



Text Localization and Recognition from Video sequence using RBF-SVM classifier

Mahendra Poojar, Anitha G

M. Tech Student, Dept. of Computer Science Engineering, UBDTC, Davanegere, India

Associate Professor, Dept. of Computer Science Engineering, UBDTC, Davanegere, India

ABSTRACT: The problem of understanding scenes semantically has been one of the challenging goals in computer vision for many decades. There is a significant attention in the last few decades in text localization and recognition in images and videos of real world. Proper indexing and retrieval of digital video data is an important aspect of video analytics. Video indexing involves content analysis of video sequences, which is generally a computationally complex process. To help users navigate libraries of video, algorithms that automatically index video based on content are needed. One approach is to extract text appearing in video. Because of complex background and noises detection of text in video is very complex thing. In this paper we are proposing an efficient method for detecting and identification of text from video sequences. Detection is done using DWT (Discrete wavelet transform) and morphological operations. For recognition process RBF-SVM is used and for feature extraction DOG (Difference of Gaussian, Gabor wavelets and Edge features are extracted).

KEYWORDS: Discrete wavelet transform, RBF-SVM classifier, Difference of Gaussian, Gabor wavelets.

I. INTRODUCTION

There is a significant attention in the last decade in text localization and reorganization in images and videos of real world. In content based retrieval systems, video event detection and human computer interaction it is very important to have an end to end system for text detection and reorganization. The challenging goal in computer vision for many decades is to understand the scenes semantically. It has gained considerable attention over the past few years, in particular, in the context of street scenes. This problem has manifested itself in various forms, namely, object detection, object recognition and segmentation. There have also been significant attempts at addressing all these tasks jointly. Although these approaches interpret most of the scene successfully, regions containing text tend to be ignored.

In existing system text provides important information about images or video sequences in a documented image, but it always remains difficult to modify the static documented image. To carry out modification in any of the text matter the text must be segmented out from the documented image, which can be used for further analysis. Taking consideration to video image sequence the isolation of text data from the isolated frame becomes more difficult due to its variable nature. Various methods were proposed for the isolation of text data from the documented image. Existing system uses MSER technique for localizing the text region from the video frame, but the segmentation depends on the predefined threshold value and also it uses trained data for detection of text and non-text region. For classification of characters traditional SVM is used. SVM is robust in classifying the data but as the number of classes increases the recognition rate may reduced. The popularity of digital video is increasing rapidly. To help users navigate libraries of video, algorithms that automatically index video based on content are needed. One approach is to extract text appearing in video. Such text often gives an indication of a scene's semantic content. Text can have arbitrary color, size, and orientation. Backgrounds may be complex and changing. It has focused only on detecting the spatial extent of text in individual video frames. A text extraction and recognition algorithm using RBF-SVM (Radial Basis Function Support Vector Machine) is proposed. This constitutes a text event that should be entered only once in the video index. Therefore it is also necessary to determine the temporal extent of text events. Such text effects are common in television programs and commercials to attract viewer attention.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

II. LITERATURE SURVEY

Ratnashil N Khobragade et.al [1] has proposed a system Segmentation of Marathi Handwritten Characters and Numerals. In this they considered Marathi words and Marathi Numerals for segmentation. The algorithm is use for Segmentation of lines and then characters. The segmented characters are then stores in result variable. First it Separate the lines and then it Separate the characters from the input image. This procedure is repeated till end of file. Mohamed Ben Halima et.al [2] has proposed a system on Arabic Text Recognition in Video Sequences. In this they proposed a robust approach for text extraction and recognition from Arabic news video sequence. The text included in video sequences is an important needful for indexing and searching system. However, this text is difficult to detect and recognize because of the variability of its size, their low resolution characters and the complexity of the backgrounds. To solve these problems, they have proposed a system performing in two main tasks: extraction and recognition of text. Shangxuan Tian et.al [3] has proposed a robust text segmentation method that employs Markov Random Field (MRF) and use graph cut algorithms to solve the energy minimization problem. To effectively select accurate seeds to boost the text segmentation performance, stroke feature transform is adopted to robustly identify text seeds and text edges. Background seeds are obtained near the text edges in order to well preserve the text boundaries. Experiments on ICDAR 2003 and ICDAR 2011 datasets show that the proposed technique obtains superior performance on both pixel level and atom level segmentation. B.H.Shekar Et.al [4] has proposed a skeleton matching based approach which aids in text localization in scene images. In this they pre-processed input image and segmented into blocks using connected component analysis. We obtain the skeleton of the segmented block using morphology based approach. The skeletonised images are compared with the trained templates in the database to categorize into text and non-text blocks. Further, the newly designed geometrical rules and morphological operations are employed on the detected text blocks for scene text localization. The experimental results obtained on publicly available standard datasets illustrate that the proposed method can detect and localize the texts of various sizes, fonts and colors.

