

Recognition of the character on the map captured by the camera using k-nearest neighbor

B Harijanto^{*}, E L Amalia and M Mentari

Dept. of Informatics Engineering, State Polytechnic of Malang, Malang, Indonesia

*budi.hijet@gmail.com

Abstract. Maps are one form of an image that often encountered in various interests. For example, many books on tourist attractions or other information that provide maps as a media of information. However, sometimes people with visual impairments such as presbyopia, hypermetropia, or astigmatism have difficulty reading the map which is usually given in small size, multi orientation (multi-scale and multi-direction). Therefore, this study tries to provide a solution through an application that converts image to text in image conditions that have many orientations and different things are briefly called heterogeneous text. The Optical Character Recognition (OCR) system that was built beginning with taking pictures made through a cell phone camera as the first step in obtaining a digital map file, then enters the pre-processing, text segmentation, feature extraction from each different character, then continues to the classification stage. This OCR system for recognizing text with multiple orientations will help people make digital maps easier to read, especially for people who have presbyopia, hypermetropia, or astigmatism vision problems. The proposed model achieves good average accuracy for classifying the characters in various orientation successfully.

1. Introduction

Documents packaged in the form of books, articles or in other forms are very important in the needs of all sides of life. One form of a document is a map. Maps are one form of an image that often encountered in various interests. For example, many books on tourist attractions or other information that provide maps as a media of information. However, sometimes people with visual impairments such as presbyopia, hypermetropia, or astigmatism have difficulty reading the map which is usually given in small size, multi orientation (multi-scale and multi-direction).

Some studies discuss similar things that explains the great advantages of digitizing documents that come from scanners or digital cameras [1]. This study discusses the decline in image quality caused by uneven lighting variations. To overcome this, a new approach is implemented which is implemented by improving the image to locate the text area. With the location information of the area, the background lighting is formed to improve the balanced final lighting. The simulation results demonstrate that this approach is superior to previous research.

The second study developed an efficiency technique for segmenting document pages that resulted from digitizing historic printouts [2]. This type of document often has problems with low quality local biases, some degradation due to the influence of old prints or ink spreading, showing complexity, and layout density. To deal with this problem, the Novel Adaptive Run Length Smoothing (ARLSA) method is used to deal with problems of layout complexity and density. Detection of noise areas, punctuation



that is normally used on typewriters, detects possible obstacles because the background area is used to separate text columns or text lines.

Documents in the form of images sometimes also contain not only writing but also contain pictures or graphics. To convert a written image document into digital writing, the first method is used to separate or delete image or graphic information that is not used. Some related studies have carried out the separation which uses the method of multi-plane segmentation, to perform textual segmentation and extraction of objects from various types of complex image documents in real life [3]. This method initially decomposes image documents into different flat objects to extract and separate homogeneous objects including textual regions of interest, objects other than writing such as graphics and images and background textures.

Subsequent research also discussed the same problem [4]. This study discusses the extraction and identification of text-lines from various drawing documents consisting of writing mixed with graphics or images. The method proposed in this research will describe the document image into different object fields (text, not text or background). The proposed system shows results with high flexibility and efficiency in the extraction of writing lines with various lighting, sizes, font types, and the complexity of blending writing with images.

After separating the writing from non-writing, digitizing the document documents will detect the writing lines and words as what has been done by the research [5]. This study presents the segmentation method of handwritten documents on different entities, i.e., written lines and words. Text line segmentation is obtained by applying Hough Transform to an overlapping subset of image documents. The post-processing stage includes correction of possible errors, detection of writing lines that failed to be overcome by Hough Transform which works by separating the vertical connection characters using a skeletonization-based development method.

Every word that has been detected in an image document still needs to go through another process until it is finally recognized as digital writing. This research has been carried out in research [6]. This research presents an algorithm of detection or detection and recognition of vehicle license plates with a dataset in the form of video traffic. The method used to detect vehicle license plates is the edge detection feature and also the local Haar feature. While the characters in the image are extracted using the blob detection development method to remove unwanted areas. The introduction of a vehicle license uses a modified OCR source. The results of the proposed method show a strong system despite bad lighting and moving vehicles.

The introduction of multi-oriented and multi-size texts was carried out in research [7]. The case study taken in the study is a map image that has writing of different sizes and different positions or directions. This will require writing recognition techniques that are more difficult than homogeneous writing.

