

Text Based Approach For Indexing And Retrieval Of Image And Video: A Review

Avinash N Bhute and B.B. Meshram

VJTI, Matunga, Mumbai-19

ABSTRACT

Text data present in multimedia contain useful information for automatic annotation, indexing. Extracted information used for recognition of the overlay or scene text from a given video or image. The Extracted text can be used for retrieving the videos and images. In this paper, firstly, we are discussed the different techniques for text extraction from images and videos. Secondly, we are reviewed the techniques for indexing and retrieval of image and videos by using extracted text.

KEYWORDS

Text Extraction, Image Text, Video Text, Image Indexing, Video Indexing, Video Retrieval, Image Retrieval

1. INTRODUCTION

Text data present in multimedia viz. video and images contain useful information for automatic annotation, indexing. The Process of Extraction of information is detection, localization, tracking, extraction, enhancement, and recognition of the text from a given image [1]. However, there are differences in text in style, orientation, size, and alignment, as well as low contrast image and complex background make the automatic text extraction problem more difficult and time consuming. While critical surveys of related problems such as document analysis face detection and image & video indexing and retrieval can be found, the problem of text extraction isn't surveyed well. A variety of approaches to text extraction from images and video have been presented for many applications like address block location [6], content-based image/video indexing [2,8], page segmentation [4,5], and license plate location [3,7]. In spite of such in critical studies, it is still not easy to design a general-purpose Text Extraction system. This is often a result of so many possible sources of variation once extracting text from complex images, or from images having difference in style, color, orientation, font size and alignment.

Although images non-inheritable by scanning book covers, CD covers, or different multi-colored documents have almost similar characteristics as the document images, Author(s) cannot be dealt with conventional document image analysis technique. Text in video images can classify into caption text and scene text. The caption text is artificially overlaid on the image and scene text exists naturally in the images. Some researchers prefer to use the term 'graphics text' for scene text, and 'superimposed text' or 'artificial text' for caption text [9,10]. It is documented that scene text is harder to detect.

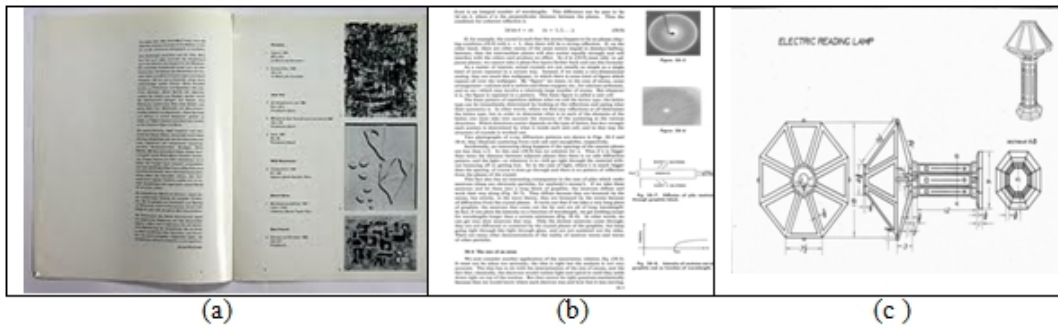


Fig. 1 Grayscale document images: (a) single-column text from a book, (b) two-column page from a book (c) Electric Lamp Drawing

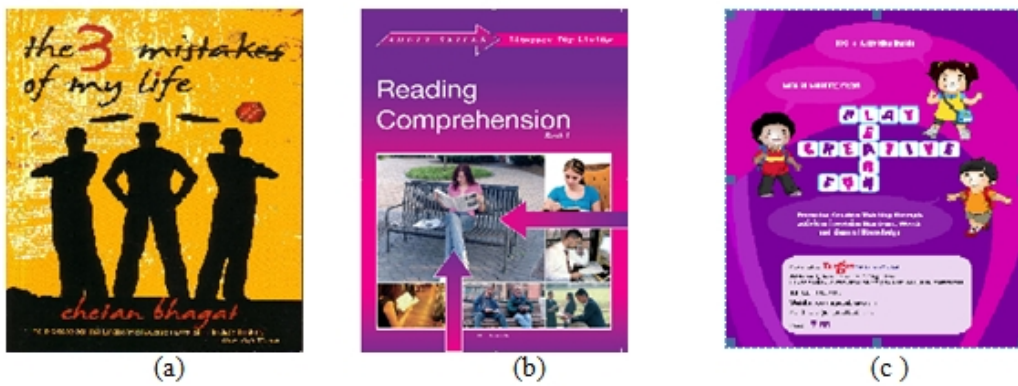


Fig. 2. Multi-color document images: each text line may or may not be of the same color.



