Recent Advances in Video Based Document Processing: A Review

[1]Nabin Sharma, [2]Umapada Pal, and [1]Michael Blumenstein
[1]Griffith University, Queensland, Australia,
Email: {m.blumenstein, nabin.sharma}@griffith.edu.au
[2]Computer Vision and Pattern Recognition Unit, Indian Statistical Unit, Kolkata, India.
Email: umapada@isical.ac.in

***Abstract*— Extraction and recognition of text present in video has become a very popular research area in the last decade. Generally, text present in video frames is of different size, orientation, style, etc. with complex backgrounds, noise, low resolution and contrast. These factors make the automatic text extraction and recognition in video frames a challenging task. A large number of techniques have been proposed by various researchers in the recent past to address the problem. This paper presents a review of various state-of-the-art techniques proposed towards different stages (e.g. detection, localization, extraction, etc.) of text information processing in video frames. Looking at the growing popularity and the recent developments in the processing of text in video frames, this review imparts details of current trends and potential directions for further research activities to assist researchers.**

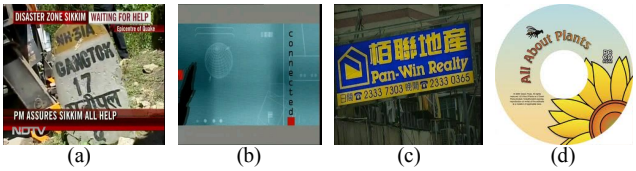
Keywords- Video OCR, Text information processing, Video Document processing Survey, Text detection, Text localization.

I. INTRODUCTION

Advancements in digital technology has gifted human beings with low priced digital imaging devices such as, digital cameras, cellular phones with digital cameras, PDAs etc. These devices are not only inexpensive, but are also highly portable, and have huge prospects to supplement traditional imaging devices such as digital scanners, etc.

Image acquisition using portable digital cameras, digital cameras attached to cellular phones, etc, has probably given birth to the Camera and Video-based document processing and recognition problem. Documents captured using traditional scanners have high resolution, contrast and less noise, and are easier to process for OCR. But the images and videos captured using a digital camera suffers from low resolution, contrast, blur, distortion, noise, etc., to mention a few. It was also noted that the images captured by cameras have a better resolution than the same document in video frames. Hence, the traditional scanner based document analysis and recognition techniques cannot be directly applied to the Video documents or documents captured using digitals cameras.

The video document processing community has classified the text in video frames based on its origin [1, 2, 3, 4]. The text information which is artificially overlaid on the image is often known as '*Caption text*' or '*graphic text*' (e.g. subtitles in news video, sports scores, etc). Whereas, the text which naturally exists in the image is known as *'Scene text'*(e.g. text on vehicles, commodities, buildings, sign boards on roads, etc.). *Scene text* has additional complexities such as multi-orientation and multi-lingual issues, whereas *Caption text* is usually horizontal or vertical. Figure 1 (a) and (b) are examples of *Caption text*, and the rest are examples of *Scene text.* A large number of techniques have been proposed by researchers towards text detection from video frames with horizontal, vertical, and non-horizontal text orientation. But arbitrarily oriented or curved text detection from video frames, as shown in Figure 1(d), has not been addressed to the best of our knowledge.



(a) (b) (c) (d)

Figure 1: Sample of video frames with different text orientation. Video frame with (a) Horizontal text from news clip; (b) Vertical text; (c) Non-horizontal and multi-lingual text; (d) Multi-oriented/arbitrarily oriented text;

A typical video frame processing system is shown in Figure 2. Text frame selection determines whether a frame contains text information. Text detection and localization finds and defines the actual location of the text present in the frame by forming bounding boxes around the text. After the bounding boxes are found, the actual text/characters are extracted and binarized for OCR.

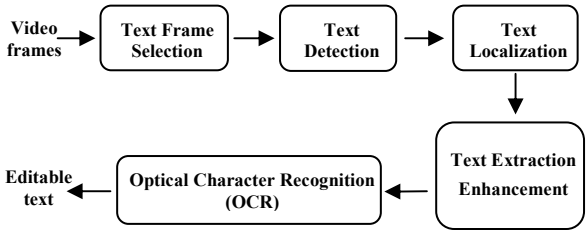


Figure 2: A typical video frame processing system

Although there are a few surveys [1, 2, 3, 4] available on camera/video-based document processing in the literature, these are outdated and do not capture recent developments. This motivates us to present a review of the recent advances in video document processing in this paper. Different techniques which include text frame selection, text detection and localization, text extraction/segmentation, binarization, and enhancement, are reviewed in Section II. In Section III, we present a brief overview of the recognition schemes proposed by various researchers in the area of video text recognition. Section IV concludes the review providing our observations, future directions, and a summary of the review.