2. Literature review

2.1. Optical Character Recognition (OCR)

OCR is a computer system that can read letters, both from handwriting or a printing machine (printer or typewriter). OCR is widely used to convert document types, such as document scans, PDF files or images taken by digital cameras to edited and searchable data.

2.2. Line detection

Detection of margins in an image is a process that produces edges of image objects. The goal is to mark the part of the image detail / image to improve the details of the image / image that is blurred, which occurs because of the effects of the image acquisition process. A point (x, y) is said to be the edge (edge) of an image if the point has a high difference with its neighbors.

2.3. Character segmentation

Character segmentation is the process of dividing regions in an image to be grouped into certain segments. In general, character segmentation is done to separate characters from the background in a

digital image. Images consisting of letters, numbers or whole sentences will be separated into characters that will stand alone and do not become part of a particular sentence so that they can be recognized.

2.4. *Thinning*

Thinning is the process of reducing an object in a digital image to a minimum size (object (region) reduced to skeleton). Thinning is a morphological operation that is used to remove the front of the pixel from a binary image, which can be used in several applications but is very useful for skeletonization. Thinning is only used in binary images and produces other binary images as output. Thinning is a form of "pre-processing" used in many image analysis techniques. The output of this process is referred to as "skeleton", therefore thinning can also be referred to as "skeletonization".

2.5. *The k-nearest neighbor algorithm (kNN)*

The k-nearest neighbor algorithm (k-NN or KNN) is a method for classifying objects based on learning data that is the closest distance to the object. Learning data is projected into multi-dimensional space, where each dimension represents the features of the data. In the learning phase, this algorithm only stores feature vectors and classifications of learning data. In the classification phase, the same features are calculated for the test data (whose classification is unknown). The distance of this new vector to all the learning data vectors is calculated, and the closest number of k is taken. The new points are predicted to be included in the most classifications of these points.

The best k value for this algorithm depends on the data; in general, a high k value reduces the effect of noise on the classification, but makes the boundaries between each classification more blurred. A good k value can be chosen with parameter optimization, for example by using cross-validation. Special cases where classification is predicted based on the closest learning data (in other words, k = 1) is called the nearest neighbor algorithm. Distance search using the Euclidean distance method in k- can be seen in the following equation

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \quad (1)$$

Where a = a₁, a₂, ..., a_n, and b = b₁, b₂, ..., b_n represents the n attribute values of the two records.

2.6. *Validation method*

Accuracy is a measurement or validation of the truth of the method used. The performance measurement uses confusion matrix, there are 4 (four) terms as a representation of the results of the classification process. The four terms are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). True Negative (TN) value is the amount of negative data that is detected correctly, while False Positive (FP) is negative data but is detected as positive data. Meanwhile, True Positive (TP) is positive data that is detected correctly. False Negative (FN) is the opposite of True Positive, so the data is positive, but detected as negative data. The accuracy formula can be seen as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (2)$$

3. Research methods

3.1. *Dataset*

The study was conducted using a dataset in the form of Map image data, which captured using the Samsung Galaxy A20 13MP phone camera, and taken from has titled Kompas Klass. 2017. 35 Destinations, Travel Bucket List. Kompas. The pictures in the book totaled 8 maps pictures in various slopes. For example, can be seen as follows:



Figure 1. Sample images of the 6th map studied.

Each picture used consists of several words, ranging from at least around 24 words to over 80 words. Then each word has characters that vary from 1 to more than 20 characters. The total test data used consists of about more than 1000 characters.

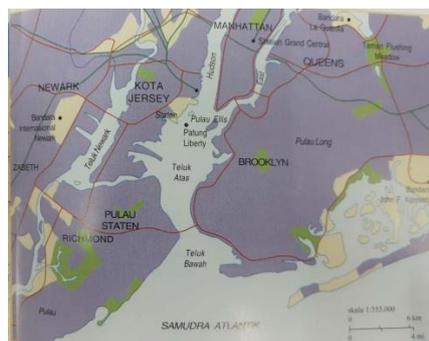


Figure 2. Sample images of the 1st map studied.

3.2. Research step

Figure 3 explains the flow that is carried out starting from the data entry stage in the form of images to the data out in the form of text. After the image is available, the initial stage is pre-processing to prepare image data that can be processed without noise and minimize object recognition errors. Pre-processing will also improve image quality edge detection of an image to obtain the coordinate axis reference for starting pixel image processing.