III. METHODOLOGY

Figure 1 represents the proposed architecture. The algorithm consists of major two parts testing and training phase respectively. In training phase the system is trained to create a knowledge base. First stage of training phase is to read the character images stored in the database. Then images are pre-processed. When the image is pre-processed feature extraction is achieved. For feature extraction DOG (Difference of Gaussian), Gabor wavelets and Edge features are used. Using RBF-SVM training algorithm a knowledge base will be created which is used in testing phase.

Testing consists of video is the system input to the top layer of the architecture, and then the video is converted into video frames. An image pre-processing is used to change the video frame into gray scale image and also resize an image. Discrete Wavelet Transform algorithm is used on video frames. Extract edges of the text from the video frames. To remove non-text region from the video frames and segment the text region morphological operation is used. When the characters are segmented, feature extraction process will be achieved. The obtained feature vector is matched with knowledge base using RBF-SVM classification. Finally the recognized texts are displayed.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

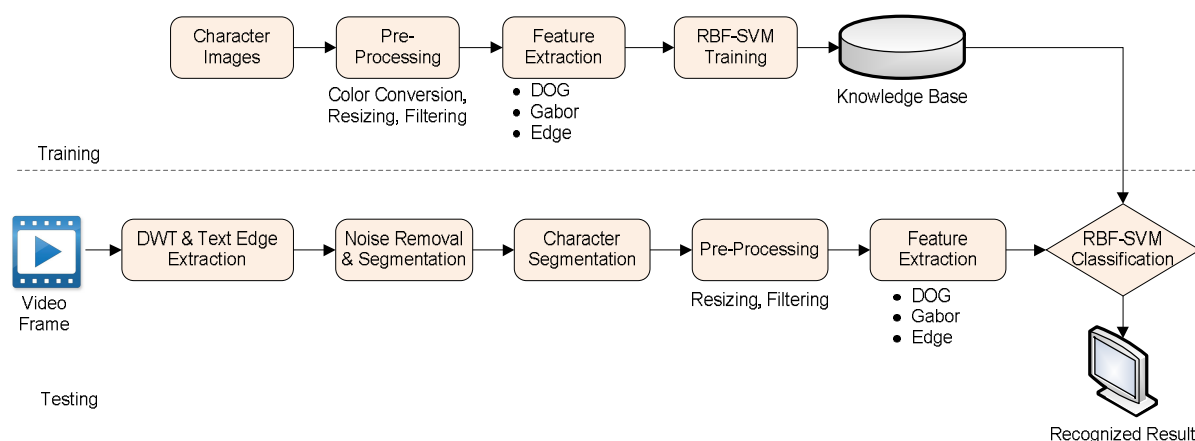


Figure 1: Block Diagram Of Proposed System.

A. PRE-PROCESSING

Pre-processing is mainly used to enhance the contrast and to adjust the size of the image, removal of noise and isolating objects of interest in the image. Pre-processing is any form of signal processing for which the output is an image or video, the output can be either an image or a set of characteristics or parameters related to image or videos to improve or change some quality of the input. This process will help to improve the video or image such that it increases the chance for success of other processes. In this paper we considered character images as input and those images are subjected to pre-processing this will resulting in color conversion, resizing and filtering of input character image.

