Robust Model for Text Extraction from Complex Video Inputs Based on SUSAN Contour Detection and Fuzzy C Means Clustering

Narasimha Murthy K N[1], Dr. Y S Kumaraswamy[2],
[1] Professor, Dept of Information Science and Engineering, V T U, City Engineering College, Bangalore, Karnataka, Pin-560062, India,

[2] Professor & HOD, Dept. of MAC (VTU), V T U, Dayananda Sagar College of Engineering, Bangalore, Karnataka, Pin-560078, India,

Abstract:

The proposed system introduces a novel approach for extracting text effectively from different types of complex video inputs. The valuable information within the text can be deployed for text indexing and localization. The proposed system uses contour based protocol like SUSAN algorithm for evaluating the contour detection. The system then explores candidate text area and refines the edges by Fuzzy C Means Clustering. The unwanted non-text portions are removed using morphological operation like dilation. The results obtained from the proposed implementation are then compared with the traditional algorithm used in prior research work for evaluating its efficiency. The result achieved outperforms all the prior algorithms for extracting text from different types of complex video input.

Keyword: *Text Extraction, SUSAN, Fuzzy C Means Clustering, Morphological Operations*

1. Introduction

With the commercial increase in the videos on the various networking and multimedia applications, various service providers are showing increasing attention towards archiving the digital contents for various value added services[1]. The presence of text in video frame is very precious source of high-level semantics and content understanding. For achieving this purpose, text localization and extraction from the videos is the mostly considered option. According to the prior work, text extraction method can be classified into connected component-based approach [2], boundary based approach [3-7], and texture based approach [8-12]. However, component based approach is not so result oriented as it doesn't operated optimally for any video images because of its assumptions that text image elements in the same area will have equivalent color as well as gray-scale strength [2]. The boundary based approach will require the text to be logically elevated contrast for identifying boundaries [3-7]. The texture based approach frequently uses FFT, DCT, wavelets, and Gabor filter for extracting features [13]. Unfortunately, such methods requires high amount of training which is definitely not cost-effective for large scale deputation. Majority of such techniques has also yielded to high rate of false positives results, which motivate the researcher to work on much more efficient algorithm with most challenging deployment environment with highest accuracy and less false positives in the area of text-extraction from videos. Moreover, extracting text from video with higher accuracy is much more challenging in comparison to text extraction from the images, because of diversified problems yielding from complex background, fast moving text in videos, videos with multiple language in one instant, contrast, etc.

Owing to the above mentioned issues, the proposed system will attempt to design a framework for extracting text using SUSAN algorithm for contour based detection, morphological operations like dilation, and Fuzzy C-Means Clustering for much accurate results. Finally the results obtained from the proposed system has been compared with the 5 significant prior research work for establishing the fact that the proposed system has outperformed the existing algorithms for text extractions. We discuss related work in Section II. The research methodology is discussed in Section-III. Proposed system is elaborated in Section IV. Implementation and Results is described in Section-V. Performance Analysis of the proposed system is discussed in Section-VI and finally conclusion and future work is described in Section-VII

2. Related Work

Pratheeba e.t. al [14] has proposed a unique technique for text localization and extraction from complex video input based on findings that there persist colors of contrast nature between text and its adjacent background.

Ghosh e.t. al [15] has proposed an analytical architecture for automated monitoring of news videos with multiple languages. The system combines the audio and visual charecteristics for identifying keywords which characterize a specific news.

Abburu [16] has proposed and analyzed DLER tool for integrating the text identification, localization, extraction, and the recognition method in a single tool

Yen e.t. al [17] has highlighted an effective text extraction algorithm using news video using the temporal information of the video and logical AND operation for removing the most irrelevant background

Qiujun [18] has described a technique for extracting news contents from the web pages based on various non-complex charecteristics seen in majority of the frequently visited websites. The significant feature is the similarity of the dual pages which are gathered from the equivalent topic of a website and published on the same date.

Vijayakumar [19] has proposed an effective text extraction algorithm from sports video. The system can only identify the text in video in the edge of the image. The author has used key frames from the video by color histogram procedure for minimizing the quantity of the video frames.