II. TEXT EXTRACTION

Text extraction from video frames can be divided into five stages, namely text frame selection, detection and localization, extraction, binarization, and enhancement. In this section we present a brief overview of the recent advances and the various techniques proposed by researchers in each of the five stages.

A. Text frame selection/classification

Text frame selection/classification attempts to label a video frame as text or non-text, before text detection and recognition. It helps to avoid the computationally expensive text detection methods on non-text frames present in the video. Moreover, text detection methods can also detect text incorrectly from a non-text frame. Ideally, the text frame selection method should be simple and fast enough to determine a text and non-text frame. To the best of our knowledge, not much progress has been made towards text frame selection, and in most of the text detection research, the text frames are assumed to have text information present in it. In this section we discuss the few techniques proposed for text frame selection.

Na and Wen [5] proposed a text tracking algorithm based on SIFT feature and Geometric constraint. Text frame classification using edge based features was proposed by Shivakumara *et al.* [6]. Height, straightness and proximity based edge features were used. The results obtained are promising.

Recently, Shivakumara *et al.* [7] proposed a text frame classification technique, using a combination of wavelet and median moment features with *k*-means clustering. The frame is divided into 16 equal non overlapping blocks and the probable text blocks are identified using Min-Max clustering, and the computed features. If one true text block is identified, the frame is considered as a text frame. They tested the method on 1220 text frames and 800 non-text frames. A text frame classification accuracy of 74.17% was achieved.

B. Text Detection and Localization

Based on the features used, text detection and localization techniques can be divided in two categories [3, 4], namely, *Region-based* and *Texture based.* The *Region based* techniques work in a bottom-up fashion, by dividing the frame into small regions and then merging the probable text regions to form bounding boxes for the text. Connected-components, color, and edge features are commonly used in the *Region based* approaches. *Texture based* methods uses the texture properties of the text to distinguish between the text and background. Wavelet transform, Gabor filters, Fourier transform, machine learning based approaches, etc. are often used in *Texture-based* techniques. Brief reviews of the techniques proposed in each of the mentioned categories are present in sub-section 1 and 2.

1. Region based methods

A stroke width similarity based technique for text detection was proposed by Dinh *et al.* [8]. A local adaptive threshold was used to eliminate the background preserving the text. Morphological dilation was applied to localize the text. To refine the text location, the multi-frame refinement method was used.

Stroke filter based methods [9, 10, 11], and Stroke Width transform [12] have been used by many researchers for text detection and localization. Jung *et al.* [9] used stroke filter for text segmentation considering the intrinsic characteristics of the text. Based on the stroke filter response and text polarity, local region growing was used to segment the text. An OCR feedback score was used to improve the text segmentation accuracy. Jung *et al.* [11] described the stroke filter in details and also its application to text localization in video frames. An SVM classifier is used for the verification of the candidate text. Based on the verification score and color distribution, the text line refinement is done. Li *et al*. [10] used stroke filter to calculate the stroke map. They used two SVM classifiers to obtain rough text region and to verify the candidate text lines. Localization was achieved by projection profile. The second SVM was used to verify localized the text lines.

Shivakumara *et al.* [13] proposed an edge based technique for text detection in images with text present in the horizontal direction. The frame was segmented into 16 non-overlapping blocks. Mean and median filter, and edge analysis was used to identify the candidate text blocks. Using block growing method, the complete text block was obtained. Finally, based on the vertical and horizontal bar feature, the true text regions are detected. In Shivakumara *et.al* [14], filters and edge analysis were used for initial text detection. The straightness and cursiveness edge features were used for false positive elimination.

A hybrid system for text detection based on the edges, local binary pattern operator, and SVM was proposed by Anthimopoulos *et al.* [15, 16]. After the detection of text block using the edge map in [15], dilation, opening, projection analysis, and a machine leaning step using SVM was introduced for refinements. In [16] multi-resolution analysis was also done to detect character of broad size range. Text detection using a cascade AdaBoost classifier with HOG and multi-scale local binary pattern feature, was proposed by Pan *et al.* [17]. Text localization was done using window grouping technique. Within each located text line, local binarization is done to extract candidate CCs and non-text CC's are filtered using Markov Random field model and MLP in order to get the final text line.

