International Journal of Latest Trends in Engineering and Technology
Vol.(7)Issue(4), pp.137-141
DOI: http://dx.doi.org/10.21172/1.74.018
e-ISSN:2278-621X

RESEARCH PAPER FOR ARBITRARY ORIENTED TEAM TEXT DETECTION IN VIDEO IMAGES USING CONNECTED COMPONENT ANALYSIS

Kiran[1] and Neeraj Jhulka[2]

Abstract : An end-to-end system for text detection and recognition is important in multiple domains such as content based retrieval systems, video event detection, human computer interaction, autonomous robot or vehicle navigation and vehicle license plate recognition. There are several commercial systems for text recognition in scanned document. However, these systems typically need cropped and binarized text regions to perform well for natural scene text. Text detection in natural scenes is a challenging problem and has gained a lot of attention recently. Such texts present low contrast with background, large variation in font, color, scale and orientation combined with background clutter. Therefore a robust and fast recognition system is desirable. In this work, we have explored the dominant text detection in video frames in which variety of images has been selected which has arbitrary alignment of text as well as varying color and size of the text. In this work, we have used wavelet based laplacian filters as well as laplacian of Gaussian filters to get the arbitrary directed text edges in a frame. After that fuzzy clustering has been applied to get the foreground text edges on which dilation and erosion operations are applied to get the pixels containing whole text. After that MSER technique has been implemented along with connected component analysis to extract the prominent text pixels in the image. Experimental results show that the present technique gives good recall and precision ratio on the tested databases.
Keywords- Image, Pixels

I. INTRODUCTION
IMAGE PROCESSING

Text detection in natural scenes is a challenging problem has gained a lot of attention recently. Such texts present low contrast with background, large variation in font, color, scale and orientation combined with background clutter. Therefore a robust and fast recognition system is desirable. In this work, we have explored the dominant text detection in video frames in which variety of images has been selected which has arbitrary alignment of text as well as varying color and size of the text. In this work, we have used wavelet based laplacian filters as well as laplacian of Gaussian filters to get the arbitrary directed text edges in a frame. After that fuzzy clustering has been applied to get the foreground text edges on which dilation and erosion operations are applied to get the pixels containing whole text. After that MSER technique has been implemented along with connected component analysis to extract the prominent text pixels in the image. Experimental results show that the present technique gives good recall and precision ratio on the tested databases.

[1] *Department of Electronics and Communication Engineering, PTU, Jalandhar*
[2] *Department of Electronics and Communication Engineering, PTU, Jalandhar*

INDENTATIONS AND EQUATIONS

An image is an array or a matrix of square pixels (picture elements) arranged in columns and rows. In a (8-bit) grayscale image, each picture element has an assigned intensity that ranges from 0 to 255. A grayscale image is what people normally call a black and white image but the name emphasizes that such an image will also include many shades of gray. A normal grayscale image has 8 bit color depth = 256 grayscales. A "true color" image has 24 bit color depth = 8 x 8 x 8 bits = 256 x 256 x 256 colors = ~16 million colors. Some grayscale images have more grayscales, for instance 16 bit = 65536 grayscales.

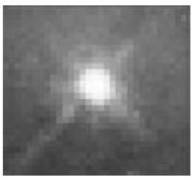


Figure 1.1: An image is an array or a matrix of pixels arranged in columns and row

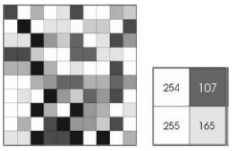


Figure 1.2: Each pixel has a value from 0 (black) to 255 (white).

An image may be defined as a two-dimensional function f(x, y), where x and y are spatial (plane) coordinates and the amplitude off(x,y) at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. When x, y and amplitude values of f are all finite, discrete quantities, then the image is called a digital image. The field of digital image processing refers to processing digital images by means of a digital computer. Note that a digital image is composed of a finite number of elements, each of which has a particular location and value. These elements are referred to as picture elements, image elements, pels or pixels. Pixel is the term most widely used to denote the elements of a digital image. A digital Image processing start with one image and produces a modified version of that image [1]. Digital image analysis is a process that transforms a digital image into something other than a digital image, such as a set of measurement data, alphabet text or a decision. Image digitization is a process that converts a pictorial form to numerical data.

