REFERENCES:

* <http://www.psut.jo/sites/osama/4.pdf>
* <http://www.tandfonline.com/doi/abs/10.1080/17489720701781905>
* <http://www.engineering.nottingham.ac.uk/icccbe/proceedings/pdf/pf49.pdf>
* <https://www.google.com/patents/US7069240>
* <https://www.google.com/patents/US20100331043>
* <http://www.hcis-journal.com/content/3/1/2>
* <http://www.sciencedirect.com/science/article/pii/S1574119210000416>

OCR REFERENCE TEXTS:

Tesseract is used for OCR. Tess4j is an extension of its implementation for Java Language. Tesseract is there for over 20 years now and is considered best for OCR :P.

sudo apt-get install tesseract-ocr

tesseract image.format outputfile // For execution

* <http://opencv-code.com/tutorials/how-to-integrate-tesseract-ocr-and-opencv/>
* <http://tess4j.sourceforge.net/tutorial/>
* <http://tess4j.sourceforge.net/codesample.html>
* <https://code.google.com/p/tesseract-ocr/>
* <http://staffhome.ecm.uwa.edu.au/~00082689/papers/Shafait-efficient-binarization-SPIE08.pdf> (Related to optimization of local binarization) \*This one seems to be quite interesting to me for binarization of image (It’s conversion to 0’s and 1’s).
* <http://www.ijcsit.com/docs/Volume%205/vol5issue02/ijcsit20140502254.pdf> (OCR algorithm and the entire process from scratch.. Not sure about implementation)