A Document Layout Analysis Method Based on Morphological Operators and Connected Components

Sebastian W. Alarcón Arenas
Universidad Católica San Pablo
Arequipa, Peru
sebastian.alarcon@ucsp.edu.pe

Graciela L. Meza-Lovón
Universidad Católica San Pablo
Arequipa, Peru
gmezal@ucsp.edu.pe

Yessenia Yari
Universidad Católica San Pablo
Arequipa, Peru
ydyari@ucsp.edu.pe

***Abstract*—During the last decades, the interest in preserving digitally historical documents have gained considerable attention. To exploit all the advantages and opportunities offered by the digitized documents, it's necessary to understand their contents. The first step toward that understanding is to determine the locations of the entities of the document, such as figures, titles, and captions, text, etc. This paper presents a new hybrid approach to analyze the structure of documents that is founded on morphological operators and connected components. The proposed method is divided into two stages, preprocessing, in which the quality of the document images is enhanced; and layout analysis, in which, we identify three types of layout. We also include a fragmentation process, in which we divide the page image into sections. Finally, We conducted the experiments on a dataset containing ancient historical newspapers.**

***Index Terms*—Document Layout Analysis, Morphological Operators, Connected Components, Old Newspapers.**

I. Introduction

Due to the fast development of mass storage systems and the interest in preserving the information contained in old printed documents (e.g., books, newspapers, magazines) and old manuscripts, the projects related to the digitization of documents have increased considerably. The digitalization of ancient documents plays a fundamental role not only for their preservation but also for their dissemination to larger audiences.

In order to exploit the benefits that a digitized document offers, it is necessary to understand its content. Traditional OCR techniques, which provide good results when applied to modern documents, do not offer the desired results for old documents. A cause of OCR low performance relies on the additional difficulties that come from the inherent characteristics of old documents, such as degradation of paper, fading of the ink, the presence of spots, among others. However, this is not the only one reason for the OCR low performance. According to Shafait [1], a crucial step for developing an OCR system is to identify the structures contained in document images, such as titles, paragraphs, images, tables, etc. This process is known as Document Layout Analysis [1].

For dealing with the problem of identifying the structures of a document, there are different methods in the literature but their effectiveness depends on the type of document. Many authors have classified them into three main categories [2]: bottom-up, top-down and hybrid methods.

Top-down methods separate the document into different regions and then use many heuristic filters to classify each region. For example Laiphangbam et al. [3], proposed to classify the regions using a histogram and a Gaussian function to find more stable peaks and valleys. Then, with this information, they calculate the optimal division between the categories (text, title, and graphics). Another top-down method is the one proposed by Naik et al. [4], who after binarizing the image, segment the titles using horizontal projections; once the headers of the documents are separated, Naik et al. [4] use vertical projections to separate the columns of a document.

Bottom-up methods start with local information such as words, text lines and then, merge them into blocks or paragraphs. These methods require a lot of memory space and are time-consuming. Some authors that use this approach are Ferilli et al. [5] [6].

Besides, exist the hybrid methods that combine the top-down and bottom-up methods and many authors used them to solve the analysis layout problem. Vasilopoulos and Kavallieratou [7] proposed a hybrid method that consists in analyzing the layouts using connected components. They obtained the height of the most frequent connected components, which was subsequently used as a threshold. In addition, they applied morphological operations such as opening operator to obtain a mask, dilatations to obtain the layouts based on the height computed before. On the other hand, Yavas and Ragot [8] designed a technique based on FAST key points. They divided the image into blocks and for each block, they computed its density of points. The denser blocks are considered as text blocks. Then, connectivity of blocks is verified in order to cluster them and obtain complete text blocks. Tran et al. [9] analyzed whitespace in maximum horizontal homogeneous regions to classify text elements and non-text elements. For the extraction of text regions, they applied mathematical morphology. Later, by using machine learning techniques, non-text elements are classified into separators, tables, images. Finally, Corbelli et al. [10] combined a classic top-down method, XY-

Cut algorithm, and a bottom-up classification method which is based on local geometry features. The regions are classified into different classes of layouts using features extracted from a Convolutional Neural Network combined with a Random Forest classifier.

In this paper, we propose a hybrid approach to document structure analysis, which is founded on morphological operators and connected components. Our method is divided into two stages, namely, 1) Preprocessing, which includes several tasks to improve the quality of the images, and 2) Processing, which split the image into subimages using section separators then for each subimage, three types of layout are identified: non-text objects, normal-sized text, and titles. We perform the experiments on a dataset containing old historical newspapers.

The rest of this paper is organized as follows. In Section II, we present the details of the proposed method, which is divided into two stages: Preprocessing, described in Section II-A, and the Analysis of the Document Structure, explained in Section II-B. Thereafter, we evaluate the proposed method applied to old newspaper documents and compare the results in Section III. Finally, the conclusions are presented in Section IV.

II. The Proposed Method

Our proposal is made up of two stages. The first stage consists in preprocessing the images and allows manipulating the images in order to ease the tasks of the second stage. This stage includes operations such as noise removal, conversion of images into grayscale, and contrast enhancement. The second stage, corresponding to the analysis of the layout itself, allows finding structures containing into the document, such as paragraphs, titles, and images. Fig. 1 shows the architecture of the proposed method.