Fig. 3 Images with Caption Text

A Text Extraction system accepts an input as a still image or a sequence of frames. The frames can be in gray scale or color, compressed or un-compressed, and also the text within the frames could or might not move. Text in frames will exhibit many variations according to their properties such as Geometry (size, alignment, inter-character distance), Color(monochrome, polychrome), Motion(static, linear moment), Edge(text boundaries, strong edges), Compression, etc.

Organization of paper In section-I we are introduced the concept of text extraction from images/video as well as text based image video retrieval. Section two discussed the text

extraction from images and videos with its process such as Text Tracking (TT), Text Detection (TD), Text Localization (TL), Text Extraction and Enhancement (TEE), Optical Character Recognition (OCR). Section-III deals with techniques and approaches for text based image and video indexing and retrieval systems. Finally Section IV has summary followed by conclusion.

2. TEXT EXTRACTION FROM IMAGES AND VIDEOS

The process for extraction carries detection, localization, tracking, extraction, enhancement, and recognition of the text [1] from a given image as shown in figure 1. Text detection refers to the identification of text in a given video frame. Text localization is refer to determine location of text in the frame or sequence of video frame. Text tracking is performed to cut back the interval for text localization and to maintain the position across adjacent frames. The precise location of text in a frame indicated by bounding boxes, the text still has to be divided from the background to useful for its recognition. Text extraction is that the stage wherever the text components are segmented from the background. Extracted text components are required enhancement because the text region usually has low-resolution and is susceptible to noise. The extracted text images are transformed into plain text using OCR technology.

2.1 Text Detection

The text of input images need to be identified as the input image contains any text, the existence or non-existence of text among the image. Several approaches assuming certain types of video frame or image contain text (e.g., recording cases or book covers). However, in the case of video, the amount of frames containing text is far smaller than the amount of frames while not text. The text detection stage detects the text in image. Kim [11] chosen a frame from shots detected by a scene-change detection methodology as a candidate containing text. Kim's scene-change detection methodology isn't delineated thoroughly in his literature, he has mention that low threshold values are required for scene-change detection as a result of a very small region of images the occupied by text regions. This approach is incredibly sensitive to scene-change detection. Text-frame choice is performed at associate interval of two seconds for caption text within the detected scene frames. This can be an easy and economical resolution for video indexing applications that only needs keywords from video clips, instead of the whole text.

Smith and Kanade [12] detected the text information based on scene-change which allows the distinction between two consecutive. Author(s) achieved accuracy of 90% in scene-change detection. Gargi et al. [13,14] performed text detection by assuming the amount of intra-coded blocks in P- and B- frames of an MPEG compressed video will increase, once a text caption appears. Lim et al. [15] made an easy assumption that text typically includes a higher intensity than the background. Author(s) counted the number pixels that are less than a threshold and exhibited a big color distinction to their neighborhood. This methodology is very simple and fast. However, issues will occur with color-reversed text. A text localization used for detecting the text in image/video.

Zhong et al. [16], Antani et al. [17] presented text localization on compressed images, that resulted in an exceedingly quicker performance. Therefore, their text localizers might even be used for detection of text. The process of text detection stage is related with the text localization and text tracking stage. Google Image labeler [63] is It is feature in the form of game which help to improve the quality of Google's image search results and allowed the user to label to images.

It was online from 2006 to 2011 design simply for fun. It was way for Google to ensure that its keywords were matched to correct images. Each matched word help Google to build an accurate database for Google Image Search.

2.2 Text Localization

Text localization techniques can be further divided into two types based on features utilization: region-based and texture-based. Region-based technique use the properties of the color/gray scale in a text region or their variations with the corresponding properties of the background. These technique can be additional divided into 2 sub-approaches: connected component(CC) and edge-based. These two approaches works in a bottom-up fashion; by distinctive sub-structures, like CCs or edges, and so integrating these structures to mark rectangles for text. CC-based methods are groups the small parts into larger parts till all regions are identified in the image. All text parts using spatial arrangement require geometrical analysis and on separate non-text parts mark the limits of the text regions.

