CHAPTER 3
RESEARCH METHODOLOGY

1. Identification Problems and Literature study

The first step of the research is to use OCR (optical character recognition) to save documents from printed paper into a file format that allows for easier document storage and for the process of reprocessing documents. research was obtained from journals regarding problems of text recognition and text extraction, which were sourced from google scholar articles from 2005 to 2011. The journal discusses the implementation of google OCR (optical character recognition) on several devices such as cellphones or computers through image processing on OCR (optical character recognition).

2. Structure algorithm

LSTM (long short term memory) is an algorithm that is almost the same as rnn (recurant neural network), LSTM (long short term memory) and RNN (recurant neural network) have the same architecture but in LSTM (long short term memory) it has a little additional activison function in the form of cells. First part namely calculating the new input data with ht-1, namely the data obtained from the output the previous one and will produce Ft (forgate) which in the future will determine whether the old memory can be used again or not. The second part is (store) which will be used to create new memory which will later be determined whether the memory can replace curent memory used before. In the store process, two steps are needed, namely calculating Çt = tanh (Xt * Uc + ht-1 Wc) and it = σ (XtUi + ht-1Wi) The third part of Ct to compare old memory with new memory Ct = σ (ft * Ct-1 + it * Çt). If the old data has a greater value (closer to 1) then

old data will be used again. The third part makes the prediction and output value which will be reused in next (ht-1). In the process of predicting the LSTM character ht = tanh (Ct) * ot. On process.

3. Pre processing image dataset

in the pre-processing image dataset, the image will be divided into 2 parts. the first part is used to train the second part is used for the test part. the train section has 15.015 images containing the alphabet character letters containing letters a to z and number 0 to 9. In train section has 9006 images containing the alphabet image letters a to z and number 0 to 9. In the pre-processing, the image train and the image test will be resized to an image size of 50 so that the incoming image has the same size. After that, each image test and image train will be identified where the image train will be labeled X and the image test will be labeled Y. Each image train and image test will be given a grayscale filter from open cv and after that it will be convert to an array value which still has 3 channels which means 3 columns and 3 rows. To simplify, normalization is done which will make each image only have 2 channels, which means it has 2 columns and 2 rows. Changing the image to be 2 channels (black & white) is done by dividing the image by the maximum value (255) of pixels. Furthermore, each image will be returned to its original size. And converted to an array value using numpy and after that it will be saved using a binary format with a pickle library which will be used as an input algorithm. The results of the train and test labeling will be saved with X_train and Y_test.

4. Implementation algorithm with structure

In applying an algorithm to train data, the first step is to load X_train, Y_test with binary format. After that enter the feature extraction image an image feature extraction will be carried out which aims to find the important parts of each image that will be trained in the algorithm. The image extractor is in the form of a batch normalization(making image train to have same width and height), convolutional layer that has conv2d (convolutional layer), max-pooling (aims to find important parts of each pixel in the train image), using relu activation (linear rectifier). use a convolutional layer to find the important parts of the image and then maxpooling to save the most important parts of the image with a kernel size [1,2,2,1] which

means the maxpooling kernel that will be run is 2x2. After the image feature extraction process is carried out, it will enter the algorithm, LSTM (long short term memory) is the development of the RNN (recurrent neural network) in the recurrent neural network algorithm section. 128 hiden layer which serves to recognize images. After that, then enter the LSTM cell in the LSTM cell, the image recognition method will be carried out using the stored memory as a benchmark so that the previously recognized image value will be saved and will be compared with the incoming image input. The next process, the train image will be identified with a test image. At this stage the algorithm process recognizes the train image and the test image will be carried out as much as epoch = 10 using the Adam optimizer and loss calculations will be carried out to calculate the inaccuracies in making the fit x, y model to be saved using the tensorflow format, the checkpoint model that will store the model on a per step basis while in the train.

5. Text detection & text ekstraction

In the detection model text that has been saved, it will be loaded again to recognize the character in the input image and make a prediction for the character of the letters in the input image. The detction text will be assisted by the Opencv library which will create a bounding box, namely as an area for detecting text in images, opencv will use the ROI (region of interest) method which uses contour calculations in the image. the first step the image will be input using im.write then the image will be given a grayscale filter, then using otsu binarization and dilation using the bounding box is done by calculating the width, height and the pixel difference between the background and text. after that opencv will create a pointer line that shows the location of the letters in the input image. for text extarction, the input image will be recognized by the loaded model. And modle will recognize letter patterns and text from the input image. What is detected will enter the predict function which has characterset from A to Z and numbers 0 to 9. The

function will print the letters that are recognized from the model and save the predicted text results in the document file format (.txt).

6. Evaluation

The evaluation method uses a step by means of the pytesseract model being trained again with the existing NIST dataset, namely for the 9006 train and for the 6009 test. Then the model from the pytesseract and the save model that has been prepared will both detect the same image to test the level of accuracy. Testing was done with the following formula . $Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)}$

To test the LSTM (long short ter memeory) algorithm, the x_test, y_train data are input without using the image feature extraction with a convolutional layer. The data set will also be reduced to train = 6017 and test 5000, train = 3000 and test = 3000. This experiment aims to test how good the LSTM (long short term memory) algorithm is in predicting from a minimal dataset.