Applied Computing and Informatics (2016) **12**, 109–116

Saudi Computer Society, King Saud University

Applied Computing and Informatics

(http://computer.org.sa)
www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

A comprehensive of transforms, Gabor filter and *k*-means clustering for text detection in images and video

V.N. Manjunath Aradhya [a,*], **M.S. Pavithra** [b]

[a] *Dept. of MCA, SJCE, Mysore, India*
[b] *Dept. of MCA, DSCE, Bangalore, India*

Received 8 December 2013; revised 3 August 2014; accepted 6 August 2014
Available online 23 August 2014

KEYWORDS

Wavelet transform;
Multilingual text;
Wavelet decomposition;
Gabor filter;
k-means clustering;
Linked list approach;
Wavelet entropy

Abstract The present paper presents one of the efficient approaches toward multilingual text detection for video indexing. In this paper, we propose a method for detecting textlocated in varying and complex background in images/video. The present approach comprises four stages: In the first stage, combination of wavelet transform and Gabor filter is applied. By applying single level 2D wavelet decomposition with Gabor Filter, the intrinsic features comprising sharpen edges and texture features of an input image are obtained. In the second stage, the resultant Gabor image is classified using *k*-means clustering algorithm. In the third stage, morphological operations are performed on clustered pixels. Then a concept of linked list approach is used to build a true textline sequence of connected components. In the final stage, wavelet entropy of an input image is measured by signifying the complexity of unsteady signals corresponding to the position of textline sequence of connected components in leading to determine the true text region of an input image. The performance of the approach is exhibited by presenting promising experimental results for 101 video images, standard ICDAR 2003 Scene Trial Test dataset, ICDAR 2013 dataset and on our own collected South Indian Language dataset.

1. Introduction

With the current multimedia technology, the captured images and understanding these images through its contents have gained lots of attention from the computer vision community. Contents of an images and video help in clear understanding the information present within. A text is one of the images and video content which carries semantic information, and may help to provide the scene description of an image. Hence, the detection and extraction of either scene or graphics text has

* Corresponding author.
E-mail addresses: aradhya.mysore@gmail.com (V.N. Manjunath Aradhya), ngspavithra@gmail.com (M.S. Pavithra).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.aci.2014.08.001
2210-8327

been widely used in content based image indexing and retrieval. The detection of text present in video is being used in video summary and in video sequence retrieval. Text detection is a pre-processing task for text recognition. Nowadays, digital images also carry useful information. Uploading these images to social networking sites is getting increased day by day. So detecting texts of digital images also plays an important and challenging role in image retrieval system. Born-digital images are generated by computer software and are saved as digital images. Born-digital images have complex foreground/background, low resolution and have severe soften edges [1]. So detecting text from such born-digital images is difficult. Texts in natural scene images also have to be robustly detected before being recognized and retrieved. Scene images contain texts such as the advertising boards, name plates, address boards of houses, and landmarks of streets, are captured naturally when the scene images are taken by the camera. Therefore scene text is embedded in the background as a part of the scene. Scene images are complex because the backgrounds are complex containing the text in different sizes, styles and alignments and the resolution of the image is low. Hence detection of text from natural scene images remains still challenging task. Text detection and recognition in images and video frames aim at integrating advanced optical character recognition (OCR) and text based searching technologies [2]. In this regard, the existing text detection methods use text features such as gradient [3], edge [4] and texture [5] information. With such efforts, detection of text in images and video remains a challenging task due to variations of text background, font and orientations.

Recent text detection in natural scene images, born-digital images and video text detection has also been surveyed. Gonzlez and Bergasa [6] described a method to read text in natural images, using geometric and gradient properties. Zeng et al. [7] presented a framework for detecting text from webpage and email images, based on maximum gradient difference values. Hence the recent literature study implies that detection of text from either natural scene or digital images and text detection in video are still in the pace of research. Our proposed algorithm implements *k*-means clustering algorithm in detecting true text regions. In this regard, earlier research works of text detection in video and images based on *k*-means clustering algorithm are surveyed herewith, Phan et al. [8] developed a text detection approach with the Laplacian operator. Then *k*-means is used to classify all the pixels into clusters. Wu et al. [5] described a text localization method based on texture segmentation by computing texture features. *k*-means algorithm is applied for classification. Shivakumara et al. [9] describe a method based on the Laplacian in the frequency domain. In this, the input image is filtered with Fourier–Laplacian. Then, *k*-means clustering is used to identify candidate text regions.

The above stated studies revealed that the text detection approaches are either region or texture based methods. Though the concept of *k*-means clustering algorithm and a connected component analysis have used, the detection accuracy of the text region can still be improved without missing text data and reducing most of the falsely detected blocks of an image. Detection of texts of south Indian language is still a challenging task. Words of such south Indian languages are framed with modifiers and compound bases. To detect such a texts in an images/video, we propose a method based on Transforms, Gabor filter and *k*-means clustering. By sustaining the development of our system [10], which describes the text detection method in color and regular images. In the first stage of the system [10], wavelet transform and Gabor filter are applied to extract sharpened edges and textural features of a given input image. In the second stage of the method wavelet entropy is calculated to get an energy value of a resulted Gabor image in order to find the high frequency texture elements of a processed one to determine the true text region of an image. As a progress of our work [11], in the present paper, we propose a multilingual text detection system with the wavelet transform, Gabor filter and *k*-means clustering. The proposed method concentrates mainly on detection of English and south Indian language texts in images/video. The system yields better results for various background complexities and texts between other dominant non-text objects. Experiments are carried out on 101 video images dataset of [8], our own collected multilingual language dataset, ICDAR 2003 scene trial test dataset, ICDAR 2013 dataset and on video frames of Kannada. Comparative studies are reported in detail. The rest of this paper is organized as follows. Section 2 describes our proposed method. Experimental results are presented and performance evaluation on considered datasets is discussed in Section 3. Finally, conclusions are drawn.

2. Proposed methodology

The proposed method is a robust multilingual text detection approach based on the sequential adoption of Wavelet transform, Gabor filter, *k*-means clustering and a measure of wavelet entropy. First, by applying single-level discrete 2-D wavelet transform, a single-level 2-D wavelet decomposition is performed. As a result approximated and detailed co-efficients are obtained. Then detail co-efficients are merged and averaged to extract efficient texture feature information. Gabor filter is further applied in order to obtain edge information of an image. The resulted Gabor output image is grouped into three clusters by applying the *k*-means algorithm to classify the background, foreground and the true text pixels of an image. In the next stage, morphological operations are performed to obtain connected components, then a concept of linked list approach is in turn used to build a true text line sequence of connected components. In the final stage, wavelet entropy is measured in an each connected component sequence in order to determine the true text region of an input image. A complete text detection procedure of the proposed work is shown in Fig. 1 and explained in the following sub-sections.

2.1. Discrete wavelet transform for texture feature extraction

The Wavelet Transform is a method convolution of the wavelet function with the signal. The ability of the discrete wavelet transform to decompose a signal at different independent scales and to do it in a very flexible way [12]. The discrete wavelet transform (DWT) is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations. In our research work wavelets are used as analytical tools for signal processing. A study has performed about the development of the discrete wavelet transform (DWT), as a series expansion of signals in terms of wavelets and scaling functions which are associated with low pass and high pass