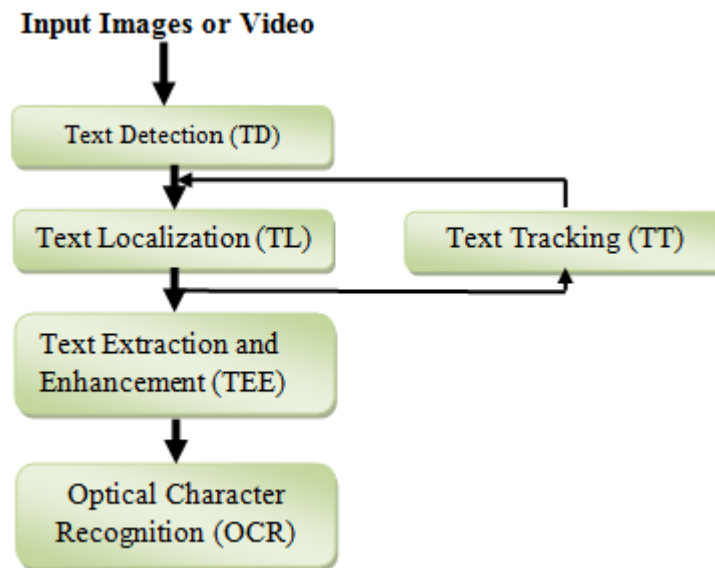


Fig. 4. Text Extraction from images/videos Architecture

Ohya et al. [18] presented a four-stage process: (i) binarization based on local thresholding, (ii) character component detection using gray-level difference, (iii) character recognition and (iv) relaxation operation to update the similarities. Author(s) were able to extract and recognize characters, together with multi-segment characters, under varied illuminating conditions, sizes, positions, and fonts once handling scene text images, like signboard, train, advt., etc.

However, binary segmentation is no appropriate for video documents, which might have many objects with completely different grey levels and high levels of noise and differences in illumination. Moreover, this approach places many restrictions associated with text alignment, like upright and not connected, also because the color of the text (monochrome). Lee and Kankanhalli [19] presented CC-based methodology to detect and recognize of text on cargo containers, Which has uneven lighting conditions and completely different sizes and shapes

characters. Edge information is employed for a coarse search before the CC generation. The boundaries of potential characters are determined by the difference between adjacent pixels in the input image. Threshold values are chosen for every text candidate based on pixels on the boundaries. Despite their claims that the method may be effectively utilized in different domains, experimental results were solely conferred for cargo container images.

Zhong et al. [20] prepared a CC-based methodology, that uses color reduction. Author(s) quantize the color using the peaks in a color histogram (RGB Color Space). It is assumed that the text regions cluster together during this color space and occupy good portion of an image. Every text component goes through a filtering stage employing a variety of heuristics, like spatial alignment, space and diameter. The performance of this technique was evaluated using CD images and book cover images.

Kim [11] presented the approach by segments an image in the RGB space using color clustering in a color histogram. Non-text components. The horizontal text lines associated text segments are extracted based on iterative projection profile analysis. After that these text segments are integrated by heuristics. Since many threshold values ought to be determined by trial and error, this approach isn't appropriate as a general-purpose text localizer. Experiments were performed with fifty video images, together with numerous character sizes and designs, and a localization rate of eighty seven was reported. Lienhart et al. [21,22] regard text regions as CCs with similar size and color, and apply motion analysis to boost the text extraction results. Segments that are too small and overlarge are filtered out. After dilation, For enhance and extracted results motion information and contrast analysis are used. Their main focus is on caption text, that exhibit the higher contrast with the background. Finally, a geometrical analysis, together with the dimension, height, and ratio, is employed to separate any non-text parts.

Shim et al. [8] proposed the homogeneity of intensity of text regions in images. Jain and Yu [23] apply a CC-based methodology once preprocessing, which incorporates bit dropping, color clump, ambiguous image decomposition, and foreground image generation. Once the input frame is decomposed into multiple foreground frames, every foreground image goes through identical text localization stage. By using block adjacency graph, CCs are generated in parallel for all the foreground images. The localized text component within the individual foreground video frame is then integrated into one output frame. The algorithm extracts solely horizontal and vertical text, and not skew text. The authors imply that their algorithm might not work well once the color histogram is sparse.