Stefanos e.t. al. [20] have introduces novel implicit interest indicator for video search and described a new procedure to designing a content similarity graph based on the implicit indicators of patterns of user interaction using SVM classifier.

3. Research Methodology

The proposed system describes a robust text localization and extraction methods from input real-time video with complex background. The text localization and extraction of the proposed system will consists of video frame identification and extraction followed by gray-scale conversion, video segmentation, binarization, contour detection and finally text region extraction. The input from running video will consists of several video frames which will be extracted. Here the video text will be classified into two types: firstly, the type of text which will not change much with the running of the video (e.g. channel name, news headings, date, time etc) and secondly, the type of text which will randomly change in very second as a matter of news update. The situation very complex, when the sports video or news video is considered. The various types of the scene text will be visible only within the frame showing all the scenes like text on hoarding, streets, shops, accessories, vehicles etc. Such types of the text will be subsidiary to the contents on the scene and will be worthy only in certain cases of applications using monitoring or scanning text visible on our known objects. This cannot be use in general text indexing and recovery. Such types of presence of text will be very challenging towards effective text extraction process as it might be observed in infinite quantity with various orientation and structures. Therefore, if such types of occurrences of complex text can be involuntarily identified and extracted, it will lead to evolution to a very higher level of text extraction algorithm which will have higher scope depending on the usage. The need of researching in this types of text extraction is of very high priority as the text extraction tool will effectively be able to capture the current contents of the text without any false positives as it has very significant illustrative charecteristics. With the implementation of such fast algorithm for text localization and extraction, algorithm can be designed for data-mining the text contents in future using certain content based text extraction approach, it can also be used in real-time monitoring system more effectively with zero error in its results. The various charecteristics of the text contents in the video can be discussed as followings:

(i). Configuration: The configuration will normally consist of size and positioning of the text. The proposed system has considered different types of real-time videos where there is a positive feasibility of occurrence of different sizes of text appearing in the same video frame. The proposed system has also consideration of diversified positioning for making the system more challenging. Majority of the prior research work has not much consideration of positioning. Normally, the textual characters are placed either horizontal or vertical (Japanese text). Certain inclined text can also be observed in most of the TV channel logos. Therefore, in order to make most robust text extraction system, we have considered all the possible orientation of the text for much contrast results. Another consideration for our proposed system is also the distance between two different characters or words, which can be maximum time uniform and sometime non-uniform.

(ii) Speed: This is one of the important consideration where we identify the characters which can exist in different frames of video sometime with or without motion. This phenomenal characteristic will be used for tracking text and its improvement.

(iii) Color: There is also a feasibility that the same text or character may appear in same or different colors in every consecutive video frames.
(iv) Boundary: Majority of the text extraction process is created for non-complex readability which results in tough boundary for textual matter and its complex background too. The proposed system has both of these factors considered.

4. Proposed System

The proposed system highlights a robust technique of text extraction from real time videos using contour detection based algorithm like SUSAN [21]. The proposed system has identified contour points as one of the significant charecteristics for video frame which is associated with maximum changes in luminance or in curve direction [22]. Here the strength of image elements inside the window will be compared with that of the center which is also known as Nucleus. In case the comparison yields smaller value than threshold, then the current picture elements will be considered having the similar strength with the center of picture elements. The USAN (Univalue Segment Assimilating Nucleus) will create region with such elements. The contour detection methodology in the proposed system can be formulated as following algorithm:

Algorithm-1: SUSAN
Objective: The SUSAN edge detector has been implemented using circular windows to give isotropic responses
Input: image, threshold for brightness and USAN
Output: Implementation of SUSAN algorithm by setting up circular mask along with removal of close corners.
Steps:
1 *Create a function for Susan*
2 *Check inputs and fill in the blank variables*
3 *Check if the inputs are empty*
4 *Convert to double image format*
5 *Prepare the inputs and outputs*
6 *Create the brightness look up table (LUT)*
7 *Set up the variables*
8 *Set up the circular masks*
9 *Create 37 pixel circular mask*
10 *37 pixel circular mask for x*
11 *37 pixel circular mask for y*
12 *Compute the USAN response*
13 *Compute correlation for the window mask w*
14 *If already too big - ignore it*
15 *Compute correlation for the window mask wx*
16 *Compute correlation for the window mask wy*
17 *Compute the sq response*
18 *Check the centre of gravity*
19 *Threshold the response to find the corners*
20 *Perform nonmaximal suppression (5x5 mask)*
21 *Find the local maxima*
22 *Corner must be local maxima*
23 *Initialized removal of any close corners*
24 *5x5 neighborhood mask (12 points)*
25 *Remove close corners*