B.FEATURE EXTRACTION

First stage of training phase is to read the character images stored in the database. Then images are pre-processed. When the image is pre-processed feature extraction is achieved. For feature extraction DOG (Difference of Gaussian), Gabor wavelets and Canny Edge algorithm are used.

IV.DOG (DIFFERENCE OF GAUSSIANS)

DOG is a feature enhancement algorithm explains the subtraction of one blurred version of an original image from another, less blurred version of the original. In the simple scenarios of gray scale images, with differing standard deviations the blurred images are obtained by convolving the original gray scale images with Gaussian kernels. Blurring an image using a Gaussian kernel suppresses only high-frequency spatial information. Subtracting one image from the other preserves spatial information that lies between the range of frequencies that are preserved in the two blurred images. Thus, the difference of Gaussians is a band-pass filter that discards all but a handful of spatial frequencies that are present in the original gray scale image.

As a feature enhancement algorithm, the difference of Gaussians can be utilized to increase the visibility of edges and other detail present in a digital image. A wide variety of alternative edge sharpening filters operate by enhancing high frequency detail, but because random noise also has a high spatial frequency, many of these sharpening filters tend to enhance noise, which can be an undesirable artifact. The difference of Gaussians algorithm removes high frequency detail that often includes random noise, rendering this approach one of the most suitable for processing images with a high degree of noise. A major drawback to application of the algorithm is an inherent reduction in overall image contrast produced by the operation.

V.GABOR WAVELET

The Gabor wavelet transform is used for feature extraction in this architecture, it has some attractive mathematical and biological properties and has been used frequently on researches of image processing. Gabor functions provide the optimal resolution in both the time and frequency domains, and the Gabor wavelet transform seems to be the optimal



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

basis to extract local features for several reasons, they are Biological motivation and Mathematical and empirical motivation.

Gabor wavelet is given as

$$\varphi(x, y) = \frac{f^2}{\pi\gamma\eta} \exp\left(-\left(\frac{f^2}{\gamma^2}x_r^2 + \frac{f^2}{\eta^2}y_r^2\right)\right) (\exp(j2\pi x_r) - k) \quad (1)$$

Where K is an offset parameter dependent on \square and \square .

Gabor wavelet is used as the discrete wavelet transform with either continuous or discrete input signal. The non-orthonormal wavelets could provide a complete representation only when they form a frame only if the Gabor wavelets are used for feature extractions. When extracting features for pattern recognition, retrieval, or computer vision purpose, the transformed coefficients are used for distance measure or compressed representation but not for reconstruction, so the orthogonal constraint could be omitted.

VI. CANNY EDGE DETECTION ALGORITHM

Canny edge detector is widely considered to be the standard edge detection algorithm in the industry. There are several steps are to be followed in the canny edge detector, firstly we have to smoothen the image with a two dimension Gaussian, next take the gradient of an considered image, then suppression is done through non-maximal, then edge threshold is calculated. The next step is to extract moving edges from sequential video frames and process the resulting edge information to obtain quantitative geometric measurements of passing vehicles. We are comparing and calculating vehicle density by using these two different edge techniques. The edges are detected by the canny edge algorithm. Using the gradient kernel approach image gradient magnitude is calculated in horizontal direction. G_x and vertical direction G_y for each pixel. And the direction of the pixel is measured by

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (2)$$

Threshold values such at the higher threshold value (HIV) and lower threshold value (LTV) are selected from the histogram of the image in order to detect the edges.

