|  |  |  |
| --- | --- | --- |
| **Credit Hours System**  **CMPN450**  **Pattern Recognition and Neural Networks** |  | **Cairo University**  **Faculty of Engineering** |

OCR

Project milestone 1

Submitted to: Eng. Hussien Fadl

|  |  |
| --- | --- |
| Student Name | ID |
| Ahmed Osama | 1152151 |
| Mina Ashraf | 1152055 |
| Omar Yousry | 1152065 |
| Wael Ashraf | 1152195 |

Submission Date: 10 / 12 / 2019

**Summary:**

We have reached a middle point where we segment the text into lines and lines into words and began character segmentation. Its accuracy is not that bad somehow. The left processes are feature extraction and building our model at the end.

**Papers used:**

* Character segmentation: An efﬁcient, font independent word and character segmentation algorithm for printed Arabic text (KING SAUDI UNIVERSITY)
* Feature Extraction: Using SIFT Descriptors for OCR of Printed Arabic (TEL AVIV UNIVERSITY)

**Algorithms used till now:**

1. FindBaselineIndex(): To find base line of the text
2. FindMaximumTransitionIndex(): To find line of maximum transitions of pixels from black to white and vice versa
3. CutPointIdentification(): To determine cut points in the word
4. SeparationRegionFilteration(): To determine the valid separation regions in the word and segment the characters

**Screenshots:** Grey points in right photo determine the valid separation regions till now for character segmentation. The accuracy is not perfect by the way.

