



Aim and Specific Objectives

- Aim:**
- This study to use of an automatic baseline detection technique, based on the use of Extreme Randomized using classifiers and clustering, for the automatic detection of baselines in Arabic handwritten documents.
- Objectives:**
- For handwritten text documents, recognition of their contents is one of the main tasks in document processing.
 - To get recognition usually needs segmentation of handwritten pages into lines.
 - A previous segmentation is needed among other techniques.
 - Automatic baseline detection technique for each line is one of the possibilities to obtain a line segmentation.



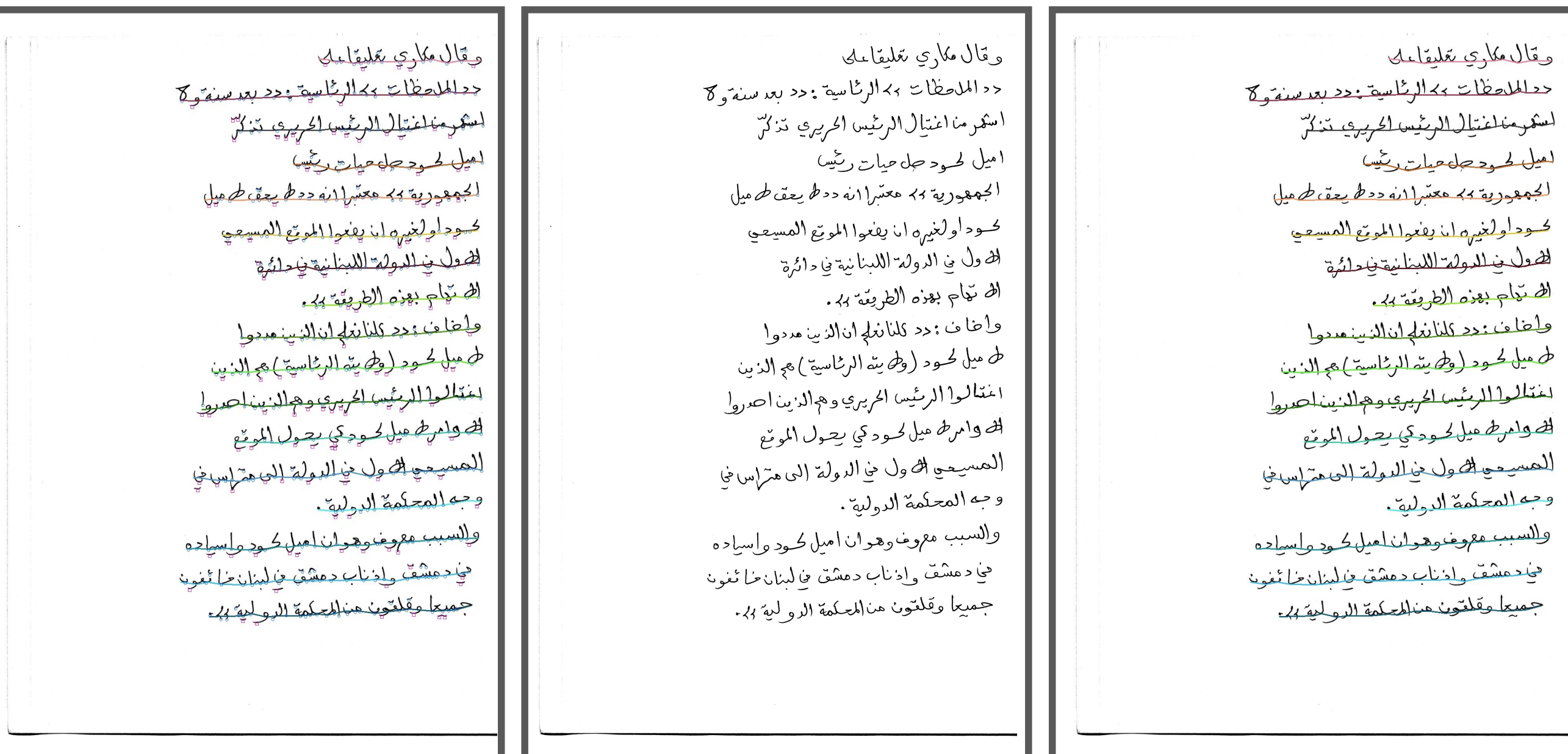
Introduction

- Document processing is one of the main interesting areas of digital image processing. Many document processing techniques are related to pattern recognition.
- The recognition of the contents of the handwritten is a task of high interest due to its applications on the automatic transcription of the documents.
- The solution of this problem is known as Handwritten Text Recognition (HTR) and most HTR applications require the previous segmentation of the document into lines.
- Baseline detection can be used to detect the number of lines of a page and to extract the different lines that can be used for handwritten text recognition



Handwritten Text and Possible Uses

- This technique has been previously tested in other types of scripting, but not for Arabic. Arabic features are different from classical Latin scripting or other Languages.
- We expect to the experiments reveal that this technique provides promising results for this task.