Messelodi and Modena's technique [24] consists of 3 serial stages: (i) extraction of elementary objects, (ii) filtering of objects, and (iii) text line selections. Preprocessing, like noise reduction, de-blurring, contrast enhancement, quantization, etc., is additionally performed. After the preprocessing, intensity normalization, image binarization, and CC generation are performed. The authors notices that the filter selection is heavily depends on and threshold values. Finally, the text line selection stage starts from one region and recursively expands, till a termination criterion is satisfied. Kim et al. [25] used cluster-based templates for sort out non-character components. The similar approach was additionally reported by Ohya et al. [34].

Hase et al. [26] proposed a CC-based technique for color documents. Author(s) assume that each character is printed in a single color. The image is then divided into many binary images and string extraction by multi-stage relaxation that was given by authors previously [27] is performed on every binary image. The character strings are selected by their

likelihoods, using conflict resolution rules by merging all results from the individual binary images, The inclusion and overlap, the conflicts between character strings are filtered using tree illustration. Contrary to other text localization techniques, Author(s) pay additional attention to the filtering stage. As a result, Author(s) deals with shadowed and curved strings

CC-based methods are widely used due to their relatively simple implementation. Almost all CC-based methods have four processing stages: (i) preprocessing, like color clustering and noise reduction, (ii) CC generation, (iii) filtering out non-text components, and (iv) Component grouping [11]. Further, the performance of a CC-based technique is affected by component grouping, like a projection profile analysis or text line selection. Additionally, many threshold values are required to filter out the non-text components, and these threshold values are depends on the image/video database.

Binarization techniques, that use global, local, or adaptive thresholding, are the simplest methods for text localization. This approach has been adopted for specific applications like address location on snail mail, courtesy amount on checks, etc., thanks to its simplicity in implementation [28].

Due to the variety of possible variations in text in different types of applications and also the limitations of any single approach to wear down such variations. Zhong et al. [20] presented integration of CC-based approach and the texture-based approach. If characters are merged or not well separated from the background, CC-based technique doesn't perform well. Additionally the disadvantage of their texture-based technique is that characters, that extend below the baseline or above the characters, are segmented into 2 Component. In the hybrid scheme the CC-based technique is applied once the bounding boxes are localized using the texture-based technique, and characters extending beyond the bounding boxes are filled in. However, the authors don't offer any quantitative analysis regarding the performance enhancement when using this hybrid approach.

Antani et al. [10, 17] proposed a multi-pronged approach for Text Information Extractions and its all processes. Author(s) used different techniques to reduce the chance of failure. The results of each detection and localization technique may be a set of bounding boxes surrounding the text regions. The results from the individual techniques are then united to provide the final bounding boxes. Author(s) utilize algorithms by Gargi et al. [13,14], Chaddha et al. [29], Schaar-Mitrea and De With [30], and LeBourgeois [31]. Chaddha et al. presented use of texture energy to classify 8×8 pixel blocks into text or non-text blocks. The sum of absolute values of a set of DCT coefficients from MPEG data is thresholded to categorised a block as text or non-text. Schaar-Mitrea and P. De With's formula [30] was originally developed for the classification of graphics and video. Antani et al. [10] modified this to classify 4×4 blocks as text or non-text. The number of pixels of pixels in block with the similar color is compared with threshold values to classify the block.

LeBourgeois [31] used a threshold value based on the sum of the gradients of every pixel's neighbors. Antani's effort to utilize and merge many traditional approaches would appear to provide an inexpensive performance for general image documents. Later, in his thesis, Antani [10] gave additional thought to the merging strategy. Gandhi et al. [32] presented planar motion model for scene text localization. The motion model parameters are then used to compute the structure and motion parameters. Perspective correction is performed using the calculable plane

normal, leading to a far better OCR performance. However, since only motion data is employed for text segmentation, the algorithm requires more processing time. The employment of additional information, like texture or tonal data, might enhance the results.