Use of temporal information for moving text detection was proposed by Huang *et al.* [18]. The frame was divided into sub-blocks and the inter-frame motion vector was computed for each sub-block. Their proposed technique worked well for scrolling text in news clips and movies. Another method to detect scrolling text was proposed by Tsai *et al* [19]. They used edge information for detection

and a two-dimensional projection profile for localization. Li *et al.* [20] proposed a four stage adaptive text detection method. Different edge detectors were used based on the background complexities. Then the CC analysis was done to find the text candidates and was refined in the fourth stage. Shi *et al.* [21] used the block change rate based techniques to detect and localize text.

Gradient difference based method for text detection was proposed by Shivakumara *et al.* [22]. Zero crossing was used to determine the bounding boxes for the detected text line. A few methods were also proposed for Arabic/Farsi [23, 24] text detection from video frames. A projection analysis based approach for Arabic video text detection and localization was proposed by Halima *et al.* [23]. While Moradi *et al.* [24] used edge and corner detection method and discarded non-text corners by histogram analysis.

Park *et al.* [25] used horizontal and vertical projection profile to detect Korean text in outdoor signboards. Shivakumara *et al*. [26] introduced the classification of low and high contract images for text detection. They analyzed the number of edges found using sobel and canny edge detector for low and high contrast images, to form the heuristic rules for classification. A method for handwritten scene text detection was first proposed by Shivakumara *et al.* [27]. The method uses maximum color difference and boundary growing method based on nearest neighbor concept, to detect multi-oriented handwritten text. A Self-Organizing Map (SOM) neural network based technique for artificial text detection in video frames was due to Yu *et al.* [28]. Three layers of supervised SOM were used to classify text, and non-text areas. Huang *et al.* [29] used the texture feature in the stroke map to detect text. For text localization, Harris' corner detection approach was used on the stroke map. Morphological operations were used to connect the corners. Guru *et al.* [30] proposed an Eigen value based technique, which performs a block wise Eigen analysis on the gradient image of the video frame. Eigen analysis helped in identifying the potential text blocks.

Zhang and Sun [31] used a Pulse Coupled Neural Network (PCNN) edge based method for locating text. The work depicted the use of PCNN in frequency domain for solving text localization problem. Anthimopoulos *et al.* [32] proposed the feature set produced by a Multilevel Adaptive Color edge Local Binary Pattern (MACeLBP) with a random forest classifier for text detection. A gradient based algorithm was then used for localization. Use of Moravec operator for text detection in images and video frames was proposed by Kumari and Shekar [33]. Zhao *et al.* [34] used StrOke unIt Connection (SOIC) operator to find the seed stroke units, and to train the SVM classifier. The method uses the stroke shape distributions for training. Edge based features for Urdu text localization in video images was proposed by Jamil *et al.* [35].

Pan *et al.* [36] proposed a hybrid method for text detection and localization using stroke segmentation, verification, and grouping. Stroke candidates are determined by a scale adaptive segmentation method and verification is done by weight Conditional Random Field (CRF). Kruskal algorithm is then used to group the strokes. Recently, Uchida *et al.* [37] established that Speeded Up Robust Features (SURF) can be used to detect character regions and to distinguish text and non-text regions with good accuracy.

2. Texture based methods

Wavelet transforms and its variations have become very popular among researchers for texture analysis. Most of the recent works on texture based text detection and localization are based on the wavelet transform [38, 39, 40, 41, 42]. Other methods such as the Gabor filter [43, 44], DCT [51], Haar wavelet [45], spatial analysis [46], Laplacian [47], Fourier [48] etc. were also used by researchers in the recent past.

In general, texture features are computed and are used to train a classifier to discriminate text and non-text. A combination of wavelet features and an SVM classifier were used in [38, 40, 45]. Ye *et al*. [38] used 2D wavelet coefficients to calculate histogram wavelet coefficients of all pixels. SVM with a RBF kernel was used for classification of text and non-text. They introduced an OCR feedback procedure to locate the final text lines. Ji *et al.* [45] used Pyramid Haar wavelet to represent an image into multiple scales. Using Directional Correlation Analysis (DCA) of Local Haar Binary Pattern (LHBP) the candidate text regions were found. LHBP histogram with SVM was used for the refinement of the text results. Ji *et al.* [40] used two texture features namely wavelet coefficients and Gray-level co-occurrence matrix for text detection along with SVM.