The contour function is formulated considering arguments like detected contour point for identification, and width and height of the window. A morphological operation like dilation is performed for getting the associated regions in video frames, once the algorithm successfully accomplishes the contour points with text. In the initial stage of processing, certain non-text area will be removed. Finally the segmentation is created for the multiple lines in the frames using vertical and horizontal ridge of the contour point. Finally, all the identified text regions are resized to initial size of the video frame.

The candidate text area captured is refined using Fuzzy C-Means Clustering algorithm.

Algorithm-2: Fuzzy C-Means Clustering Algorithm
Objective: In fuzzy clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be in the cluster to a lesser degree than points in the center of cluster.
Input: initialization of clusters.
Output: creation of clusters for text refinement
Steps:
1 *Create a function for Fuzzy C-Means Cluster*
2 *Choose a number of clusters.*
3 *Assign randomly to each point coefficients for being in the clusters.*
4 *Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than* ε*, the given sensitivity threshold) :*
5 *Compute the centroid for each cluster*
6 *For each point, compute its coefficients of being in the clusters*
7 *Create cluster*
8 *Calculate number of rows and columns for subplot*
9 *Display clusters*

Algorithm-3: Text Extraction using Fuzzy C-Means Clustering.
Objective: The main objective of this program is to extract text region from the input video and display the sub-images along with implementation of Fuzzy C-Means clustering
Input: video and initialization of threshold for brightness and USAN

Output: The output shows sub-images, with separate letters

Steps:

1 *Input the video file*
2 *Use multimedia reader method to read the video*
3 *Initialize brightness threshold.*
4 *Initialize USAN threshold*
5 *Defined Matrix for non text corner points removal*
6 *Convert the video to frames and read the frames*
7 *Call process frame function*
8 *Use image resize method*
9 *Use sub plot method*
10 *Display the sub-image*
11 *Plot a rectangle*
12 *Region Merging*
13 *Separate colors*
14 *Quantize the colors to 16 x 16 x 16 values*
15 *Plot histogram for R, G and B*
16 *merge all color matrix to form one image*
17 *Extract text region*
18 *Use image crop method*
19 *Apply Fuzzy C-means clustering*
20 *Create cluster*
21 *Calculate number of rows and columns for subplot*
22 *Display clusters*
23 *Show separate letters*

The processing of the algorithm will result in candidate text area. Finally, the clustering of the associated regions are estimated. Finally creation of the sub-images will result in merger image which will also give rise to the quantized image, i.e. the characters are merged in order to generate the text line.

5. Implementation and Results

The framework project work is designed in Matlab in 32 bit system 1.8 GHz with dual core processor where different real time videos are considered for the experiment. The basic graphics video display card of DIAMOND AMD ATI Radeon is used for experimenting on both OS of Windows Vista and Windows 7. The implementation also considers videos with single text, multiple text, text with different sizes of fonts, text with complex and simple background, text with different languages.

For the purpose of the experiment, different types of real time video inputs were considered:

1. Video clips of short / long interval
2. Video clips with fast and slow moving text.
3. Video clips with text appearing in different orientations (horizontal, vertical, slanted).
4. Video clips with multiple languages (English, Hindi)

In the preliminary experiment, a video file in AVI format is chosen from a news channel which was the 1st consideration of the video input. The real time video is read by multimedia reader object. The program initialized the brightness threshold and USAN threshold as 20 and 2000 respectively (in Algorithm 2 and 3). A matrix is created for non-text corner point removal. Various sub-images are created which will all identify the dynamic text appearing at every duration.

Fig 1. A video frame.

Fig 2(a) Sub-image-1

Fig 2(b) Sub-image-2

Fig 2(c) Sub-image-3

Fig 2. Creation of 3 sub-images from the respective video input as shown in Fig 1.