Niblack's [33] adaptive thresholding method is used to separate non-text regions. Author(s) investigated the relationship between the OCR accuracy and the resolution and found that the best OCR results were obtained when using a factor of 4 in image interpolation. Li and Doermann [34] also used a multiple frame-based text image enhancement technique, wherever consecutive text blocks are correctly tracked registered using a pure translational motion model. The text blocks are correctly tracked, the mean square errors of two consecutive text blocks and motion trail information are used. A sub-pixel accuracy registration is additionally performed bi-linear interpolation to reduce the noise sensitivity of the two low-resolution text blocks. However, this could reduce non-text regions only if text and background have different movements. Lastly, a super-resolution-based text enhancement scheme is additionally given for de-blurring scene text, onto convex sets (POCS)-based method [35].

2.3 Tracking, Extraction and Enhancement

This section discusses tracking, extraction, and enhancement Techniques. Due to many reasons, utility, enhancement speedup, etc., tracking of text in video has not been studied in great extends. Locating text in images, such as low resolution and complex backgrounds, these topics need to be more investigation.

To enhance the system performance, it's necessary to consider temporal changes in a frame sequence. The text tracking stage will serve to verify the text localization results. Additionally, if text tracking may be performed in a shorter time than text detection and localization, this might speed up the general system. In cases wherever text is occluded in different frames, text tracking can be facilitate recover the original image.

Lienhart [21,22] represented a block-matching algorithm, that is an international standard for video compression like H.261 and MPEG, and used temporal text motion info to refine extracted text regions. The minimum mean absolute difference is used as the matching criterion. Each localized block is checked on whether or not its fill factor is higher than a given threshold value. For every block that meets the specified fill factor, a block-matching algorithm is performed.

Antani et al. [17] and Gargi et al. [14] utilize motion vectors in an compressed video MPEC-1 bit stream for tracking the text, based on the strategies of Nakajama et al. [62] and Pilu [63]. This methodology is implemented on the P and that I frames in MPEG-1 video streams. The original bounding box is then moved by the total of the motion vectors of all the macroblocks that correspond to current bounding box. Li et al. [34] presented a approach or many circumstances, including scrolling, captions, text printed on an athlete's jersey, etc. Author(s) used the sum of the square difference for a pure translational motion model, based on multi-resolution matching, to reduce the computational complexity. The text contours are used to stabilize for additional complex motions for tracking process. Edge maps are generated using the canny operator for larger text block. Once a horizontal smearing method to cluster the text blocks, the new text position is extracted. However, since a

translational model is used, this methodology isn't appropriate to handle scale, rotation and perspective variations.

Sawaki et al. [38] projected a techniques for adaptively acquiring templates of degraded characters in scene images involving the automated creation of 'context-based image templates' from text line images. Zhou et al. [39,40,41] use their own OCR algorithm based on surface fitting classifier and n-tuple classifier. we address the issue of text extraction enhancement for generating the input to an OCR algorithm.

Text enhancement techniques are often divided into two categories: single frame-based or multiple frame-based. Several thresholding techniques are developed for still images. However, these strategies don't work well for video sequences. Based on the fact that text typically spans several frames, varied approaches that use a tracking operation in consecutive frames are proposed to enhance text quality. Such enhancement strategies are often effective when the background movement is completely different from the text movement.

Sato et al. [42] used a linear interpolation technique to magnify small text at a better resolution for commercial OCR software package. This methodology is unable to clean the background when both the text and background moving at the same time. After the image enhancement stages, four specialized character-extraction filters are applied based on correlation and a recognition-based segmentation methodology is used for character segmentation.

Li et al. [43,44] given many approaches for text enhancement. Author(s) use the Shannon interpolation technique to enhance the image resolution of video frames. The image resolution is increased an extension of the Nyquist sampling theorem and it's determined whether or not the text is normal or inverse by comparing it with a global threshold and background color.

Antani et al. [10] used number of algorithms for extracting text. Author(s) developed a binarization algorithm similar to Messelodi and Modena's approach [24]. To refine the binarization result, a filtering stage is utilized based on color, size, spacial location, topography, and shape. These extraction algorithms operate on both the original image and its inverse image to detect text regions. Chen et al. [45] proposed a text enhancement methodology that uses a multi-hypotheses approach.