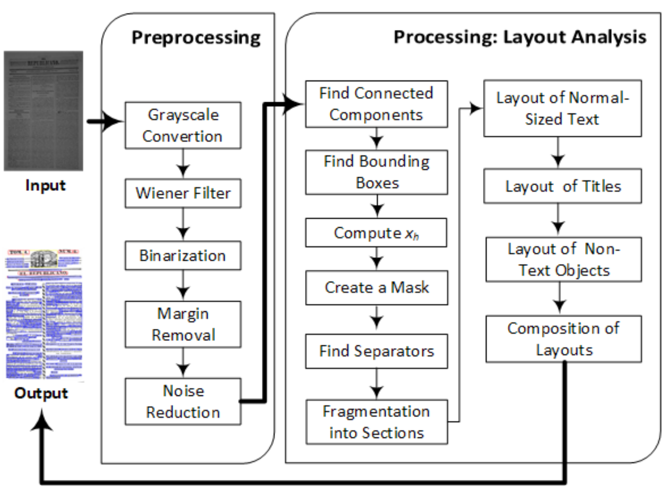
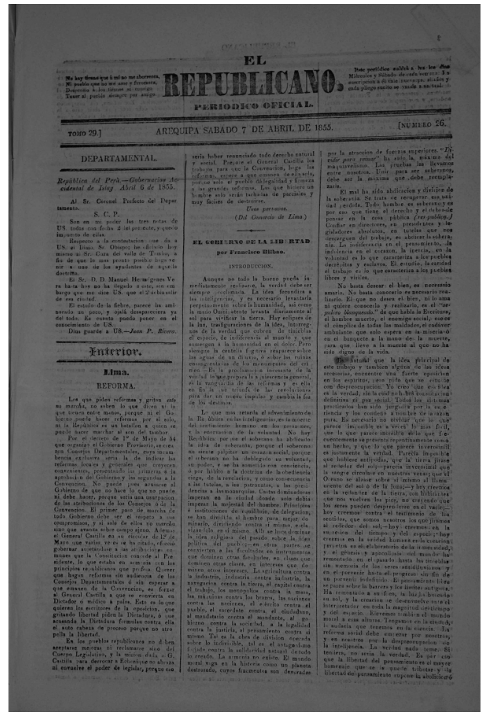


Fig. 1. Architecture of the proposed method.

A. Preprocessing

The processing stage aims at increasing the contrast between the background and the foreground by applying several image processing techniques described in the following.

Firstly, we convert the original image, in RBG format, to grayscale, and then, we utilize the Wiener filter with a window 3×3. Then, by applying the Otsu algorithm [11], the ink-printed regions are set to 0 (black) and the rest of the pixels of the image are set to 255 (white) (see Fig. 2).



EL REPUBLICANO.

PERIODICO OFICIAL.

TOMO 29.] AREQUIPA SABADO 7 DE ABRIL DE 1855. [NUMERO 26.

(a) Original image in grayscale format.

EL REPUBLICANO.

PERIODICO OFICIAL.

TOMO 29.] AREQUIPA SABADO 7 DE ABRIL DE 1855. [NUMERO 26.

(b) Binarized image.

Fig. 2. Noise reduction using the Wiener filter with a window 3×3, and binarization using the Otsu algorithm [11].

As can be noted in Fig. 3(a), there is a great amount of noise introduced in the areas corresponding to the margins during the process of digitalization; because of that, the preprocessing stage also removes the margins of the documents, as shown in Fig. 3(b). The method applied for this purpose is based on computing horizontal and vertical projection profiles, which contain peaks and valleys. The margins areas are visualized in the profiles as extended valleys, so that, by identifying the beginning and the endpoints of these valleys, we obtain cuts points, which are used to separate the areas containing text from the areas that correspond to margins.

A projection profile can be represented as a vector $\mathbf{P} = [p_1, p_2, \ldots, p_L]$, where L is the length of the vector, and p_i is the i-th element estimated by summing the intensity values of the pixels in a specific direction. In particular, the vertical projection profile of an image with height H and width W, is computed via,

$$p_i = \sum_{j=1}^{H} p(i, j). \tag{1}$$

Then, we find the connected components in the inverted image using the connectivity parameter set to 8, and subsequently, we eliminate those components having a reduced number of pixels (according to the experiments, less than 30 pixels) for being considered noisy components.

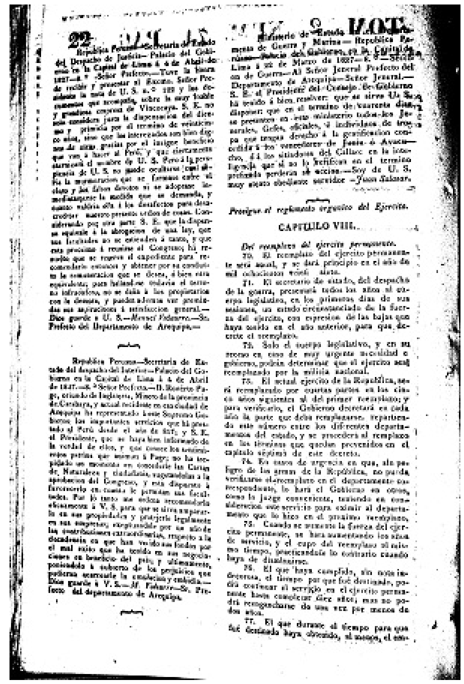
B. Analysis of the Document Structure

As mentioned in Section I, there exist different methods for analyzing the structure of the documents; many of them are based on projection profiles. Nevertheless, for documents that have a complex structure, this technique does not provide the desired results. Since the dataset used in this work contains images with complex structure, we propose a new method based on morphological operations and the identification of connected components.

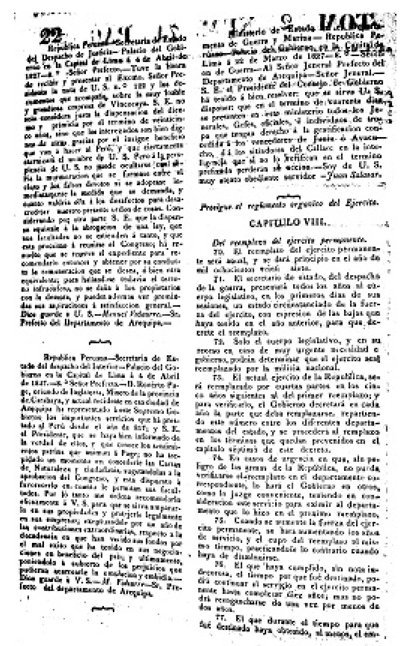
Our proposal is similar to the methodology by Vasilopoulos and Kavallieratou [7]; nonetheless, the images of our dataset contain separating structures that divide the document into different regions. For that reason, we have included an operation that detects those separators, which are used to fragment the image.

1) Detection of Section Separators and Fragmentation into Sections: First, we need to compute the parameter x_h. For doing so, we find the connected components using 8 connectivity (see Fig. 4) and we delimit their bounding boxes. Then, we compute the height of those boxes and we obtain the most frequent height. The parameter x_h is the half of the most frequent height.