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

Fig 3. Merger Region output

Fig 4. Text Localization and Extraction

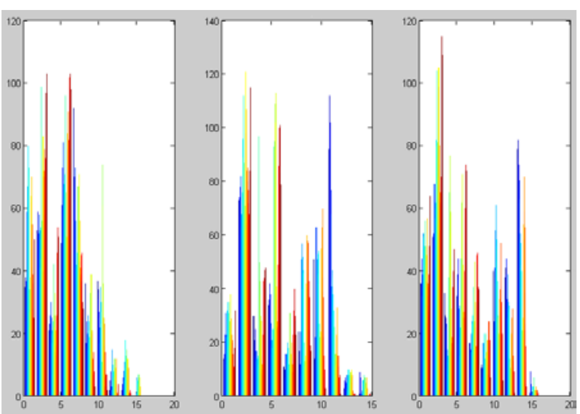

Fig 5. Graph created for Red, Green, and Blue Histogram based color clustering

A text region merging algorithm then identifies the text in each sub-image which will be finally magnified to original size. The R, G, B colors are separated and quantized to 16 x 16 x16 values which gives rise to histogram as shown in Fig 5. Finally all the color matrix are quantized to form one image. As seen in sample news video, that there will exist certain text e.g. Logo text, video caption, time, date, update headings etc, which will not change much for longer duration of play, it might give rise to overlapping and interleaving as the same set of text will be extracted every time by the algorithm. Therefore the intersection regions are selected to avoid this. The final extraction of the text region is done using Fuzzy C-Means Clustering. The segmentation considering all the color constituents is designed with classification based on text and non-text regions. The fuzzy clustering algorithm is one of the efficient data clustering algorithm [23], which is an iteratively most favorable protocol normally with consideration of least square method to design a fuzzy partition of the datasets. The process of iteration halts when the variation between the two consecutive iterations becomes very insignificant. The advantage of this approach is that the two average vectors can be confirmed as the two leading groups e.g. text and non-text regions. Therefore the edge histogram distribution for the detected text regions is estimated, as shown in Fig 5, in order to binarize the text region for effective text extraction.

The second set of the implementation is conducted from the second consideration of video input where the video are sometime very slow and suddenly very fast. This experiment is conducted to measure the efficiency of the algorithm for text extraction for fast moving text in the video. The video input taken for this experiment shows a text “DRENCHING RAIN” and “HOLLYWOOD” which appears onscreen in less than 2 seconds.

Fig 6. Frame Captured from the video

The application created for this set of evaluation reads the video which calls SUSAN function to identify the corner points using morphological structuring elements. The masking application as shown in Fig 7(b) is created from the corner for each generated frames from the input video. The dilation operation is used here which is used for morphologically open binary image for removing unwanted text region.

Finally, the effective text region is located and extracted as shown in Fig 7(c)-(d).

Fig 7(b) Mask Created from corners

Fig 7(c). Detected Text Region

Fig 7(d). Final Output

The next set of the experiment is conducted with 3rd and 4th consideration of the video input. The input video is captured from the TV tuning card for the live broadcasting of the TV show. The significant fact here is the text appearing in this set of video are in different size, style, orientation, as well as language (English and Hindi).

Fig 8 (a). A Video Frame captured

Fig 8(b) Creation of Mask from corners

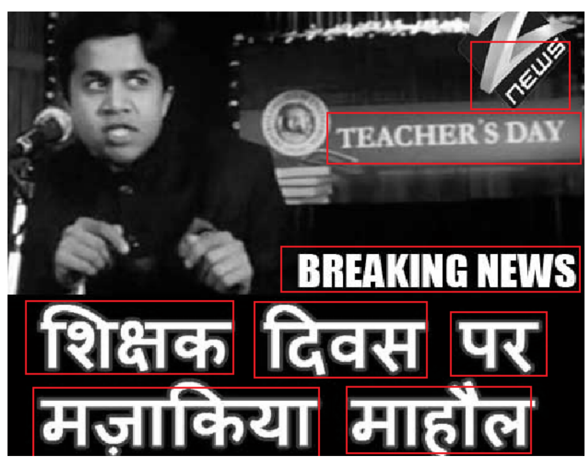

Fig 8 (c). Effective Text Localization

It can be noticed here that almost all the text (both in Hindi and English) is 100% accurately extracted in all implementation of masking. Interesting fact is the English text on the top right side of the channel logo which has different orientation or inclination in comparison to other text is also extract with 95.2 % accuracy. Therefore, the system design for the proposed text extraction can be eventually considered as robust, reliable, and efficient in text extraction in multiple scenarios as understood from the final output as shown in Fig 8(d).