Zhao *et al.* [41] used wavelet transform and sparse representation with discriminative dictionaries for text detection. Shivakumara *et al.* [39, 42] also used haar wavelet in both the works. In [39] *k-means* clustering was used to classify text and background. In [42] they also used color features along with Wavelet-Laplacian method to detect text. Sivakumara *et al.* [49] used wavelet-median-moment feature with k-means clustering to obtain text pixels. Angle projection based boundary growing was used to handle multi-oriented text. Recently, Shivakumara *et al.* [47] proposed a Laplacian approach for multi-oriented text detection in videos. A Fourier-Laplacian filter along with k-means clustering is used to determine the text and non-text clusters. Straightness and Edge density features were used for false positive elimination. Phan *et al.* [50] used the same Laplacian approach as in [47] to identify text candidates, but used CC analysis to form simple CCs. Using the straightness and edge density feature the text blocks were finalized. Fourier-statistical features in RGB space were used for text detection in video frames by Shivakumara *et.al.* [48]. The Fourier-statistical features were subjected to K-means clustering to classify text pixels from the background. Yi *et al*. [43] used Gabor filters to describe the stroke components in the text characters. They defined

Stroke Gabor Words (SWGs) and used it with image window classification techniques to detect text regions.

The use of Conditional Random Field (CRF) [44] along with texture feature is also becoming popular among the researchers. Peng *et al.* [44] computed the features using 2-D Gabor filters and Harris corner detection. Based on the confidence of text and background labeling by SVM, CRF framework is defined, and isolated text blocks are merged by heuristic reasoning. Hanif *et al.* [46] used gray-level co-occurrence matrix for texture analysis, and three classifiers namely maximum a posteriori, neural network, and mean spatial histogram were used to discriminate text and non-text. Discrete Cosine transform (DCT) was used by Qian *et al.* [51]. Texture intensities were used to verify horizontal and vertical text. Horizontal and Vertical projection profiles were used for text localization.

C. *Extraction, Binarization, and Enhancement*

Not much work has been towards text extraction, binarization and enhancement in the recent past. Extraction and binarization is often used synonymously. They aim towards the extraction of the individual characters from the detected and localized text blocks, for OCR. A wide range of binarization techniques have been used by the researchers in last few decades in order to get a two tone image. It often results in touching characters, individual characters with broken segments, missing part of character, etc to mention a few. In order to get a better OCR accuracy, enhancement of the extracted characters is required, and not much research has been done for the enhancement of the broken characters for video frames.

Methods based on Tensor voting [52], Fourier moments [53], Conditional Random Fields (CRF) [53, 54], and Gradients [55, 56], were proposed towards character extraction/segmentation. Few binarization techniques [57, 58, 59, 60] were also proposed in the recent past.

A Tensor voting based text segmentation technique was proposed by Lim *et al.* [52]. In their approach the image was first decomposed into chromatic and achromatic regions. Using tensor voting the text layers are identified and noise removal was done by adaptive median filter. K-means clustering algorithm was used for segmentation. Cho *et al.* [53] also used CRFs for text extraction by superpixel representation of the image. Character features namely, color, edge strength, stroke width and contextual feature, were used. Zhang *et al.* [54] also used CRF for scene text extraction. A two-step iterative CRF method with OCR as a region filtering module was used in [54]. Recently, Shivakumara *et al.* [55] proposed a gradient based character segmentation scheme. Bresenham's line drawing algorithm was used for handling multi-oriented text, and gradient features are then extracted. Min-Max clustering was used to separate text and non-text cluster. Segmentation was achieved based on the height difference, top distance and bottom distance vector of the union operation. Phan *et al.* [56] proposed a Gradient Vector Flow (GVF) based techniques for character segmentation from the localized text in video frame. GVF was used to find the pixels which are potentially part of non-vertical cuts. Then, multiple least cost paths were found from top row to bottom row to the image, and finally the cuts which pass through the middle of the character are eliminated. Rajendran *et al.* [61] proposed a Fourier-moment based feature for the extraction of words and characters from the video textline. They used similar method as proposed in [55].