Secondly, we create a mask for finding the structures that can be separators. To this end, we perform an opening operation, which consists of eroding and dilating the image, using a particular morphological structure. Since the sections separators are line-shaped, the employed structure is a line, or more precisely, a square structure with the shape of a line. The opening operation is performed twice: one for finding vertical



(a) Binarized image.



(b) Image without margins.

Fig. 3. Removal of the margins of a page image.

separators and other for finding horizontal separators, by using square structures whose dimensions are $\frac{x_h}{4} \times \frac{W}{2}$ and $\frac{H}{4} \times \frac{x_h}{2}$, respectively.

For creating the mask, we also include a complementary heuristic to locate line-shaped separators: the bounding box of a component that represents a line has a disproportionate relation between its height h and its width w, i.e., the more disproportionated is the relation between the height and the width of the bounding box, the higher the certainty of that the bounding box circumscribes a line. Based on this rationale, if the ratio r between m_l and m_g is less than a threshold $tr_l = 0.1$, the bounding box is considered as a line-shaped separator, where, if $h > w$ then $m_g = h$ and $m_l = w$; otherwise, $m_g = w$ and $m_l = h$. A mask containing all the possible separators of a page image is shown in Fig. 5. The background areas are colored in white while the inked areas in black.

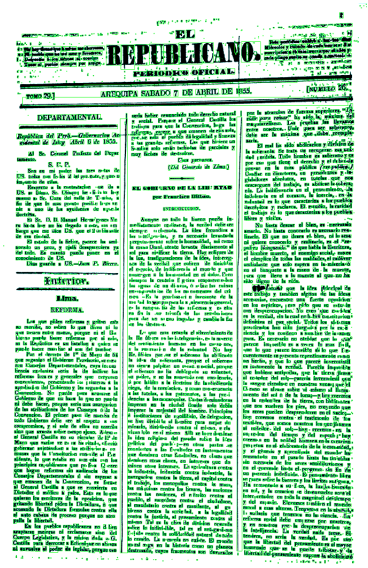

EL REPUBLICANO.
PERIODICO OFICIAL.
AREQUIPA SABADO 7 DE ABRIL DE 1855.
DEPARTAMENTAL.
Interior.
Lima.
REFORMA.

Fig. 4. Bounding boxes of the connected components obtained using 8 connectivity.

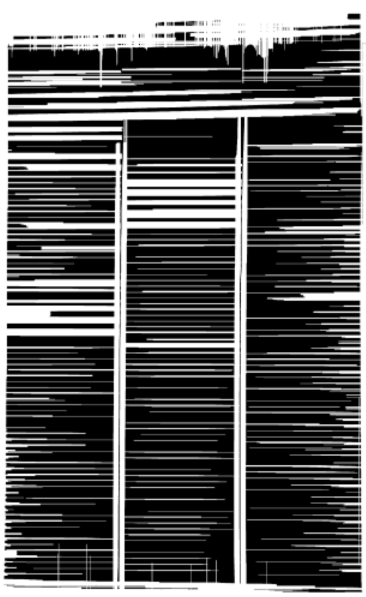
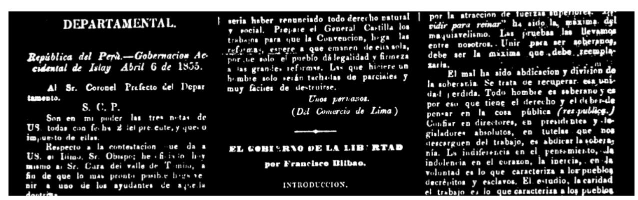


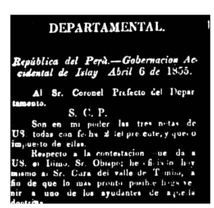
Fig. 5. Mask containing all the possible separators of a page image.

Subsequently, based on the mask computed before, we need to decide the order of the fragmentation (horizontal-vertical or vertical- horizontal) using real separators. For this purpose, we compute the horizontal $\mathbf{PH}=[ph_1, ph_2, \ldots, ph_W]$ and the vertical $\mathbf{PV} = [pv_1, pv_2, \ldots, pv_H]$ projection profiles of the mask. Both $\mathbf{PH}$ and $\mathbf{PV}$ are functions with peaks and valleys, which correspond to background areas and inked areas, respectively. The vertical and horizontal projections are truncated such that, if $ph_i > 0.75W$, then $ph_i = 0.75W$ and if $pv_i > 0.75H$, then $pv_i = 0.75H$. By doing so, the peaks became plateaus which could represent separators. The criterion for deciding the order of the fragmentation is based on the number of plateaus: if the number of horizontal plateaus is greater than the vertical one, then we fragment horizontal-vertical; otherwise, we fragment vertical-horizontal.

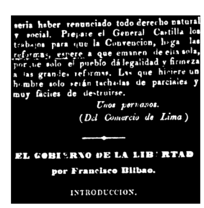
First, let's assume that we fragment horizontally first and then vertically. For doing so, we need to discriminate between plateaus representing real separators from plateaus that represent background areas (e.g., the space between text lines), by following this heuristic: if the distance between two consecutive plateaus is greater than $2x_h$, then the plateau is considered as a real separator. Based on the obtained separators, the page image is split into fragments, as shown in Fig. 6(a). Subsequently, for each fragment, it is needed to split the fragment vertically. To this end, we compute its vertical projection profile $\mathbf{PV'} = [pv'_1, pv'_2, \ldots, pv'_{h_f}]$, where h_f is the height of the fragment; and we truncate $\mathbf{PV'}$ using the following rule: if $pv'_i > 0.98h_f$, then $pv'_i = 0.98h_f$. Once more, we need to determine real vertical separators using this rule: if the distance between two consecutive plateaus is greater than $10x_h$, then the plateau is considered as a real separator and it is used to split each fragment into subimages, as shown in Fig. 6(b)(c)(d).


DEPARTAMENTAL.
República del Perú.—Gobernacion Accidental de Islay Abril 6 de 1855.
Al Sr. Coronel Prefecto del Departamento.
S. C. P.
(Del Comercio de Lima)
EL GOBIERNO DE LA LIBERTAD
por Francisco Bilbao.
INTRODUCCION.

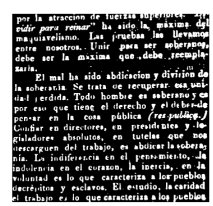
(a) A fragment obtained after finding the horizontal separators.