Fig 8 (d). Effective Text Extraction

6. Performance Comparison

For the purpose of comparative analysis, various prior significant researches works in the area of text localization and text extraction has been considered. The proposed system has also being compared and experimented with all the major significant previous research work like:

- *Prior-Work-1*: Qixiang e.t. al (2004) [24] approach has used SVM edge and wavelet features. This is one of the significant work which has better performance in comparison to K-Means clustering. It also has consideration of eastern and western languages.
  - *Comparison to our Approach*: The above discussed approach has no consideration of text orientation as well as fast moving videos. Besides the approach has used multiple number of complex algorithm which consumes much time and definitely not cost effective in terms of text extraction. Our approach has simple use of Morphological operations and Fuzzy C Means Clustering, which effectively works for fast moving videos as well as with text with different orientations.
- *Prior-Work 2*: Matko e.t. al. (2008) [25] approach has worked on player number localization using HSV Color Space and Internal Contours.
  - *Comparison to our Approach*: The result of the above approach has only localization rate (83%) which is considerably higher than recognition rate (52%). This difference is caused by sensitivity of OCR software to non-rigid deformation, noisy character borders etc. Errors in number localization occur has occurred due to skewed numbers, folded jerseys and especially blur which is caused by player or camera motion. But our approach with efficient use of SUSAN algorithm for contour detection has effectively able to overcome all the issues and complexities found in this work.
- *Prior-Work 3*: Shivakumara e.t. al (2009) [26] approach of text detection from video using heuristics rules, initial text block identification, text portion segmentation and new edge features for false positive elimination.
  - *Comparison to our Approach*: The above discussed approach has not considered complex background images. Moreover all the experiment result were derived from testing with English Text and no consideration of text orientation. Whereas our approach has clearly outperformed this results.
- *Prior-Work 4*: Phan e.t. al. (2009) [27] approach of text detection method based on the Laplacian operator and use of K-means Clustering.
  - *Comparison to our Approach*: The work discussed has no consideration of arbitrary orientation. Moreover, the text detection step shows white patches even for non-horizontal text. Where our approach facilitates much more accurate results of text localization and extraction in much more challenging scenarios.
- *Prior-Work 5*: Ghorpade e.t. al (2011) [28] approach with neural network pattern matching technique for text localization, segmentation, and recognition.
  - *Comparison to our Approach*: In this work, as, the characters are recognized on run-time basis, there may be a few cases found in which one or two characters may get misrecognised i.e. a character may get recognized as some other character. Moreover only English text is considered. The process is time consuming due to inclusion of neural network and may not give

the best result in this work. Our proposed work obviously is in much contrast results in comparison to this prior work.

The empirical effectiveness of our proposed algorithm can be derived by estimating the accuracy in detection of the text contents from the input videos in the application designed. The following parameters are considered for this purpose:

- Actually Recognized Text Region (ARTR) consisting of text line.
- False Identified Text Region (FITR) which do not contain any text.
- Missing Text (MT) which ignores some text characters.

The approach manually estimates the Original text region (OTR) where the Text Identification Rate (TIR) can be evaluated as:

$$TIR = ARTR / OTR$$

And Error in Text Identification rate (ETIR) as given by,

$$ETIR = FITR / (ARTR + FITR)$$

and

$$\text{Non-Text Identification Rate (NTIR)} = MT / ARTR$$

The comparative analysis is done by estimating the above empirical parameters for the proposed work with all the significant research work specified in this paper. Same sets of the input type and considerations are made towards the analysis.

Table 1. Comparative Performance analysis Parameters with 5 prior research work and proposed system.

