ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology
Volume 1, Issue 4, June 2012

Text Extraction From Images

Satish Kumar[1], Sunil Kumar[2], Dr. S. Gopinath[3],

[1] Research Scholar of Singhania University, India, Reg. No.0950102580
satishcrrit@rediffmail.com
[2]Maharaja Agrasen Institute of Technology, GGSIPU
Sunil.mathur10@gmail.com
[3] ABB GISL, Corporate Research, Bangalore, India
Email: selgopi@gmail.com

Abstract—**Images and videos on webs and in databases are increasing. It is a pressing task to develop effective methods to manage and retrieve these multimedia resources by their content. Text, which carries high-level semantic information, is a kind of important object that is useful for this task.When a machine generated text is printed against clean backgrounds, it can be converted to a computer readable form (ASCII) using current optical character recognition (OCR) technology. However, text is often printed against shaded or textured backgrounds or is embedded in images. Examples include maps, photographs, advertisements, videos, etc. Current document segmentation and recognition technologies cannot handle these situations well.**
Our system takes advantage of the distinctive characteristics of text that make it stand out from other image material i.e. text possesses certain frequency and orientation information; text shows spatial cohesion—characters of the same text string (a word, or words in the same line) are of similar heights, orientation, and spacing.

Keywords- binarization, connected components, filters, text reading system.

I. INTRODUCTION

Today the most information is available either on paper or in the form of photographs or videos. Large information is stored in images. The current technology is restricted to extracting text against clean backgrounds. Thus, there is a need for a system to extract text from general backgrounds. There are various applications in which text extraction is useful. These applications include digital libraries, multimedia systems, Information retrieval systems, and Geographical Information systems. The role of text detection is to find the image regions containing only text that can be directly highlighted to the user or fed into an optical character reader module for recognition.

In this paper a new system is proposed which extracts text in images. The system takes colored images as input. It detects text on the basis of certain text features: text possesses certain frequency and orientation information; text shows spatial cohesion—characters of the same text string (a word, or words in the same line) are of similar heights, orientation, and spacing. The image is then cleaned up so that the text stands out.

II. RELATED WORK

Various methods have been proposed in the past for detection and localization of text in images and videos. These approaches take into consideration different properties related to text in an image such as color, intensity, connected-components, edges etc. These properties are used to distinguish text regions from their background and/or other regions within the image. In the algorithm based on color clustering, the input image is first pre-processed to remove any noise if present. Then the image is grouped into different color layers and a gray component. This approach utilizes the fact that usually the color data in text characters is different from the color data in the background. The potential text regions are localized using connected component based heuristics from these layers. Also an aligning and merging analysis (AMA) method is used in which each row and column value is analyzed. The experiments conducted show that the algorithm is robust in locating mostly Chinese and English characters in images; some false alarms occurred due to uneven lighting or reflection conditions in the test images.
The text detection algorithm in [6] is also based on color continuity. In addition it also uses multi-resolution wavelet transforms and combines low as well as high level image features for text region extraction. Texture based segmentation is used to distinguish text from its background. Further a bottom-up ‘chip generation’ process is carried out which uses the spatial cohesion property of text characters. The chips are collections of pixels in the image consisting of potential text strokes and edges. The results show that the algorithm is robust in most cases, except for very small text characters that are not properly detected. Also in the case of low contrast in the image, misclassifications occur in the texture segmentation.
A focus of attention based system for text region localization has been proposed before. The intensity profiles and spatial variance is used to detect text regions in images. A Gaussian pyramid is created with the original image at different resolutions or scales. The text regions are detected in the highest resolution image and then in each successive lower resolution image in the pyramid.

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology
Volume 1, Issue 4, June 2012

Some approaches utilize a support vector machine (SVM) classifier to segment text from non-text in an image or video frame. Initially text is detected in multi scale images using edge based techniques and morphological operations. These detected text regions are then verified using wavelet features and SVM.

III. Method

A. *Converting colored image to grayscale*

A digital color image is a color image that includes color information for each pixel. There are various color models which are used to represent a color image. These are RGB color model, in which red, green and blue light is added together in various ways to reproduce a broad array of colors. The other models are CMY color model which uses cyan, magenta and yellow light and HSI model which uses hue, saturation and intensity variations.

Grayscale images have range of shades of gray without apparent color. These are used as less information needs to be provided for each pixel. In an 8 bit image, 256 shades are possible. The darkest possible shade black is represented as 00000000 and lightest possible shade white is represented as 11111111.

B. *Binarization*

A Binary image is a digital image that can have only two possible values for each pixel. Each pixel is stored as single bit 0 or 1. The name black and white is often used for this concept. To form a binary image we select a threshold intensity value. All the pixels having intensity greater than the threshold value are changes to 0 (black) and the pixels with intensity value less than the threshold intensity value are changed to 1 (white). Thus the image is changed to a binary image.