(b) Subimage 1.



(c) Subimage 2.



(d) Subimage 3.

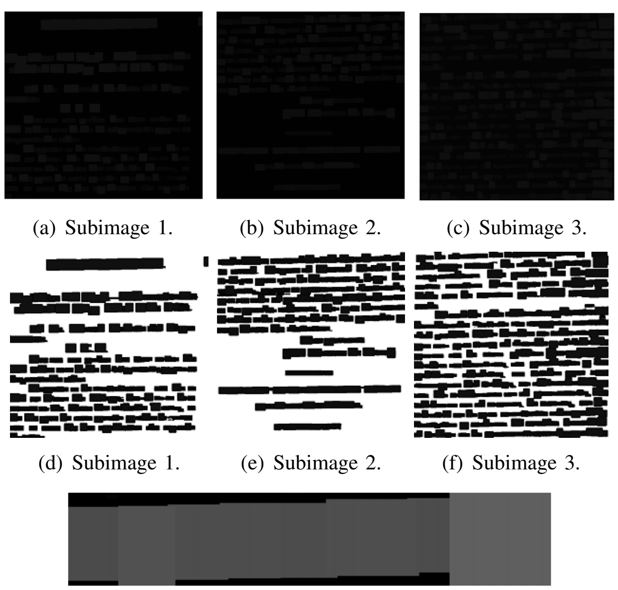
Fig. 6. Fragmentation process: (a) The fragment was obtained after applying a horizontal fragmentation, and then that fragment was split into three subimages (b) (c) (d) by applying vertical fragmentation.

On the contrary, if the vertical separators are located first, and the horizontal ones later, the plateau whose distance to the next one is greater than $10x_h$ is considered as a real separator. The real separators are used to break down vertically the image page into fragments. Analogous to what described for the horizontal-vertical case, the horizontal projection profile

$\mathbf{PH}'$ is found for each fragment, and then $\mathbf{PH}'$ is truncated using the same value, i.e., 0.98. If the distance between two consecutive plateaus is greater than $2x_h$ then the plateau is considered as a horizontal separator.

2) Detection of Normal-Sized Text: In this step, we need to detect the normal-sized text areas. We based our rationale on the assumption that the normal-sized text objects have bounding boxes whose heights are smaller than those of the images and titles. So that, we change the intensity of each pixel according to the height of the bounding box the connected component belongs to. This results in that the higher the component is, the clearer (close to white) the intensity value will be. The change of the intensity of pixel $p(x,y)$ is given by, $p(x,y) = (\frac{h(p(x,y))}{2} \mod 255)$, where, $h(p(x,y))$ is the height of the bounding box that contains the pixel $p(x,y)$.

Figs. 7(a)(b) and (c) show the subimages presented in Figs. 6 (b)(c) and (d) after changing the intensity values of their pixels. As noted, the bounding boxes are almost black since the height of the bounding boxes is small; however, the bounding boxes of the titles, presented in Fig. 7(g), are less black and more. In Figs. 7(d)(e) and (f), we changed the color of the bounding boxes and the background areas of Figs. 7(a)(b) and (c) only for visualization purposes.



(a) Subimage 1. (b) Subimage 2. (c) Subimage 3.

(d) Subimage 1. (e) Subimage 2. (f) Subimage 3.

(g) Subimage 4.

Fig. 7. Update of the intensity values of the pixels of different subimages. In (a), (b) and (c), the pixels inside the bounding boxes are almost black. In (d), (e) and (f), the color of the pixels inside the bounding boxes and the background areas was modified for visualization purposes. (g) shows that the bounding boxes of the titles are less black.

Then, the dilation operation is applied with the purpose of getting the letters together by using a square structure with dimensions $\frac{x_h}{3} \times 2x_h$, and subsequently, the connected components are found again along with their bounding boxes. Then, for each connected component, the following rule is used in order to determine whether it is related to normal-sized text: if at least the 51% of its pixels have the intensity value lower than $2x_h \mod 255$, then the connected component is labeled as normal-sized text. The layout of normal-sized text is presented in Fig. 8.

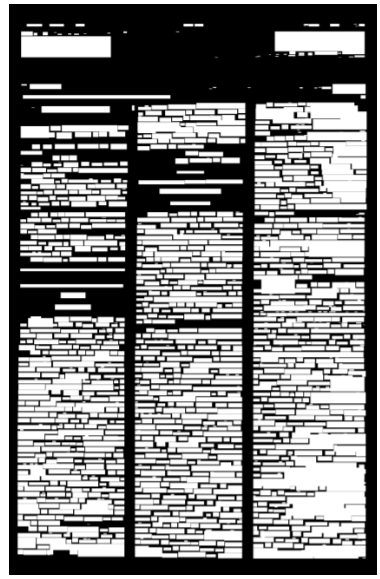


Fig. 8. Layout of normal-sized text.

3) Detection of Titles: With the aim of locating the titles of the document, the dilation operator is applied another time, using a square structure, whose measurement is equal to $x_h \times 4x_h$. The connected components and the corresponding bounding boxes are determined again. Then, for finding the titles, the following rule is employed: if the relation between the height and the width of a bounding box is greater than 0.2 and the half of its height is greater than x_h then the bounding box is considered as a title. Fig. 9 shows the layout of normal-sized text.