II. CONCLUSION

Acknowledgements

In this work, we explored text detection in video by selecting dominant text pixels and text candidates with the help of the hybrid technique which uses number of algorithms to get the output results. Dominant text pixels are found in frames of the video using Lab color space in which connected component labeling has been used by using sMSER technique after applying wavelet laplacian as well as Laplace of Gaussian filters on input image. This dominant text pixel selection helps in removing non text information in complex background of video frames. Results showed that the proposed method works well for text detection irrespective of contrast, orientation, background, script, fonts and font size. However, the proposed method may not give good accuracy for text lines with no spacing between text lines. However it gives improved results than existing techniques in terms of segmentation as well as in terms of variety of text location in the video frames I.e. horizontal, vertical, rotatory etc. The existed algorithm can be improved further to eliminate the false positives found after segmentation. The main drawback of the presented technique is that it works only on the English test and the alphabets inside the image should be detached from one another.

As we have used the height and width of the individual alphabets, the algorithm fails on the Indian languages as it is hard to disconnect the individual letters in them. In future, there is a good scope to enhance it further for all types of text.

Binary Image: Each pixel is just black or white. Since there are only two possible values for each pixel, we only need one bit per pixel. Such images can therefore be very efficient in terms of storage. Images for which a binary representation may be suitable include text (printed or handwriting), fingerprints, or architectural plans. An example was the image shown in figure 1.4. In this image, we have only the two Colors: white for the edges, and black for the background. See figure 1.4 below.

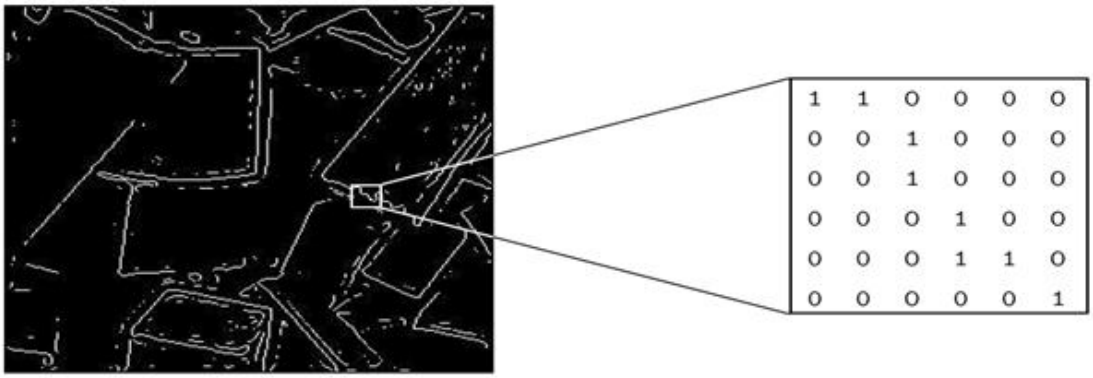


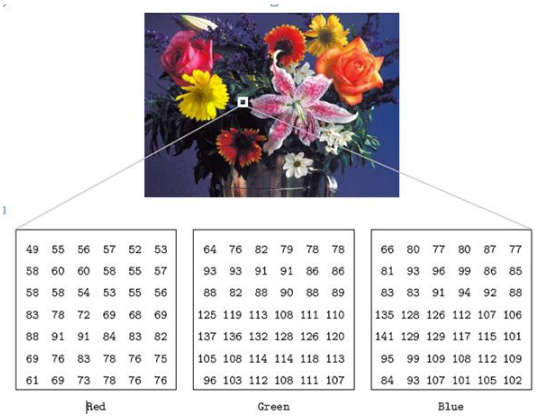
Figure 1.4: A binary image

2 Gray scale Image: Each pixel is a shade of grey, normally from 0 (black) to 255 (white). This range means that each pixel can be represented by eight bits, or exactly one byte. This is a very natural range for image _le handling. Other gray scale ranges are used, but generally they are a power of 2. Such images arise in medicine (X-rays), images of printed works, and indeed different grey levels are sufficient for the recognition of most natural objects.

Figure 1.5: A grayscale image

True color or RGB Image. Here each pixel has a particular Color; that Color being described by the amount of red, green and blue in it. If each of these components has a range 0-255 this gives a total of 255×255×255=16,777,216 different possible Colors in the image. This is enough Colors for any image. Since the total number of 24 bits required for each pixel is such images are also called 24-bit Color images.