Prior Work	**OTR**	**ARTR**	**FITR**	**MT**
Qixiang e.t. al (2004) [24]	500	393	87	79
Matko e.t. al. (2008) [25]	500	349	50	35
Shivakumara e.t. al (2009) [26]	500	251	94	94
Phan e.t. al. (2009) [A27	500	458	39	55
Ghorpade e.t. al (2011) [28]	500	350	55	70
Proposed Work	500	485	38	37

Table 2. Estimation of TIR, ETIR, and NTIR for all the 5 prior research work with proposed system.

Method	**TIR**	**ETIR**	**NTIR**
Qixiang e.t. al (2004) [AR]	78.6	18.12	2.01
Matko e.t. al. (2008) [AR]	69.8	12.53	1.00
Shivakumara e.t. al (2009) [AR]	50.2	27.72	3.74
Phan e.t. al. (2009) [AR]	91.6	7.8	1.2
Ghorpade e.t. al (2011) [AR]	70.6	13.5	2
Proposed Approach	**97.0**	**7.0**	7.0

From the above table, it can be easily identified that proposed system has better Text identification rate (97%) as well as less Error in Text Identification rate (7%) as compared with the prior research work stated.

7. Conclusion

The proposed system highlights a unique technique for text extraction system using contour based SUSAN algorithm and Fuzzy C Means Clustering algorithm. The proposed system is potential enough to extracting text from different types of complex and fast moving text in videos. Even multiple language is also considered for the checking the effectiveness of the algorithm. The proposed system has achieved 97% of Text Identification rate and only 7% of Error in Text Identification Rate. The achieved result is much in contrast compared to majority of prior significant research in this area.

Reference

[1] Dimitris N. Kanellopoulos, Adapting Multimedia Streaming to Changing Network Conditions, IGI Global. 2011
[2] A.K. Jain and B. Yu, “Automatic Text Location in Images and Video Frames”, Pattern Recognition, Vol. 31(12), pp. 2055-2076, 1998.