Binarization technique proposed by Zhou *et al.* [57] used the contour of the text along with local thresholding to determine the inner side of the contour. The contour is then filled up to form the characters. Mishra *et al.* [58] presented an MRF based technique of binarization of natural scene text. The pixels in the image were represented as random variables in an MRF, and quality of the binarization is determined by the value of energy function. The energy function is minimized using an interactive graph cut scheme to find the optimal binarization. A K-means clustering and SVM based method for binarization was proposed by Wakahara *et al.* [59], which is a four step method. HSI color space was used, and was helpful in contrasting characters against the backgrounds. SVM was used to determine the character or non-character images. Character-likeness estimates are used to achieve optimal binarized result. Ntrirogiannis *et al.* [60] used the upper and lower baseline of the text, stroke width, and convex hull analysis for binarization of the text in video frames.

III. CHARACTER RECOGNITION

The aim of video document processing is to recognize the text content in the video frame, which can be used for indexing and information retrieval purposes. The recognition of text in video frames is still in its infancy, and not much work has been done on it. A number attempts [62, 63, 64, 65, 23, 25, 66, 67, 68] on video text recognition have been made, recently.

Uchida *et al.* [62] presented a Mosaicing-by-recognition technique for video based text recognition. The video mosaicking and text recognition problem is formulated as a unified optimization problem, which is solved by dynamic programming-based optimization algorithm. Character recognition accuracy of more than 95% was reported. A recognition scheme for Korean characters present in the outdoor signboards was proposed by Park *et al.* [25]. The System uses a minimum distance classifier with a shape based statistical feature, for character recognition. An Arabic video text recognition system was proposed by Halima *et al.* [23]. The feature used for recognition include projection feature, transition feature, occlusion features, number of components in the character and location of the dots. A *k*-nearest neighborhood classifier was used for classification, and best results were obtained for k=10. Iwamura *et al.* [66] proposed a non-learning based technique for camera captured characters. It tries to find the most similar example of an input character.

Saidane and Gracia [63] used a convolutional neural network for character recognition and achieved an average recognition accuracy of 84.53% from the text extracted from ICDAR 2003 dataset. Features they used include oriented edges, corners, end points that were extracted directly from the three color channels. A Subspace method was used for low-resolution character recognition by Ohkura *et al.* [64]. The resolution of the images were enhanced using a super-resolution scheme before the recognition phase. There was a significant increase in the recognition accuracy from 90.35% (for 7x7 pixels size character) to 99.97% (for 9x9 pixels size character).

Saidane *et al.* [65] presented a graph based technique named image Text Recognition Graph (iTRG) for color text recognition from images and videos. The graph consists of five modules, namely, text segmentation, graph connection builder, character recognition, graph weight calculator and optimal path search module. ICDAR 2003 data set was used of performance evaluation. Recently, Shivakumara *et al.* [67] proposed a video character recognition scheme using hierarchical classification based on voting method. They used structural features to classify 62 character classes into different smaller classes. Only 10% of the samples of each class were used for training, which yields a recognition accuracy of 94.5%. Their dataset was chosen from 2005 and 2006 TRECVID database. Coates *et al.* [68] presented a scheme based on unsupervised feature learning. Using a combination of K-means clustering and linear SVM, 81.7% accuracy was obtained on a 62 class problem. The ICDAR 2003 dataset was used for testing.

IV. Summary and Conclusions

In this paper we presented a brief review of the recent techniques proposed by researchers towards video based document processing. The survey reveals that most of the work was done towards text detection and localization. However, it is noted that text frame detection is an area which is still in its infancy. Not much work has been done towards text frame selection/segmentation from video, although this has a huge potential in saving the computational time for text detection in a non-text frame. There were recent advances towards text extraction, binarization, and recognition, which indicate that text detection and localization steps have a significant number of established methods available with satisfactory results being obtained. But still text detection and localization enjoys more than a decade of popularity, as there is a huge scope of possible improvement by incorporating temporal information. The drawback which video document researchers might be facing is the non-availability of the standard benchmark datasets for research. The review depicts that, researchers have used different datasets to evaluate their techniques, which makes it difficult to perform a comparative study of the results. A benchmark dataset provides a common platform for the researcher to evaluate their methodologies, which will help in establishing the results. We have noted that [31, 32, 33, 41, 47, 48, 49] have reported better results on text detection. Better accuracy on segmentation was reported by [55, 56]. The methods [58, 60] have reported better binarization results. The character recognition accuracies reported by various researchers, are discussed in Section III.