3. TEXT BASED IMAGE VIDEO INDEXING AND RETRIEVAL

The unique properties of video collections (e.g., multiple sources, noisy features and temporal relations) examine the performance of these retrieval methods in such a multimodal environment, and identify the relative importance of the underlying retrieval Components. Li and Doermann [47], presents text-based video indexing and retrieval by expanding the semantics of a query and using the glimpse matching method to perform approximate matching. Cees et al.[49], identify three strategies to select a relevant detector from thesaurus. Jawahar, et. al. [51] presents search approach that based on the textual information present in the video. Textual information regions are identified within the frames of the video. Based on textual content present in the images, video is annotated. OCRs are used to extract the text from the video. Author(s) proposed an approach that provides matching at the image-level and avoided an OCR. Based on query string matching videos are retrieve from database and sort it based on relevance.



Fig 5. Result for Query video Sania in devnagari script (Courtesy Jawahar, C. V, et. al. [51])

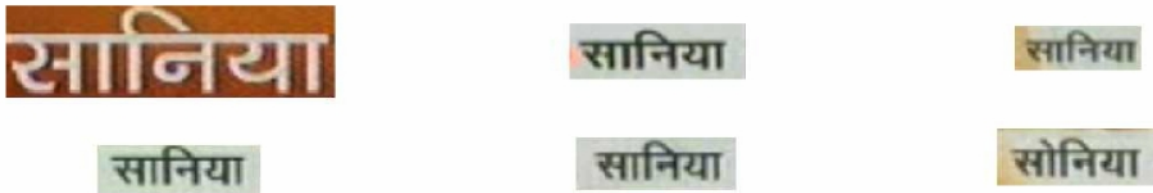


Fig 6. Query word “Sania” in devnagari script from Fig 5. frames (Courtesy Jawahar, C. V, et. al. [51])

4. CONCLUSION

In this paper we are provided the comprehensive literature review of text extraction in images and video as well as text based image and video retrieval. In the literature, a large number of algorithms have been presented but didn't find any techniques which provide satisfactory performance. The different information Sources (e.g., color, texture, motion, shape, geometry, etc). are used for text. By merging the different sources of information we and enhance the performance of a text extraction system and text based video retrieval systems.

REFERENCES

- [1] Jung, Keechul, Kwang In Kim, and Anil K Jain. "Text information extraction in images and video: a survey." Pattern recognition 37, no. 5 (2004): 977-997.
- [2] H. J. Zhang, Y. Gong, S. W. Smoliar, and S. Y. Tan, Automatic Parsing of News Video, Proc. of IEEE Conference on Multimedia Computing and Systems, 1994, pp. 45-54.

