-288. https://doi.org/10.1109/tau.1971.1162202

[12] J. Sauvola and M. Pietikinen, Adaptive document image binarization, *Pattern Recognition*, **33** (2000), no. 2, 225-236. https://doi.org/10.1016/s0031-3203(99)00055-2

[13] V. I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals, *Cybern Control Theory*, **10** (1966), 707-710.

[14] H. Lin, Review of scene text detection and recognition, *Archives of Computational Methods in Engineering*, **27** (2019), 433-454.

[15] J. Margaronis, M. Christou, E. Kavallieratou and T. Tzouramanis, Gcdb: A character database system, *In Proceedings of the International Workshop on Multilingual OCR*, MOCR '09, New York, NY, USA, 2009. ACM. https://doi.org/10.1145/1577802.1577821

[16] W. Min, S. Jiang, L. Liu, Y. Rui and R. Jain, A survey on food computing, *ACM Computing Surveys*, **52** (2018), no. 5, 1-36. https://doi.org/10.1145/3329168

[17] C. Nabendu, S.H. Soharab Shaikh and K. Saeed, A comprehensive survey on image binarization techniques, *In Exploring Image Binarization Techniques*, (2014), 5-15.

[18] F. Orieux, J.F. Giovannelli and T. Rodet, Bayesian estimation of regularization and point spread function parameters for wiener-hunt deconvolution, *Journal of the Optical Society of America A*, **27** (2010), no. 7, 1593-1607. https://doi.org/10.1364/josaa.27.001593

[19] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics*, **9** (1979), no. 1, 62-66. https://doi.org/10.1109/tsmc.1979.4310076

[20] J.R. Parker, *Algorithms for Image Processing and Computer Vision*, Wiley Computer Publishing, 2010.

[21] E.D. Pisano, S. Zong, B.M. Hemminger, M. DeLuca, R.E. Johnston, K. Muller, M.P. Braeuning and S.M. Pizer, Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms, *Journal of Digital Imaging*, **11** (1998), no. 4, 193-200. https://doi.org/10.1007/bf03178082

[22] J. W. Ratcliff and D. Metzener, Pattern matching: The gestalt approach, *Dr. Dobb's Journal*, **13** (1988), no. 7, 1-46.

[23] S. Rice, F. Jenkins and T. Nartker, The fourth annual test of OCR accuracy, 2012.

[24] B. Sainz-De-Abajo, J. M. Garca-Alonso, J. J. Berrocal-Olmeda, S. Laso-Mangas and I. De La Torre-Dez, Foodscan: Food monitoring App by scanning the groceries receipts, *IEEE Access*, **8** (2020), 227915-227924. https://doi.org/10.1109/access.2020.3046031

[25] B. Sankur and M. Sezgin, Image thresholding techniques: A survey over categories, *Pattern Recognition*, **34** (2001), 1573-1583.

[26] R. Smith and Inc. Google, An overview of the tesseract OCR engine, *In Proc. 9th IEEE Intl. Conf. on Document Analysis and Recognition (ICDAR)*, (2007), 629-633. https://doi.org/10.1109/icdar.2007.4376991

[27] W. Swastika, Yoko, H. Setiawan and M. Subianto, Android based application for recognition of indonesian restaurant menus using convolution neural network, *International Conference on Sustainable Information Engineering and Technology (SIET)*, (2017), 3034. https://doi.org/10.1109/siet.2017.8304104

[28] R. Ullah, A. Sohani, F. Ali and R. Messier, OCR engine to extract food-items,prices,quantity, units from receipt images, heuristics rules based approach, *International Journal of Scientific and Engineering Research*, **9** (2018), no. 2, 1334-1341.

[29] K. Zuiderveld, *Contrast Limited Adaptive Histogram Equalization*, Graphics gems IV, 1994.

Received: November 29, 2021; Published: January 5, 2022