We also noted that there is no work on the classification of graphics text and scene text. Work in this area will also be helpful for video document handling.

We hope that this review not only encourages the current research on video document processing but also provides appropriate directions for future research.

References

[1] D. Doermann, J. Liang, and H. Li., "Progress in Camera-Based Document Image Analysis". ICDAR, 2003, pp.606-616.

[2] J. Liang, D. Doermann, and H. Li., "Camera-Based analysis of text and document: a survey". IJDAR, 2005, pp.84-104.

[3] K. Jung, K. I. Kim, and Anil. K. Jain, "Text information extraction in images and video: a survey", Pattern Recognition, vol. 37, 2004, pp.977-997.

[4] J. Zhang and R. Kasturi, "Extraction of Text Objects in Video Documents:Recent Progress", DAS, 2008, pp.5-17.

[5] Y. Na, and D. Wen, "An Effictive Video Text Tracking Algorithm based on SIFT Feature and Geometric Constraint", PCM, LNCS 6297, Part I, 2010, pp.392-403.

[6] P. Shivakumara, and C. L. Tan, "Novel Edge Features for Text Frame Classification in Video", ICPR, 2010, pp.3191-3194.

[7] P. Shivakumara, A. Dutta, T. Q. Phan, C. L.Tan, and U. Pal, "A novel mutual nearest neighbor based symmetry for text frame classification in video", Pattern Recognition, vol.44, 2011, pp. 1671-1683.

[8] V. C. Dinh, S. S. Chun, S. Cha, "An Efficient Method for Text Detection in Video Based on Stroke Width Similarity", ACCV, LNCS 4843, part-1, 2007, pp.200-209.

[9] C. Jung, Q. Liu, and J. Kim, "A new apporoach for text segmentation using a stroke filter", Signal Procesing, vol.88, 2008, pp.1907-1916.

[10] X. Li, W. Wang, S. Jiang, Q. Huang, and Wen Gao, "Fast and Effective text detection", ICIP, 2008, pp.969-972.

[11] C. Jung, Q. Liu, and J. Kim, "A stroke filter and its application to text localization", Pattern Recognition Letters, vol.30, 2009, pp.114-122.

[12] B. Epshtien, E. Ofek, Y. Wexler, "Detecting text in natural scenes with Stroke Width Transform", CVPR, 2010, pp. 2963 - 2970.

[13] P. Shivakumara, W. Huang, C. L. Tan, "An Efficient Edge Based Technique for Text Detection in Video Frames", DAS, 2008, pp.307-314.

[14] P. Shivakumara, T. Q. Phan, and C. L. Tan, " Video text detection based on filters and edge features", ICME, 2009, pp.514-517.

[15] M. Anthimopoulus, B. Gatos, and I. Pratikakis, "A Hybrid System for Text Detection in Video Frames", DAS, 2008, pp.286-292.

[16] M. Anthimopoulos, B. Gatos, and I. Prakakis, " A two-stage scheme for text detection in video images", Image and Vision Computing, vol.28, 2010, pp.1413-1426.

[17] Yi-Feng Pan, X. Hou, and C. L. Liu, "A Robust System to Detect and Localize Texts in Natural Scene Images", DAS, 2008, pp.35-42.

[18] W. Huang, P. Shivakumara, and C. L. Tan, "Detecting Moving Text in Video Using Temporal Information", ICPR, 2008, pp.1-4.

[19] Tsung-Han Tsai, Yung-Chien Chen, and Chih-Lun Fang, "2DVTE: A two-directional videotext extractor for rapid and elaborate design", Pattern Recognition, vol.42, 2009, pp.1496-1510.

[20] M. Li, C. Wang, "An Adaptive text detection approach in Images and Video Frames", IJCNN, 2008, pp.72-77.

[21] S. Shi, T. Cheng, S. Xiao, and X. Lv, " A smart Approach for text detection, localization, and extraction in video frames", Int. Conf. on IT and CS, 2009, pp.158-161.

[22] P. Shivakumara, T. Q. Phan, and C. L. Tan, "A gradient difference based technique for Video text detection", ICDAR, 2009, pp.156-160.