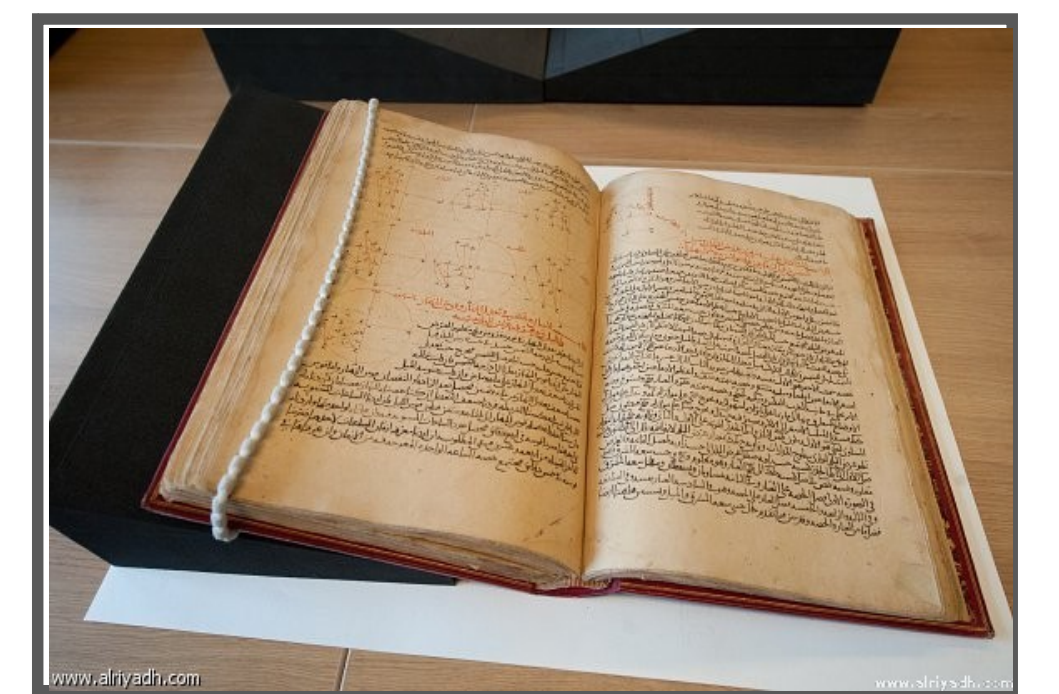


Sample pages from the dataset that we used it.

- The results of this research could be used to create assistance systems of handwriting.

Some utilities might be:

- 1- The results of this research could be used to create assistance systems of handwriting.



<http://www.ancientresource.com>



<http://www.wacom.com>

- 2 - Used also in others applications during the drafting of documents with graphics tablets or touch Screens.



Conclusion

- We have tested with the technique for the behavior of a baseline detection on Arabic handwriting.
- The baseline detection was performed with the use of Extremely Randomized clustering to create the corresponding baselines.
- Results show that the proposed technique provides good results in Arabic handwritten text
- This technique can be applied in a realistic framework where initially only a small amount of training data is manually annotated.



Main References

- 1- Likforman-Sulem, L., Zahour, A., Taconet, B.: Text line segmentation of historical documents: a survey. International Journal of Document Analysis and Recognition (IJ DAR) 9(2-4) (2007) 123-138.
- 2- Romero, V., Sanchez, J.A., Bosch, V., Depuydt, K., de Does, J.: Influence of text line segmentation in handwritten text recognition. In: Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, IEEE (2015) 536-540.



Contact Info



Ahmed Fawzi Ali
Email: ahfawal@dsic.upv.es
Phone: (+34)631653688

Supervisor

Dr. Carlos D. Martínez Hinarejos
Email: cmartine@dsic.upv.es
Phone: 963877069 (Ext. 73529)

Dr. Moisés Pastor
Email: mpastorg@dsic.upv.es
Phone: 963873518 (Ext. 73518)