CHAPTER 3 RESEARCH METHODOLOGY

1. Literature Study

The process of collecting and reading were 10 journals relating to text detection algorithms using ocr, libraries Tesseract and by different methods. All of this journal as a reference for projects and programs that will be created.

2. Review Topic

This project is a text detection or a converter from image to text so that the dataset used is in the form of an image itself which wants to be converted into a txt file, but as an example I provide 10 datasets in the form of text images The image will be taken and will experience various pre-processing processes such as image denosing, erosion and also so that it can be better read by the tesseract itself.

3. Program design

Make a code program to detect text in the image then the image will be eliminated Noise so that the image is clearer, each word will also be in an erosion so that no word that is tested and then the algorithm will work to convert the text on the image into a digital document that can be processed in the notepad output will automatically appear on the notepad.

4. Coding

This program uses Python as a programming language for applying the pre-processing and detect text in the image. algorithms used in this program is ocr algorithm that converts the text in the image into a digital document in the form of a txt file that can be reprocessed.

5. Implementation and Analyze

This uses 100 sample images, where there is a text image from Google, there is also a text image taken from my own photo.

6. Conclusion

So the conclusion of the end of this project is to prove that how important the preprocessing process is for tesseract in detecting and converting text.