[23] M. B. Halima, H. Karry, and A. M. Alimi, "A Comprehensive method for Arabic Video text detection, Localization, extraction and recognition", PCM, LNCS 6298, part II, 2010, pp.648-659.

[24] M. Moradi, S. Mozaffari, A. A. Orouji, " Farsi/Arabic text extraction from Video Images by Corner detection", MVJP, 2010, pp.1-6.

[25] J. Park, G. Lee, E. Kim, J. Lim, S. Kim, H. Yang, M. Lee, and S. Hwang, "Automatic detection and recognition of Korean text in outdoor signboard Images", Pattern Recognition Letters, vol.31, 2010, pp.1728-1739.

[26] P. Shivakumara, W. Huang, T. Q. Phan, C. L. Tan, " Accurate video text detection through classification of low and high contrast Images", Pattern Recognition, vol.43, 2010, pp.2165-2185.

[27] P. Shivakumara, A. Dutta, U. Pal, and C. L. Tan, "A New method for Handwritten scene text detection in video", ICFHR, 2010, pp.387-392

[28] J. Yu and Y. Wang, "Apply SOM to Video Artificial text area detection", Int. Conf. Internet computing for Science.and Engg., 2010, pp.137-141.

[29] X. Huang and H. Ma, "Automatic Detection and Localization of Natural scene text in Video", ICPR, 2010, pp.3216-3219.

[30] D.S. Guru, S. Manjunath, P. Shivakumara, C.L.Tan, "An Eigen Value based Approach for Text Detection in Video",DAS, 2010, pp.501-505

[31] X. Zhang, and F. Sun, " Pulse Coupled Neural Network Edge-Based Algorithm for Image text locating", Tsinghua Science and Technology, vol.16, no.1, 2011, pp.22-30.

[32] M. Anthimopoulos, and B. Gatos, " Detection of artifical and scene text in images and video frames", Pattern Analysis and Applic., 2011, pp.1-16.

[33] M. S. Kumari, and B. H. Shekar, " On the use of Moravec Operator for Text Detection in Documents Images and Video Frames", ICRTIT, 2011, pp.910-914.

[34] Yan Zhao, Tong Lu, Wujun Liao,"A Robust Color-Independent Text Detection Method from Complex Videos",ICDAR, 2011, pp.374-378.

[35] A. Jamil, I. Siddiqi, F. Arif, and A. Raza "Edge-Based Features for Localization of Artificial Urdu Text in Video Images", ICDAR, 2011, pp.1120-1124.

[36] Yi-Feng Pan, Y. Zhu, J. Sun, and S. Naoi, "Improving Scene Text Detection by Scale-Adaptive Segmentation and Weighted CRF Verification", ICDAR, 2011, pp.759-763.

[37] S. Uchida, Y. Shigeyoshi, Y. Kunishige, and F. Yaokai, "A Keypoint-Based Approach toward Scenery Character Detection", ICDAR, 2011, pp.819-823.

[38] Q. Ye, J. Jiao, J. Huang, and Hua Yu, "Text detection and restoration in natual scene images", J. Visual Comm. & Image Rep., vol.18, 2007, pp.504-513.

[39] P. Shivakumara, T. Q. Phan, and C. L. Tan, "A Robust Wavelet transform based technique for Video text detection", ICDAR, 2009, pp.1285-1289.

[40] Z. Ji, J. Wang, and Yu-Ting Su, " Text Detection in Video frames using Hybrid Features", ICMLC, 2009, pp.318-322.

[41] M. Zhao, S. Li, and J. Kwok, "Text detection in Images using sparse representation with discriminative dictionaries", Image and Vision Computing, vol.28, 2010, pp.1590-1599.

[42] P. Shivakumara, T. Q. Phan, C. L. Tan, "New wavelet and color features for text detection in Video", ICPR, 2010, pp.3996-3999.

[43] Chucai Yi and Yingli Tian, "Text Detection in Natural Scene Images by Stroke Gabor Words", ICDAR, 2011, pp.177-181.

[44] X. Peng, H. Cao, R. Prasad, and P. Natarajan "Text Extraction from Video Using Conditional Random Fields", ICDAR, 2011, pp.1029-1033.

[45] R. Ji, Pengfei Xu, H. Yao, Z. Zhang, X. Sun, T. Liu, "Directional Correlation Analysis of Local Haar Binary Pattern for Text Detection", ICME, 2008, pp.885-888.