- [3] Y. Cui and Q. Huang, Character Extraction of License Plates from Video, Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 1997, pp. 502–507.
- [4] A. K. Jain, and Y. Zhong, Page Segmentation using Texture Analysis, Pattern Recognition, 29 (5) (1996) 743-770.
- [5] Y. Y. Tang, S. W. Lee, and C. Y. Suen, Automatic Document Processing: A Survey, Pattern Recognition, 29 (12) (1996) 1931-1952.
- [6] B. Yu, A. K. Jain, and M. Mohiuddin, Address Block Location on Complex Mail Pieces, Proc. of International Conference on Document Analysis and Recognition, 1997, pp. 897-901.
- [7] D. S. Kim and S. I. Chien, Automatic Car License Plate Extraction using Modified Generalized Symmetry Transform and Image Warping, Proc. of International Symposium on Industrial Electronics, 2001, Vol. 3, pp. 2022-2027.
- [8] J. C. Shim, C. Dorai, and R. Bolle, Automatic Text Extraction from Video for Content-based Annotation and Retrieval, Proc. of International Conference on Pattern Recognition, Vol. 1, 1998, pp. 618-620.
- [9] S. Antani, D. Crandall, A. Narasimhamurthy, V. Y. Mariano, and R. Kasturi, Evaluation of Methods for Detection and Localization of Text in Video, Proc. of the IAPR workshop on Document Analysis Systems, Rio de Janeiro, December 2000, pp. 506-514.
- [10] S. Antani, Reliable Extraction of Text From Video, PhD thesis, Pennsylvania State University, August 2001
- [11] H. K. Kim, Efficient Automatic Text Location Method and Content-Based Indexing and Structuring of Video Database, Journal of Visual Communication and Image Representation 7 (4) (1996) 336-344.
- [12] M.A. Smith and T. Kanade, Video Skimming for Quick Browsing Based on Audio and Image Characterization, Technical Report CMU-CS-95-186, Carnegie Mellon University, July 1995
- [13] U. Gargi, S. Antani, and R. Kasturi, Indexing Text Events in Digital Video Database, Proc. of International Conference on Pattern Recognition, 1998, vol. 1. pp. 1481-1483.
- [14] U. Gargi, D. Crandall, S. Antani, T. Gandhi, R. Keener, and R. Kasturi, A System for Automatic Text Detection in Video, Proc. of International Conference on Document Analysis and Recognition, 1999, pp. 29–32.
- [15] Y. K. Lim, S. H. Choi, and S.W. Lee, Text Extraction in MPEG Compressed Video for Content-based Indexing, Proc. of International Conference on Pattern Recognition, 2000 pp. 409-412.
- [16] Yu Zhong, Hongjiang Zhang, and Anil K. Jain, Automatic Caption Localization in Compressed Video, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, (4) (2000) 385-392.
- [17] S. Antani, U. Gargi, D. Crandall, T. Gandhi, and R. Kasturi, Extraction of Text in Video, Technical Report of Department of Computer Science and Engineering, Penn. State University, CSE-99-016, August 30, 1999
- [18] J. Ohya, A. Shio, and S. Akamatsu, Recognizing Characters in Scene Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16 (2) (1994) 214-224.
- [19] C.M. Lee, and A. Kankanhalli, Automatic Extraction of Characters in Complex Images, International Journal of Pattern Recognition Artificial Intelligence, 9 (1) (1995) 67-82.
- [20] Yu Zhong, Kalle Karu, and Anil K. Jain, Locating Text In Complex Color Images, Pattern Recognition, 28 (10) (1995) 1523-1535.
- [21] R. Lienhart and F. Stuber, Automatic Text Recognition In Digital Videos, Proc. of SPIE, 1996, pp. 180-188.
- [22] R. Lienhart and W. Effelsberg, Automatic Text Segmentation and Text Recognition for Video Indexing, Technical Report TR-98-009, Praktische Informatik IV, University of Mannheim, 1998
- [23] A. K. Jain, and B. Yu, Automatic Text Location in Images and Video Frames, Pattern Recognition, 31 (12) (1998) 2055-2076.
- [24] S. Messelodi and C.M. Modena, Automatic Identification and Skew Estimation of Text Lines in Real Scene Images, Pattern Recognition, 32 (1992) 791-810.
- [25] E. Y. Kim, K. Jung, K. Y. Jeong, and H. J. Kim, Automatic Text Region Extraction Using Cluster-based Templates, Proc. of International Conference on Advances in Pattern Recognition and Digital Techniques, 2000, pp. 418-421