[3] M. Anthimopoulos, B. Gatos and I. Pratikakis, “A Hybrid System for Text Detection in Video Frames”, The Eighth IAPR Workshop on Document Analysis Systems (DAS2008), Nara, Japan, , pp 286-293, September 2008.
[4] M. R. Lyu, J. Song and M. Cai, “A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 2, pp 243-255, February 2005.
[5] C. Liu, C. Wang and R. Dai, “Text Detection in Images Based on Unsupervised Classification of Edge-based Features” , pp. 610-614, ICDAR 2005.
[6] E. K. Wong and M. Chen, “A new robust algorithm for video text extraction”, Pattern Recognition 36, pp. 1397-1406, 2003.
[7] P. Shivakumara, W. Huang and C. L. Tan, “An Efficient Edge based Technique for Text Detection in Video Frames”, The Eighth IAPR Workshop on Document Analysis Systems (DAS2008), Nara, Japan, pp 307-314, September 2008.
[8] Y. Zhong, H. Zhang and A.K. Jain, “Automatic Caption Localization in Compressed Video”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, No. 4, pp. 385-392, 2000.
[9] K. L Kim, K. Jung and J. H. Kim, “Texture-Based Approach for Text Detection in Images using Support Vector Machines and Continuous Adaptive Mean Shift Algorithm”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 12, pp 1631-1639, December 2003.
[10] Q. Ye, Q. Huang, W. Gao and D. Zhao, “Fast and robust text detection in images and video frames”, Image and Vision Computing 23, pp. 565-576, 2005.
[11] H. Li, D. Doermann and O. Kia, “Automatic Text Detection and Tracking in Digital Video”, IEEE Transactions on Image Processing, Vol. 9, No. 1, pp 147-156, January 2000.
[12] W. Mao, F. Chung, K. K. M. Lam and W. Siu, “Hybrid Chinese/English Text Detection in Images and Video Frames”, ICPR, Volume 3, pp 1015- 1018, 2002.
[13] H. B. Kekre, V. A. Bharadi, P. Roongta, S. Khandelwal, V. I. Singh, S. Gupta, P. P. Janrao , Performance Comparison of DCT, FFT, WHT, Kekre’s Transform & Gabor Filter Based Feature Vectors for On-Line Signature Recognition, International Conference and Workshop on Emerging Trends in Technology, 2011
[14] T.Pratheeba , Dr.V.Kavitha , S.Raja Rajeswari, Morphology Based Text Detection and Extraction from Complex Video Scene, International Journal of Engineering and Technology Vol.2(3), pp.200-206, 2010
[15] Hiranmay Ghosh, Sunil Kumar Kopparapu, Tanushyam Chattopadhyay, Ashish Khare, Sujal SubhashWattamwar, Amarendra Gorai, and Meghna Pandharipande, Multimodal Indexing ofMultilingual News Video, International Journal of Digital Multimedia Broadcasting Volume 2010.
[16] Sunitha Abburu, Multi Level Semantic Extraction for Cricket Video By Text Processing, International Journal of Engineering Science and Technology Vol. 2(10), 2010
[17] Shwu-Huey Yen, A, Hsiao-Wei Chang, B, Chia-Jen Wang,C, Chun-Wei Wang,d, Robust News Video Text Detection Based on Edges and Line-deletion, WSEAS Transactions on SIGNAL PROCESSING, Issue 4, Volume 6, October 2010
[18] Qiujun Lan, Extraction of News Content for Text Mining Based on Edit Distance, Journal of Computational Information Systems 6:11, pp. 3761-3777, 2010
[19] V.Vijayakumar, R.Nedunchezhian, A Novel Method for Super Imposed Text Extraction in a
Sports Video, International Journal of Computer Applications (0975 – 8887) Volume 15– No.1, February 2011
[20] Stefanos Vrochidis, Ioannis Kompatsiaris, and Ioannis Patras, Utilizing Implicit User Feedback to Improve Interactive Video Retrieval, Advances in Multimedia Volume, 2011.
[21] Shenghua Xu; Litao Han; Lihua Zhang, An Algorithm to Edge Detection Based on SUSAN Filter and Embedded Confidence, Intelligent Systems Design and Applications, ISDA '06. Sixth International Conference, 2006
[22] M. Bertini , C. Colombo , A. Del Bimbo , Via S. Marta, Automatic caption localization in videos using salient points, Proceedings of IEEE International Conference on Multimedia and Expo ICME, 2001
[23] Moh'd Belal Al-Zoubi, Amjad Hudaib, Bashar Al-Shboul, A fast fuzzy clustering algorithm, Proceeding AIKED'07 Proceedings of the 6th Conference on 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases - Volume 6, 2007
[24] Qixiang Ye, Wen Gao1, Weiqiang Wang, Wei Zeng, A Robust Text Detection Algorithm in Images and Video Frames, Information, Communications and Signal Processing, the Fourth Pacific Rim Conference on Multimedia. 2003
[25] Matko Šari, Hrvoje Dujmi, Vladan Papi and Nikola Roži, Player Number Localization and Recognition in Soccer Video using HSV Color Space and Internal Contours, World Academy of Science, Engineering and Technology 43, 2008
[26] Palaiahnakote Shivakumara, Trung Quy Phan and Chew Lim Tan, Video Text Detection Based On Filters and Edge Features, Pattern Recognition, 19th International Conference ICPR, 2008.

[27] Trung Quy Phan, Palaiahnakote Shivakumara and Chew Lim Tan, A Laplacian Method for Video Text Detection, 10th International Conference on Document Analysis and Recognition, 2009
[28] Jayshree Ghorpade, Raviraj Palvankar, Ajinkya Patankar and Snehal Rathi , Extracting Text From Video, Signal & Image Processing : An International Journal (SIPIJ) Vol.2, No.2, June 2011

Mr. K N Narasimha Murthy is an experienced teacher for more than two decades and presently working as a Professor and HOD in department of Information Science and Engineering at City Engineering college Bangalore, Affiliated to Visvesvaraya Technological University, India. He got BE degree in 1989, M.Tech degree in 1993, MCA degree in 2003 and M.Tech degree in 2006. His areas of interest are Algorithms, Image processing and Computer Vision and also a member of MIE, MISTE, IACSIT.

Dr. Y S Kumaraswamy working presently as Sr. professor and HOD Department of MCA (VTU) at Dayandasagar college of Engineering Bangalore, India. He has published more than 150 research papers and guided 38 PhD students in the area of Computer Science/ Mathematics and also a selection committee member for ISRO/UGC/DSI profiles