C. *Connected components*

For two pixels to be connected they must be neighbors and their gray levels must specify a certain criterion of similarity. For example, in a binary image with values 0 and 1, two pixels may be neighbors but they are said to be connected only if they have same values. A pixel p with coordinates (x, y) has four horizontal and vertical neighbors known as 4-neighbors of p, given as: (x, y+1), (x, y-1), (x+1, y), (x-1, y) and four diagonal neighbors given as: (x+1, y-1), (x-1, y-1), (x+1, y+1), (x-1, y+1). Together these are known as 8-neighbors of p. If S represents subset of pixels in an image, two pixels p and q are said to be connected if there exists a path between them consisting entirely of pixels in S. For any pixel p in S, the set of pixels that are connected to it in S is called a connected component of S.

D. *Horizontal and Vertical Projections*

The method is performed on binary images. It starts scanning from left side of every line and records a change in case of facing the pixel change from zero to one and again to zero. Counting the change does not depend on number of pixels in this method. Robustness in noisy condition is an advantage of this method.

After finding the connected components, we check the transitions in the values of pixels horizontally. Transitions can be either zero (black) to one (white) or one (white) to zero (black). The text regions will have larger number of transitions from black to white or vice versa, whereas the background region will have lesser number of transitions.

If the allocated amount of changes for each row is between two thresholds (low and high thresholds), the row potentially would be considered as text area and the up and down of this row would be specified. Next, we search vertically for finding the exact location of the text and ignoring these rows as a text.

For finding the exact location of the text, we use some heuristics. These heuristics include height and length of the text and the ratio of height to length and enough number of pixels in this area.

E. *Reconstruction*

After the extraction of text regions from images, the text regions become a bit distorted and difficult to read, thus we recover these components using the original image. The distorted and original images are compared with each other and the pixels which are erased or disfigured are recovered.

IV. Conclusion

Current technologies do not work well for documents with text printed against shaded or textured backgrounds or those with non-structured layout. In contrast, our technique works well for normal documents as well as with the documents described above situations. The system is stable and robust. All the system parameters remain the same throughout all the experiments.

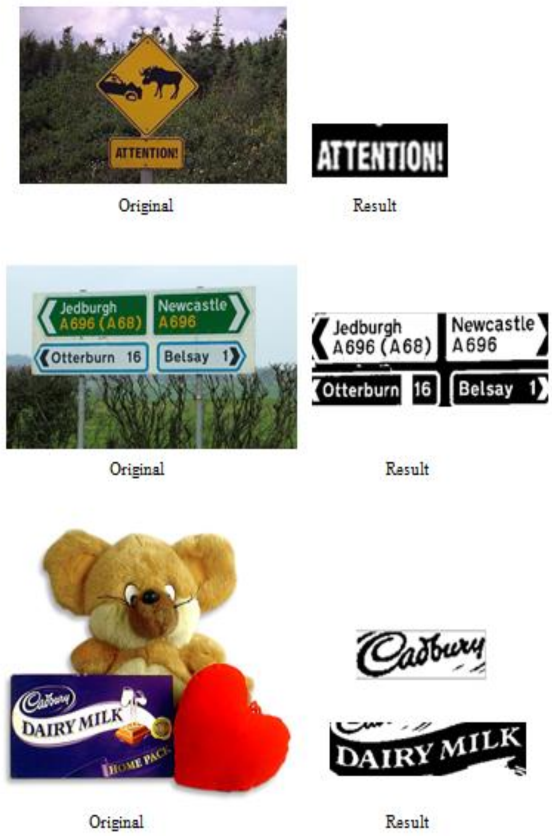

Figure 1. Text Extraction Results

ACKNOWLEDGMENT

This paper gives us an opportunity to convey our gratitude to all those who have helped us to reach a stage where we have immense confidence to launch our career in the competitive world of electronics and communication.

We would like to thank Mr. Sunil Mathur who provided us the much needed guidance on every step.We will always be grateful to him who has been mentoring us throughout.Last but not the least we would like to thank our family and friends who encouraged us and advised us constantly.

REFERENCES

[1] Victor Wu, R. Manmatha, Edward M. Riseman.:“Text Finder: An Automatic System to Detect and Recognise Text in Images.”
[2] Miriam León, AntoniGasull.:“Text Detection in Images and Video Sequences.”
[3] Sneha Sharma.:“Extraction of Text Regions in Natural Images.”
[4] KhaledHammouda, Prof. ED Jernigan.:“Texture Segmentation using Gabor Filters.”
[5] HaidarAlmohri, John S. Gray, HishamAlnajjar.:” A real-time dsp-based Optical Character Recognition System for isolated arabic characters.”
[6] AmerDawoud and Mohamed Kame1. : ”Binarization of Document Images Using Image Dependent Model.”
[7] J. Gllavata, R. Ewerth and B. Freisleben, “Finding Text in Images via Local Thresholding”, International Symposium on Signal Processing and Information Technology, Darmstadt, Germany, 2003.
[8] Q. Ye, Q. Huang, W. Gao and D. Zhao, “Fast and robust text detection in images and video frames”, Image Vision Comput. 23 (2005).
[9] G. Aghajari, J. Shanbehzadeh, and A. Sarrafzadeh, “A Text Localization Algorithm in Color Image via New Projection Profile”