VII. DWT (DISCRETE WAVELET TRANSFORM)

A wavelet transform is a type of transform in which the wavelets are discretely sampled. We are using DWT for text edge extraction process. In analysing a real world signals DWT will provide a multi-resolution description of a signal. Essentially, a discrete multi-resolution description of a continuous-time signal is obtained by a DWT. DWT converts a series $a_0, a_1, a_2, \dots, a_m$ into one low pass coefficient series known as approximation and one high pass coefficient series known as detail. Length of each series is $m/2$. In real life situations, such transformation is applied recursively on the low-pass series until the desired number of iterations is reached. The discrete wavelet transform has applications ranging from data compression to signal coding. In our paper Gobar wavelet was used to filter a noisy signal to extract information from the signal. For noise removal and segmentation process will use canny edge detection method for image segmentation, after segmentation of each text image will going for pre processing the segmented image to do resizing and filtering. Next, these pre-processed segmented images are then subjected to RBF-SVM classifier which is explained in next section.

VIII. RBF-SVM CLASSIFIER

RBF-SVM stands for Radial Basis Function Support Vector Machines. In order to achieve the optimal separating hyper-plane in the higher dimensional feature space, SVM first maps the input vector into a higher dimensional feature space. The RBF kernel on two samples X and X' , represented as feature vectors in some input space, is defined as

$$K(X, X') = \exp\left(-\frac{\|X-X'\|^2}{2\sigma^2}\right) \quad (3)$$

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

$\|X - X'\|^2$ is the squared Euclidean distance between the two feature vectors. σ is a free parameter.

Furthermore, a decision boundary, i.e. the separating hyper-plane, is determined by support vectors rather than the whole training samples and thus is extremely robust to outliers. Exactly an SVM classifier is designed for binary classification. That is, to separate a set of training vectors which belong to two different classes. Note that a decision boundary similar to the support vector i.e. training samples. To provide the required mapping to ice-water labels a soft-margin SVM classifier is used. An SVM works by computing a linear decision boundary in a high dimensional space using the subset of labeled training samples near the decision boundary (called the support vectors). The SVM decision boundary equation is

$$f(x) = \sum_{v_i} y_i \alpha_i K(x_i, x) \quad (4)$$

Where $K(x_i, x)$ Kernel function is defined by

$$K(x_i, x) = \exp(-\gamma|x_i - x|^2) \quad (5)$$

Then according to the result which we stored in the knowledge base at training phase SVM classifier will classify the segmented image by comparing it with character image.

IX.EXPERIMENTAL RESULT

Figure 2 represent the experimental result of the proposed system. Firstly will consider the video frames as input image, which is shown in Figure 2(a) after considering video frames from the input video those frames are pre-processed, in that gray scale converted images, next in those images noise should be removed in order to get noise free clear image for next process as shown in Figure 2(b), from that detected images are tracked as shown in Figure 2(c) next by applying canny edge detection algorithm we will go for segmentation of pre-processed image as shown in Figure 2(d).The effective experimental results shows how effectively our proposed system will work.