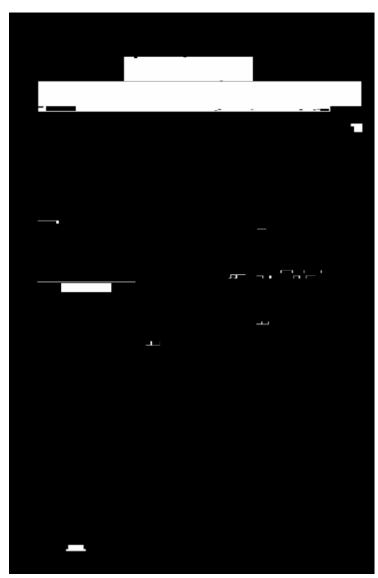
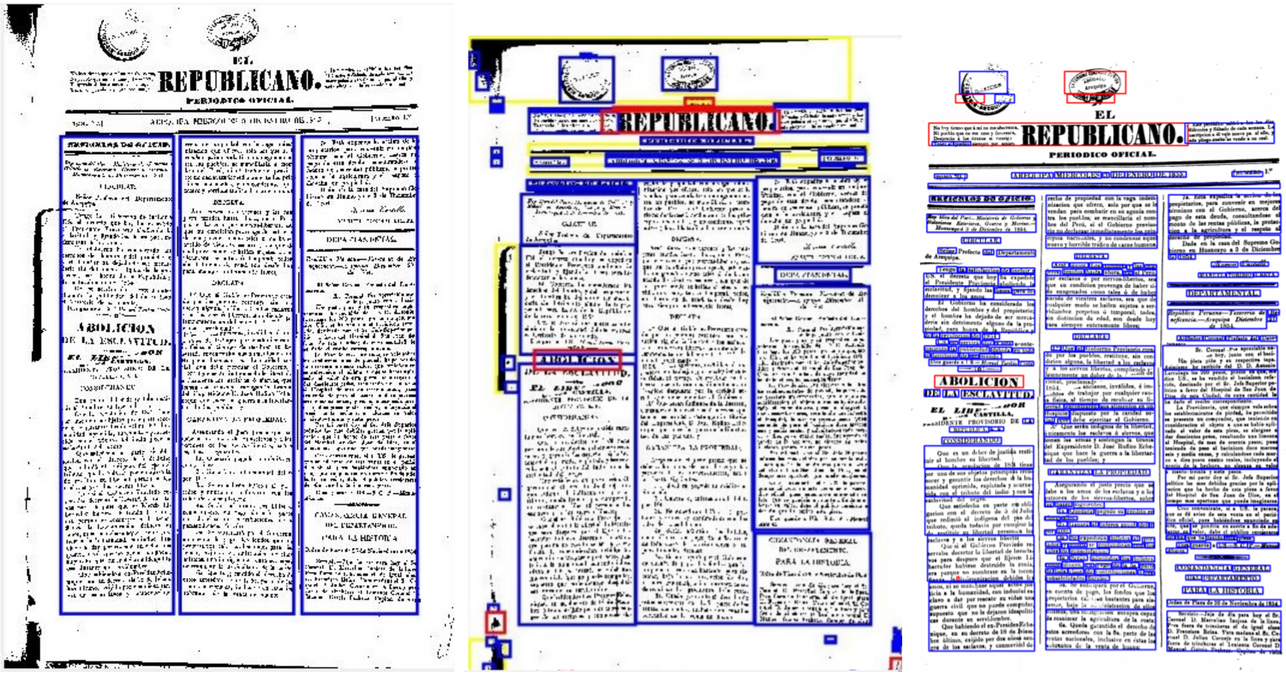


Fig. 9. Layout of titles.



(a) SofUP - Naik et al. [4]. (b) CCMO - Vasilopoulos and Kavallieratou [7]. (c) Our Proposed Method.

Fig. 10. Example 1: Comparative results of SofUP, CCMO and our proposed method.

4) Detection of Non-Text Objects: Objects such as figures, pictures, stamps are considered as non-text objects in this paper. In order to locate them, we proceed as follows. For each bounding box of the remaining image, i.e., the image without normal-sized text and without titles, we follow this rule: if its width is three times lower than its height then the bounding box is considered as a non-text object.

Finally, the titles are marked in red, the normal-sized texts are delimited in blue, while the non-text objects in are delineated in yellow.

III. Experiments and Results

In this section, we present preliminary results obtained after executing some experiments, which were conducted using a small dataset containing 45 images from two old newspapers, namely, “El Republicano” and “El Deber”. The resolution of “El Republicano” images is 2209×3697 while the one of “El Deber” is 2665×4265 pixels. Both newspapers are in RGB format.

We compare our proposed method with the method of Segmentation of Unstructured Newspapers (SofUP) proposed by Naik et at. [4]; and the method by Vasilopoulos and Kavallieratou [7], which is based on contour classification and morphological operations (CCMO). We present the layouts delimited by each method in Figs. 10, 11 and 12. As mentioned before, the titles are marked in red, the normal-sized texts in blue, and the non-text objects in yellow.

In order to evaluate the methods, we calculated the precision, recall and F1 score for the three layouts, namely, titles, non-text objects, and normal-sized text. The precision, recall and F1 score are defined according Eqs. 2, 3 and 4, respectively.

$$precision = \frac{VP}{VP + FN}, \tag{2}$$

$$recall = \frac{VP}{VP + FP}, \tag{3}$$

$$F1 \text{ score} = 2 \times \left(\frac{precision \times recall}{precision + recall}\right), \tag{4}$$

where VP represents the true positives and corresponds to the titles, the normal-sized text and non-text objects correctly identified; FN, the false negatives, corresponds to the images incorrectly rejected, and FP, the false positives, to those incorrectly identified.

Figs. 13, 14 and 15 show the precision, recall, and F1 score, respectively, after applying the three techniques using the dataset; while Figs. 16, 17 and 18 present the results for the three metrics: Titles, Normal-Sized Text, and Non-Text Objects, respectively.

As can be noted, the technique based on morphological operations (CCMO) proposed by Vasilopoulos and Kavallieratou [7] surpasses the first technique (SofUP) by Naik et al. [4] in terms of precision, but our proposal which combines both approaches is better than both. In relation to titles and non-text elements, we obtain lower precision. This is due to the fact that in some cases when normal-sized text elements, such as characters, are very separate from each other, our proposal labels them as non-text elements, i.e., as figures, which contributes to increasing the error of precision.

(a) SofUP - Naik et al. [4]. (b) CCMO - Vasilopoulos and Kavallieratou [7]. (c) Our Proposed Method.

Fig. 11. Example 2: Comparative results of SofUP, CCMO and our proposed method.

In Table I, we present the overall results for the three techniques using precision, recall and F1 score, while in the Tables II, III and IV, we show the results for each type of layout.

TABLE I
OVERALL RESULTS: PRECISION, RECALL AND F1 FOR THE SOFUP, CCMO AND OUR PROPOSAL.

Technique	Precision	Recall	F1
SofUP - Naik et al. [4]	63.21%	73.96%	61.06%
CCMO - Vasilopoulos and Kavallieratou [7]	74.61%	81.00%	76.57%
Our Proposal	89.22%	87.40%	87.46%

TABLE II
RESULTS FOR TITLES: PRECISION, RECALL AND F1 FOR THE SOFUP, CCMO AND OUR PROPOSAL.