Such an image may be considered as consisting of a stack of three matrices; representing the red, green and blue values for each pixel. This means that for every pixel there correspond three values.



Properties defining a text in video

Text in images can exhibit many variations with respect to the following properties :

1. Geometry:

• Size: Although the text size can vary a lot, assumptions can be made depending on the application domain.

• **Alignment:** The caption texts appear in clusters and usually lie horizontally, although sometimes they can appear as non-planar texts as a result of special effects. This does not apply to scene text, which has various perspective distortions.

4**. Inter-character distance**: characters in a text line have a uniform distance between them.

5. **Color**: The characters tend to have the same or similar colors. This property makes it possible to use a connected component-based approach for text detection. Most of the research reported till date has concentrated on finding 'text strings of a single color (monochrome)'. However, video images and other complex color documents can contain 'text strings with more than two colors (polychrome)' for effective visualization, i.e., different colors within one word.

6. **Motion:** The same characters usually exist in consecutive frames in a video with or without movement. This property is used in text tracking and enhancement. Caption text usually moves in a uniform way: horizontally or vertically. Scene text can have arbitrary motion due to camera or object movement.

7. **Edge:** Most caption and scene texts are designed to be easily read, thereby resulting in strong edges at the boundaries of text and background.

8. **Compression**: Many digital images are recorded, transferred, and processed in a compressed format. Thus, a faster TIE system can be achieved if one can extract text without decompression.

REFERENCES

[1] Bhavadharani R., Sowmya P., Thilagavathy A., "A Dynamic Approach to Extract Texts and Captions from Videos" IJCSMC, Vol. 3, Issue. 4, April 2014, pg.1047 – 1052.

[2] Choksi A., Desai N., Chauhan A., Revdiwala V., Patel K., " Text Extraction from Natural Scene Images using Prewitt Edge Detection Method" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 12, December 2013.

[3] Liu X., Wang W., "Robustly Extracting Captions in Videos Based on Stroke-Like Edges and Spatio-Temporal Analysis" IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 14, NO. 2, APRIL 2012.

[4] Murthy N., Kumaraswamy Y., "Robust Model for Text Extraction from Complex Video Inputs Based on SUSAN Contour Detection and Fuzzy C Means Clustering" International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011.

[5] Palma D., Ascenso D., Pereira F., "Automatic Text Extraction in Digital Video Based on Motion Analysis", Springer-Verlag Berlin Heidelberg 2004, pp. 588–596.

[6] Shivkumara P., Phan T., Tan C., "Gradient Vector Flow and Grouping-based Method for Arbitrarily Oriented Scene Text Detection in Video Images", IEEE transactions on circuits and systems for video technology, vol. 23, no. 10, october 2013.

[7] Shivakumara P., Sreedhar R., Phan T., Lu S., Tan L., "Multioriented Video Scene Text Detection Through Bayesian Classification and Boundary Growing" IEEE transactions on circuits and systems for video technology, vol. 22, no. 8, august 2012.

[8] Sonam., Kumar M., "Implementation of MD algorithm for Text Extraction from Video" Nirma University International Conference on Engineering IEEE 2013.

[9] Zhong Y., Zhang H., Jain A., "Automatic Caption Localization in Compressed Video" IEEE transactions on pattern analysis and machine intelligence, vol. 22, no. 4, april 2000.

[10] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, "Digital Image Processing Using MATLAB", Third Edition Tata McGraw Hill Pvt. Ltd., 2011

[11] Palaiahnakote Shivakumara, Trung Quy Phan, Chew Lim Tan," New Wavelet and Color Features for Text Detection in Video." Published in International Conference on Pattern Recognition 1051-4651/10 $26.00 © 2010 IEEE, DOI 10.1109/ICPR.2010.972.

[12] M Sharmila Kumari, and B H Shekar, "On the Use of Moravec Operator for Text Detection in Document Images and Video Frames." Published in IEEE-International Conference on Recent Trends in Information Technology, ICRTIT 2011 978-1-4577-0590-8/11/$26.00 ©2011 IEEE MIT, Anna University, Chennai. June 3-5, 2011.

[13] Nabin Sharma, Palaiahnakote Shivakumara, Umapada Pal, Michael Blumenstein and Chew Lim Tan,"A New Method for Arbitrarily-Oriented Text Detection in Video." Published in 2012 10th IAPR International Workshop on Document Analysis Systems.