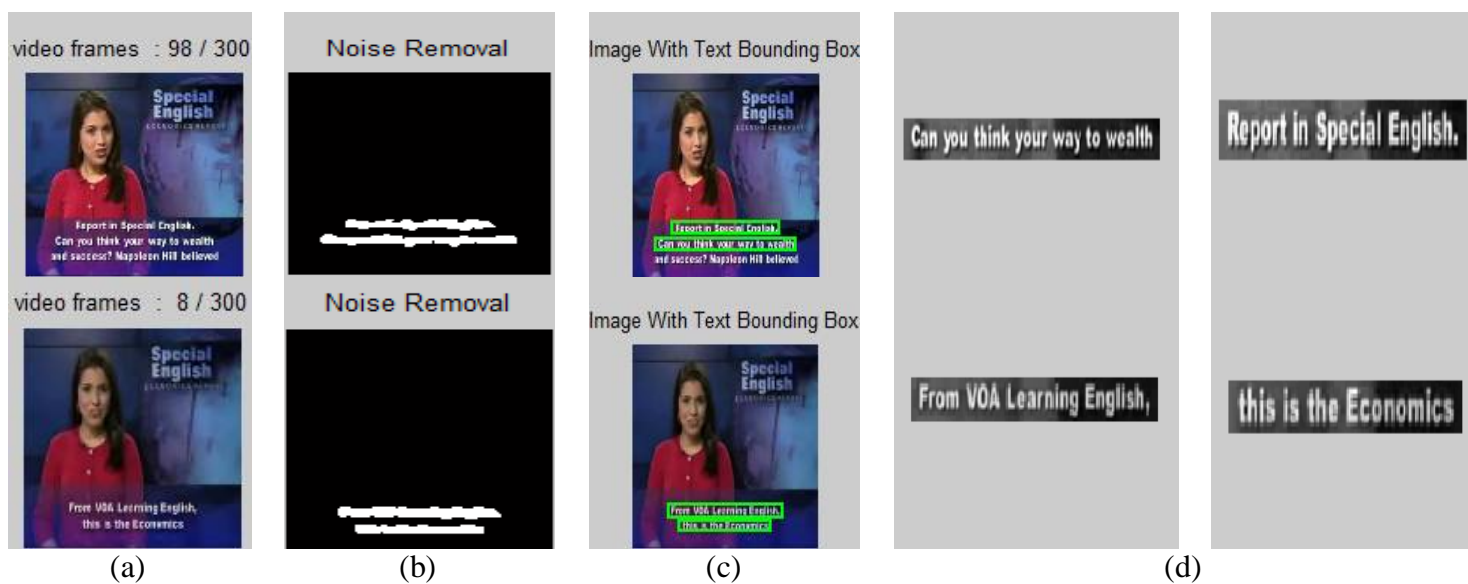


Figure 2: (a) Video Frames/Input Images; (b) Noise Removal Image; (c) Detected Image; (d) Segmented Images.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

X.CONCLUSION

In this paper propose an end-to-end text detection and recognition architecture. The text detection component uses RBF-SVM classifier based on rich shape descriptors such as DOG, Gabor and edge features for improved performance, and DWT technique for filtering and text edge detection, leading to SVM speed improvement. Detection is done using DWT (Discrete wavelet transform) and morphological operations. For recognition process RBF-SVM is used and for feature extraction DOG (Difference of Gaussian, Gabor wavelets and Edge features are extracted). Extensive evaluation on a large dataset illustrates the efficacy of our approach in both pixel-level text detection and word recognition tasks

REFERENCES

- [1] Ratnashil N Khobragade, Dr. Nitin A. Koli and Mahendra S Makesar, "Segmentation of Marathi Handwritten Characters and Numerals", International Journal of Innovative Research in Computer and Communication Engineering, Volume 3, 2015.
- [2] Mohamed Ben Halima, Hichem Karray and Adel M. Alimi, "Arabic Text Recognition in Video Sequences", arXiv, 2013.
- [3] Shangxuan Tian, Shijian Lu, Bolan Su and Chew Lim Tan, "Robust Text Segmentation using Graph Cut", International Conference on Document Analysis and Recognition, 2015.
- [4] B.H.Shekar and Smitha M.L., "Skeleton Matching based approach for Text Localization in Scene Images", 2015.
- [5] B. Gangamma, Srikanta Murthy K and Arun Vikas Singh, "Restoration of Degraded Historical Document Image", Journal of Emerging Trends in Computing and Information Sciences, Volume 3, 2012.
- [6] Cunzhao Shi, Chunheng Wang, Baihua Xiao, Yang Zhang, and Song Gao, "Scene text detection using graph model built upon maximally stable external regions," Pattern Recognition Letters, vol. 34, no. 2, pp. 107 –116, 2013.
- [7] Serge Belongie Kai Wang, Boris Babenko, "End-to-end scene text recognition," in International Conference on Computer Vision, 2011.
- [8] Milyaev S, Barinova O, Novikova T, Lempitsky V, and Kohli P, "Image binarization for end-to-end text understanding in natural images," in International Conference on Document Analysis and Recognition (ICDAR), 2013.
- [9] Xuwang Yin, Xu-Cheng Yin, Hong-Wei Hao, and Khalid Iqbal, "Effective text localization in natural scene images with mser, geometry-based grouping and adaboost," in International Conference on Pattern Recognition, 2012.