Technique	Precision	Recall	F1
SofUP - Naik et al. [4]	55.39%	62.42%	28.07%
CCMO - Vasilopoulos and Kavallieratou [7]	49.67%	79.95%	45.56%
Our Proposal	84.24%	80.38%	71.76%

In tables presented before, it can be noted again that in most of the cases, the technique proposed in this paper provides superior results than those of the other techniques, except for 0.10% with respect to precision of normal-sized text in which CCMO is better. This is due to the fact that in some cases the separators computed by our technique fails to divide correctly the paragraphs that are very close to the area of the titles.

TABLE III
RESULTS FOR NORMAL-SIZED TEXT: PRECISION, RECALL AND F1 FOR THE SOFUP, CCMO AND OUR PROPOSAL.

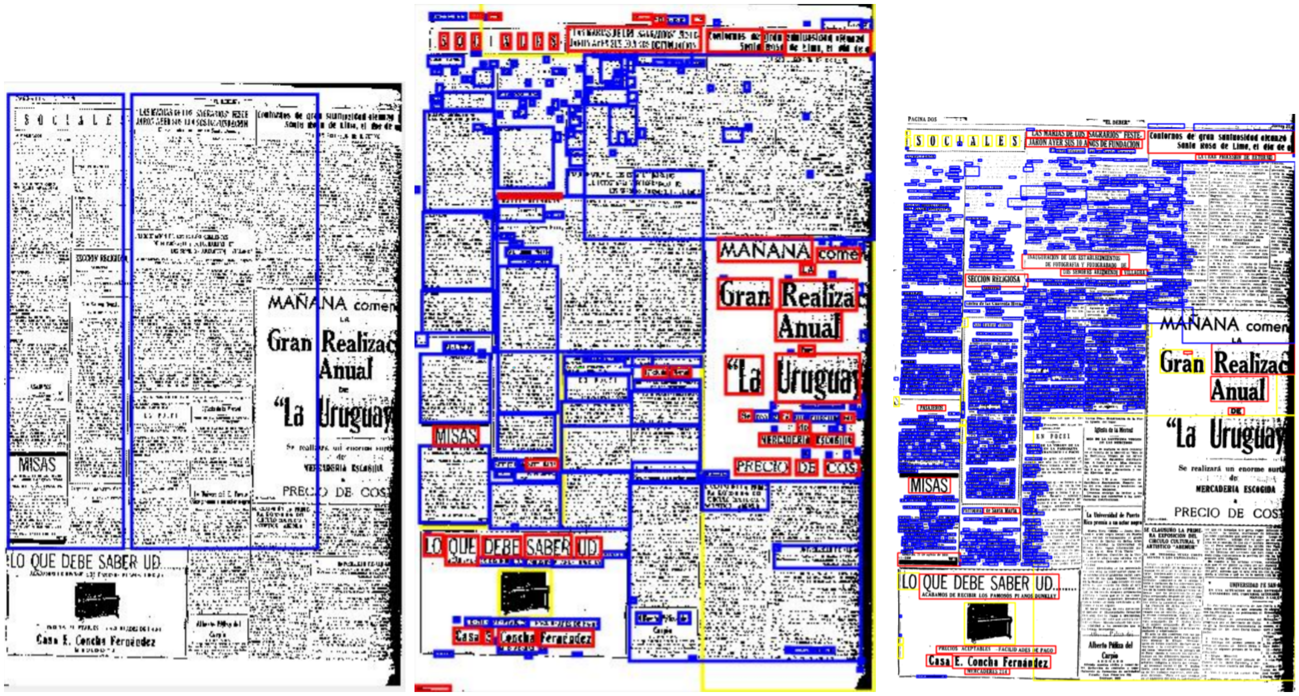
Technique	Precision	Recall	F1
SofUP - Naik et al. [4]	79.32%	86.56%	80.62%
CCMO - Vasilopoulos and Kavallieratou [7]	97.21%	92.00%	92.71%
Our Proposal	97.11%	93.32 %	93.78%

TABLE IV
RESULTS FOR NON-TEXT ELEMENTS: PRECISION AND RECALL FOR THE SOFUP, CCMO AND OUR PROPOSAL.

Technique	Precision	Recall	F1
SofUP - Naik et al. [4]	63.64%	54.55%	18.18%
CCMO - Vasilopoulos and Kavallieratou [7]	70.65%	59.66%	41.62%
Our Proposal	68.99 %	65.15 %	57.39%

Taking into account these results, our proposal is more appropriate to use for this type of old newspapers, however, we think that the precision and recall need to be improved for the layouts of titles and non-text objects. We attribute these failures to several factors, such as the presence of noise and the slant that some page images present. These factors cause the image to be fragmented incorrectly.

Fig. 19 and Table V present the execution time in seconds. Table V shows that SofUP is faster than the other techniques used for comparison, but it does not correctly identify the entities of the documents. On the other hand, our proposal



(a) SofUP - Naik et al. [4]. (b) CCMO - Vasilopoulos and Kavallieratou [7]. (c) Our Proposed Method.

Fig. 12. Example 3: Comparative results of SofUP, CCMO and our proposed method.

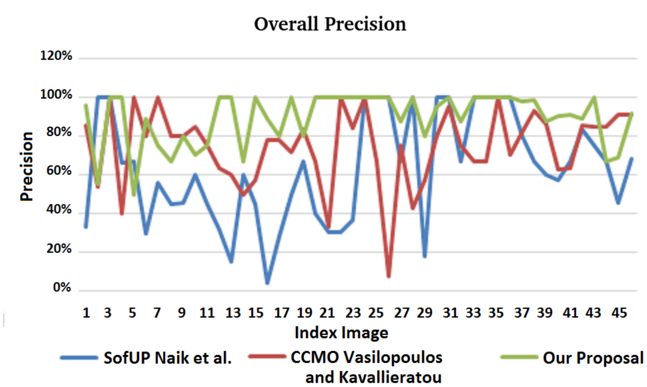


Fig. 13. Overall precision.