[46] S. Muhammad Hanif, L. Prevost, "Text Detection in Natural Scene Images using Spatial Histograms", CBDAR, 2007, pp.122-129.

[47] P. Shivakumara, T. Q. Phan, C. L. Tan, "A Laplacian Approach to Multi-Oriented Text Detection in Video" IEEE Trans. on PAMI, vol.33, no.2, 2011, pp.412-419.

[48] P. Shivakumara, T. Q. Phan, C. L. Tan, "New Fourier-Statistical Features in RGB Space for Video Text Detection," IEEE Trans. On CSV, vol.20, no.11, 2010, pp.1520-1532.

[49] P. Shivakumara, A. Dutta, C. L. Tan, U. Pal, "A New Wavelet-Median-Moment based method for Multi-oriented Video text Detection", DAS, 2010, pp-279-286.

[50] T. Q. Phan, P. Shivakumara, C. L. Tan, "A skeleton-based Method for Multi-oriented Video Text Detection", DAS, 2010, pp.271-278.

[51] X. Qian, G. Liu, H. Wang, and R. Su, "Text Detection, Localization, and tracking in compressed video", Signal Processing: Image Comm., vol.22, 2007, pp.752-768.

[52] J. Lim, J. Park, and G. G. Medioni, "Text segmentation in color images using tensor voting", Image and Vison Computing, vol.25, 2007, pp.-671-685.

[53] M. Su. Cho, Jae-Hyun Seok, S. Lee, and J. H. Kim, "Scene Text Extraction by Superpixel CRFs Combining Multiple Character Features", ICDAR, 2011, pp.1034-1038.

[54] H. Zhang, C. Liu, C. Yang, X. Ding, and K. Wang, "An Improved Scene Text Extraction Method Using Conditional Random Field and Optical Character Recognition", ICDAR, 2011, pp.708-712.

[55] P. Shivakumara, S. Bhowmick, B. Su, C. L. Tan, U. Pal, "A New Gradient based character segmentation Method for Video text Recognition", ICDAR, 2011, pp.-126-130.

[56] T. Q. Phan, P. Shivakumara, B. Su, and C. L. Tan, "A Gradient Vector Flow-Based Method for Video Character Segmentation", ICDAR, 2011, pp.1024-1028.

[57] Z. Zhou, L. Li, C. L. Tan, "Edge based Binarization for video text images", ICPR, 2010, pp.133-136.

[58] A. Mishra, K Alahari, and C. V. Jawahar ,"An MRF Model for Binarization of Natural Scene Text", ICDAR, 2011, pp-11-16.

[59] Toru Wakahara and Kohei Kita, "Binarization of Color Character Strings in Scene Images Using K-Means Clustering and Support Vector Machines", ICDAR, 2011, pp.274-278.

[60] K. Ntirogiannis, B. Gatos, and I. Pratikakis "Binarization of Textual Content in Video Frames", ICDAR, 2011, pp.673-677.

[61] D. Rajendran, P. Shivakumara, B. Su, S. Lu, and C. L. Tan, "A new Fourier-Moments based Video Word and Character Extraction Method for recognition", ICDAR, 2011, pp.1165-1169.

[62] S. Uchida, H. Miyazaki, H. Sakoe, "Mosaicing-by-recognition for video-based text recognition", Pattern Recognition, vol.41, 2008, pp.1230-1240.

[63] Z. Saidane and C. Gracia, "Automatic Scene Text Recognition using a Convolutional Neural Network", CBDAR, 2007, pp.100-107.

[64] A. Ohkura, D. Deguchi, T. Takahashi, I. Ide, and H. Murasse, "Low-resolution Character Recognition by Video-based Super-resolution", ICDAR, 2009, pp.191-19.

[65] Z. Saidane, C. Garcia, and J. L. Dugelay, "The image Text Recognition Graph (iTRG)", ICME, 2009, pp.266-269.

[66] M. Iwamura, T. Tsuji, K. Kise, "Memory-Based Recognition of Camera-captured characters", DAS, 2010, pp.89-96.

[67] P. Shivakumara, T. Q. Phan, S. Lu, and C. L. Tan, "Video Character Recognition through Hierarchical Classification", ICDAR, 2011, pp.131-135.

[68] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu, and A. Y. Ng, "Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning", ICDAR, 2011, pp.440-445.