After pre-processing in figure 3 explained that the color filter is applied to the color range of the text. After that the next process is carried out for MSER Regions with the maximum constants according to the map image used. After that the MSER regions detection results still allow non-text areas to be recognized as text, therefore this non-text portion will be removed from the focus object. After that proceed with the process of thinning and deleting the focus area of objects other than text with data output from thinning. The bounding box for each character is then generated, and the character's closeness is calculated with the overlap detection system. Overlap characters will be considered as one word. This one word will also be generated in the bounding box and then continued to take morphological features and invariant moments as input to the classification process.

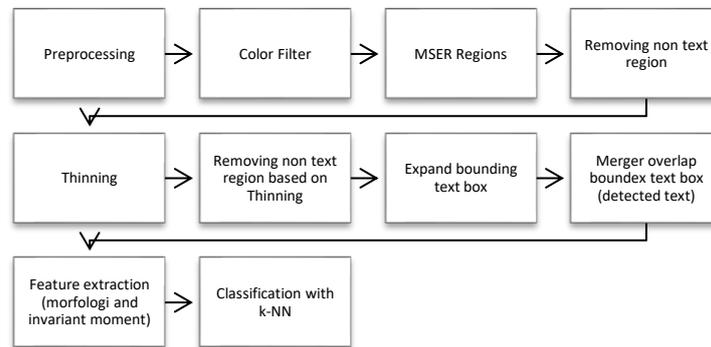


Figure 3. System design.

4. Results and discussion

The first step taken is to capture the area that is considered to be a character located. Then check again and estimate which areas of the character area in the previous process are not true characters (incorrect detection). After meeting the non-character region, then delete the area. After that a merger is done for the adjacent characters. The final stage is character detection using k-NN classification. This classification needs to be done because the direction of writing varies, not always horizontal. The average accuracy obtained in this study is under 50% for $k=24$. This is because the average word in the map image is too close, so it is difficult to break down each character. So the character is truncated imperfect (one cutting area can consist of more than one character). For the results of cutting words consisting of a few characters, ranging between 1, 2 and 3 characters. Then the detection results character reaches 100%. Figure 5 is detecting text based on the box around the characters on the map one example of the original image input (figure 2).

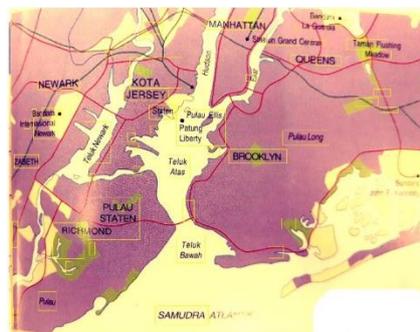


Figure 4. Detected text.

5. Conclusion

Character recognition in this study requires an extra process at a pre-processing in order to get a good standard image. This is because the shooting is done with unstable light. Even some photos seem unclear. Besides the symbols on the map are sometimes recognized as characters, for example, the circle symbol is recognized as o. This will also make the accuracy is still less than 50%. The accuracy detection character can reaches 100% for words which consists of only a few characters. For further research, the choice of k values is more precise and varied. In addition, the process of detecting bounding boxes for each character, then the process of overlap bounding boxes to determine the relationship between letters into one word must also be improved.

Acknowledgments

Authors want to thank all colleagues from State Polytechnic of Malang in the Information Technology Department, State Polytechnic of Malang for the participation in this research.

References

- [1] Chen K N 2012 Efficient illumination compensation techniques for text images *Digital Signal Processing* **22** 726-733
- [2] Nikolaou N 2010 Segmentation of historical machine-printed documents using Adaptive Run Length Smoothing and Skeleton segmentation path *Image and Vision Computing* **28** 590-604
- [3] Chen Y L 2009 A multi-plane approach for text segmentation of complex document images *Pattern Recognition* **42** 1419-1444
- [4] Chen Y L 2012 A knowledge-based system for extracting text-lines from mixed and overlapping text/graphics compound document images *Expert System with Applications* **39** 494-507
- [5] Referensi Teknologi dan Elektronika Indonesia *Image Processing* [Online] Retrieved from: <https://ndoware.com/image-processing.html> accessed on 1 November 2018
- [6] Louloudis G 2009 Text line and word segmentation of handwritten documents *Pattern Recognition* **42** 3169-3183
- [7] Chiang Y Y Recognition of Multi-Oriented, Multi-Sized, and Curved Text *International Conference on Document Analysis and Recognition* Reference lists