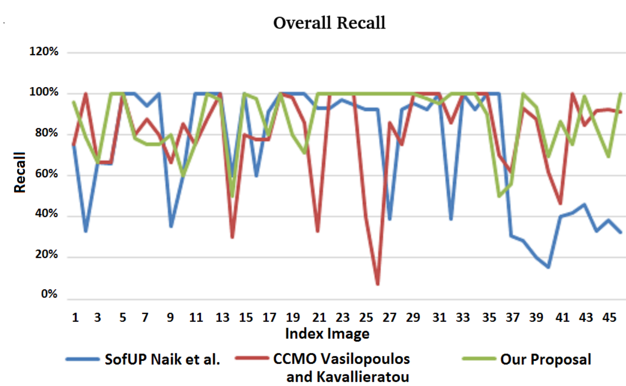


Fig. 14. Overall recall.

provides lower processing time in most of the cases when compared to CCMO. Note also that for the images between 37 and 45, the time is much longer. This is because those images belong to the newspaper El Deber, which present a more complex structure (i.e., it contains more columns of text, letters are smaller and are close to each other) causing the time for finding the connected components to increase considerably.

TABLE V
EXECUTION TIME IN SECONDS.

Technique	Execution Time (seconds)
SofUP - Naik et al. [4]	1.51
CCMO - Vasilopoulos and Kavallieratou [7]	176.77
Our Proposal	132.03

IV. CONCLUSIONS

In this paper, we propose a new hybrid method for layout analysis. This method is based on morphological operations and connected components.

An important process for analyzing the structure of the documents is the preprocessing stage, since the better the quality of the images, the better the obtained results. Preprocessing includes tasks to reduce the noise, to transform the image to grayscale and to binarize the image applying the Otsu algorithm.

Our method identifies three types of layouts: titles, normal-sized text, and non-text objects, in a dataset composed of images of “El Republicano” and “El Deber” newspapers.

The results show that the proposed method is superior

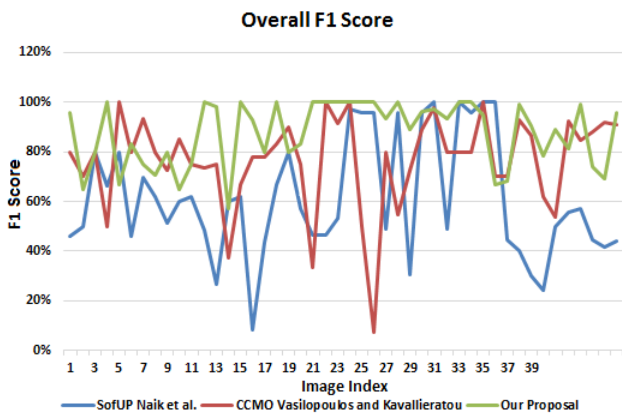


Fig. 15. Overall F1 score.

to the method by Naik et al. [4] and Vasilopoulos and Kavallieratou [7]. Using our technique, we obtain 89.22% of precision. However, due to the used thresholds, in some cases, our proposal does not identify correctly the layouts of titles.

Although our proposed method consumes more time than Naik et al. [4] technique, our method can still reduce the processing time, since the fragmentation makes it possible for the analysis of each section to be carried out in parallel. This is possible because our technique includes a process by which documents are divided into sections before labeling the different structures of page images.

ACKNOWLEDGMENT

The authors would like to thank Universidad Católica San Pablo for the financial support.

REFERENCES

[1] F. Shafait, D. Keysers, and T. M. Breuel, "Performance comparison of six algorithms for page segmentation," in *Document Analysis Systems*, 2006.

[2] A. M. Namboodiri and A. K. Jain, "Document structure and layout analysis," in *Digital Document Processing*. Springer, 2007, pp. 29–48.

[3] L. Melinda, R. Ghanapuram, and C. Bhagvati, "Document layout analysis using multigaussian fitting," in *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, vol. 1. IEEE, 2017, pp. 747–752.

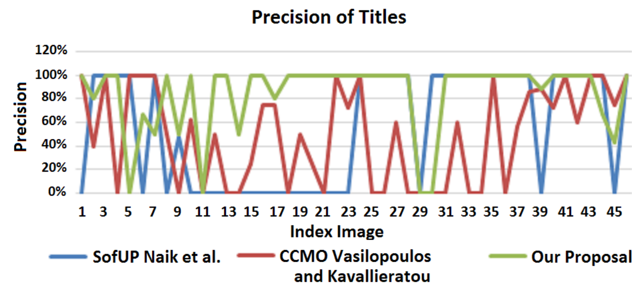
[4] S. Naik, R. Dinesh, and S. Prabhanjan, "Segmentation of unstructured newspaper documents," *International Journal of Advanced Engineering Research and Science*, vol. 4, no. 5.

[5] S. Ferilli, T. M. A. Basile, and F. Esposito, "A histogram-based technique for automatic threshold assessment in a run length smoothing-based algorithm," in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, ser. DAS '10. New York, NY, USA: ACM, 2010, pp. 349–356.

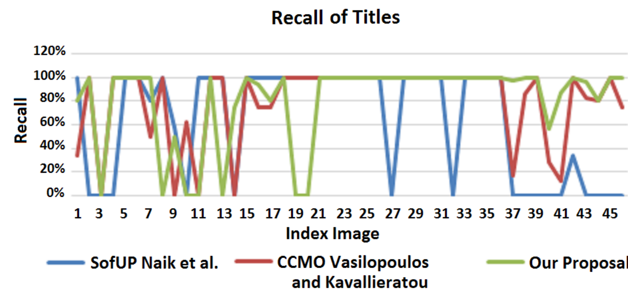
[6] S. Ferilli, M. Biba, F. Esposito, and T. M. A. Basile, "A distance-based technique for non-manhattan layout analysis," in *2009 10th International Conference on Document Analysis and Recognition*, July 2009, pp. 231–235.

[7] N. Vasilopoulos and E. Kavallieratou, "Complex layout analysis based on contour classification and morphological operations," *Engineering Applications of Artificial Intelligence*, vol. 65, pp. 220–229, 2017.

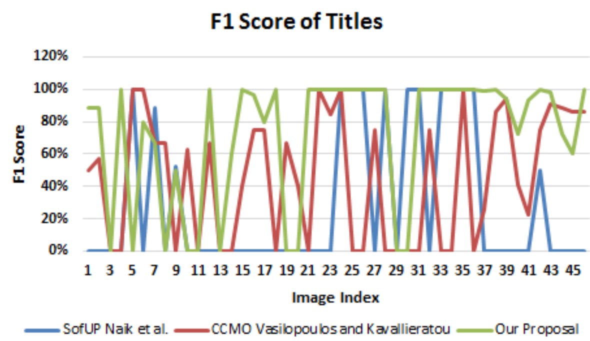
[8] V. Yadav and N. Ragot, "Text extraction in document images: highlight on using corner points," in *Document Analysis Systems (DAS), 2016 12th IAPR Workshop on*. IEEE, 2016, pp. 281–286.



(a) Precision.



(b) Recall.



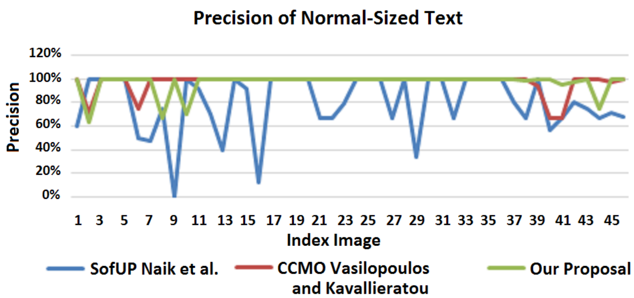
(c) F1 score.

Fig. 16. Precision, Recall and F1 metrics computed for the layout of titles.

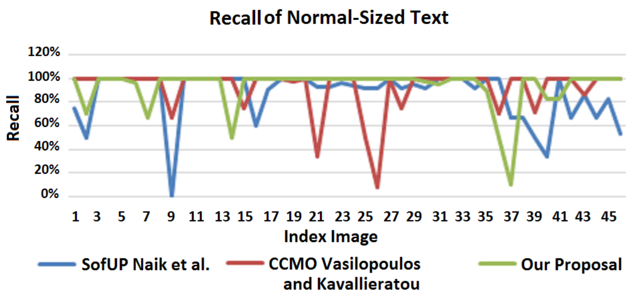
[9] T. A. Tran, K. Nguyen-An, and N. Q. Vo, "Document layout analysis: A maximum homogeneous region approach," in *2018 1st International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, April 2018, pp. 1–5.

[10] A. Corbelli, L. Baraldi, C. Grana, and R. Cucchiara, "Historical document digitization through layout analysis and deep content classification," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, Dec 2016, pp. 4077–4082.

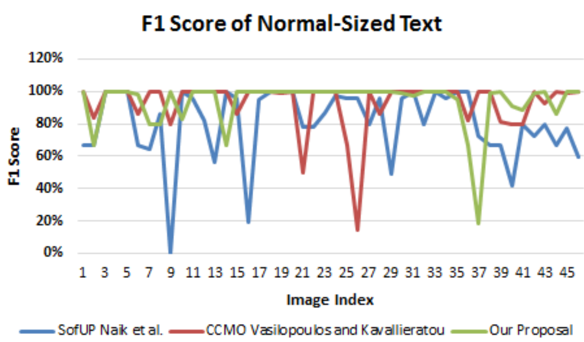
[11] N. Otsu, "A Threshold Selection Method from Gray-level Histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.



(a) Precision.

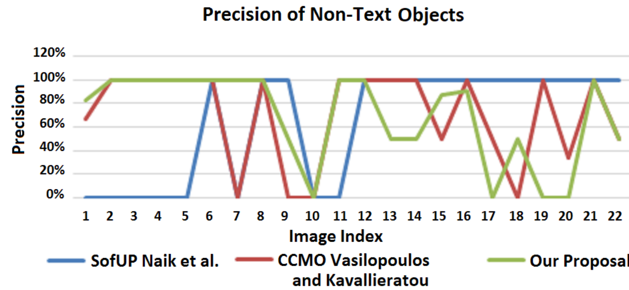


(b) Recall.

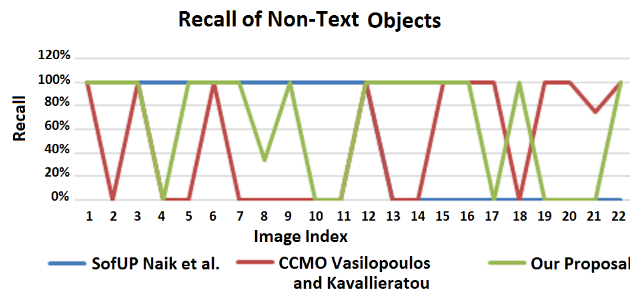


(c) F1 score.

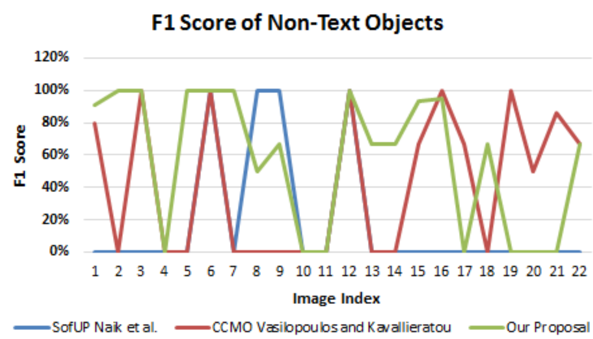
Fig. 17. Precision, Recall and F1 metrics computed for the layout of normal-sized text.



(a) Precision.



(b) Recall.



(c) F1 score.

Fig. 18. Precision, Recall and F1 metrics computed for the layout of non-text elements.

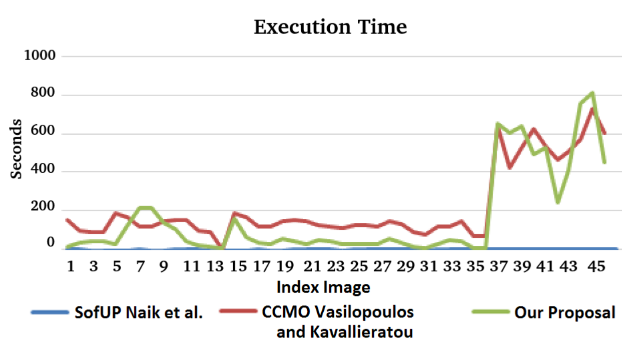


Fig. 19. Execution Time.