- [26] H. Hase, T. Shinokawa, M. Yoneda, and C. Y. Suen, Character String Extraction from Color Documents, *Pattern Recognition*, 34 (7) (2001) 1349-1365.
- [27] H. Hase, T. Shinokawa, M. Yoneda, M. Sakai, and H. Maruyama, Character String Extraction by Multi-stage Relaxation, *Proc. of ICDAR'97*, 1997, pp. 298-302.
- [28] D. Chen, J. Luetin, and K. Shearer, A Survey of Text Detection and Recognition in Images and Videos, Institut Dalle Molle Intelligence Artificielle Perceptive (IDIAP) Research Report, IDIAP-RR 00-38, August 2000.
- [29] N. Chaddha, R. Sharma, A. Agrawal, and A. Gupta, Text Segmentation in Mixed-mode Images, *Proc. of Asilomar Conference on Signals, Systems and Computers*, 1994, pp. 1356-1361.
- [30] M.v.d.Schaar-Mitrea and P. de With, Compression of Mixed Video and Graphics Images for TV Systems, *Proc. of SPIE Visual Communication and Image Processing*, 1998, pp. 213-221.
- [31] F. LeBourgeois, Robust Multifont OCR System from Gray Level Images, *Proc. of International Conference on Document Analysis and Recognition*, 1997, Vol. 1, pp. 1-5.
- [32] T. Gandhi, R. Kasuturi and S. Antani, Application of Planar Motion Segmentation for Scene Text Extraction, *Proc. of International Conference on Pattern Recognition*, 2000, Vol. 1, pp. 445-449.
- [33] YL. Niblack, *An Introduction to Image Processing*, Englewood Cliffs, N. J.:Prentice Hall, 1986.
- [34] H. Li, O. Kia, and D. Doermann, Text Enhancement in Digital Video, *Proc. of SPIE, Document Recognition IV*, 1999, pp. 1-8.
- [35] H. Li and D. Doermann, Superresolution-based Enhancement of Text in Digital Video, *Proc. of International Conference of Pattern Recognition*, 2000, Vol. 1, pp. 847-850.
- [36] Y. Nakajima, A. Yoneyama, H. Yanagihara, and M. Sugano, Moving Object Detection from MPEG Coded Data, *Proc. of SPIE*, 1998, Vol. 3309, pp.988-996.
- [37] M. Pilu, On Using Raw MPEG Motion Vectors to Determine Global Camera Motion, *Proc. of SPIE*, 1998, Vol. 3309, pp. 448-459.
- [38] M. Sawaki, H. Murase, and N. Hagita, Automatic Acquisition of Context-based Image Templates for Degraded Character Recognition in Scene Images, *Proc. of International Conference on Pattern Recognition*, 2000, Vol. 4, pp. 15-18.
- [39] J. Zhou and D. Lopresti, Extracting text from WWW Images, *Proc. of International Conference on Document Analysis and Recognition*, 1997, Vol. 1, pp. 248 –252.
- [40] J. Zhou, D. Lopresti, and Z. Lei, OCR for World Wide Web Images, *Proc. of SPIE on Document Recognition IV*, Vol. 3027, 1997, pp. 58-66.
- [41] J. Zhou, D. Lopresti, and T. Tasdizen, Finding Text in Color Images, *Proc. of SPIE on Document Recognition V*, 1998, pp. 130-140.
- [42] T. Sato, T. Kanade, E. K. Hughes, and M. A. Smith, Video OCR for Digital News Archive, *Proc. of IEEE Workshop on Content based Access of Image and Video Databases*, 1998, pp. 52-60.
- [43] H. Li, D. Doerman, and O. Kia, Automatic Text Detection and Tracking in Digital Video, *IEEE Transactions on Image Processing*, 9 (1) January (2000) 147-156.
- [44] H. Li, D. Doerman, and O. Kia, Automatic Text Detection and Tracking in Digital Video, *IEEE Transactions on Image Processing*, 9 (1) January (2000) 147-156.
- [45] D. Chen, J. Odobez, and H. Bourlard, Text Segmentation and Recognition in Complex Background Based on Markov Random Field, *Proc. of International Conference on Pattern Recognition*, 2002, Vol. 4, pp. 227-230.
- [46] D. Chen, J. Odobez, and H. Bourlard, Text Segmentation and Recognition in Complex Background Based on Markov Random Field, *Proc. of International Conference on Pattern Recognition*, 2002, Vol. 4, pp. 227-230.
- [47] Li H.P., and Doermann D., 2002, Video indexing and retrieval based on recognized text, in *Proceedings IEEE Workshop Multimedia Signal Process*, pp.245–248.
- [48] Caroline Lacoste, Joo-Hwee Lim, Jean-Pierre Chevallet, and Diem Thi Hoang Le, 2007, Medical-Image Retrieval Based on Knowledge-Assisted Text and Image Indexing, *IEEE Transactions On Circuits And Systems For Video Technology*, Vol.17, No.7, pp.889-900.
- [49] Cees G. M. Snoek, Bouke Huurnink, Laura Hollink, Maarten De Rijke, Guus Schreiber and Marcel Worring, 2007, Adding Semantics To Detectors For Video Retrieval, *IEEE Transactions on Multimedia*, Vol.9, No.5, pp.975-986.

- [50] Rong Yan, and Alexander G. Hauptmann, 2007, A review of text and image retrieval approaches for broadcast news video, *Computer Science Information Retrieval*, Vol.10, No.4-5, pp.445-484.
- [51] Jawahar, C. V., BalaKrishna Chennupati, Balamanohar Paluri, and Nataraj Jammalamadaka. "Video retrieval based on textual queries." In *Proceedings of the Thirteenth International Conference on Advanced Computing and Communications*, Coimbatore. 2005.
- [52] A.N. Bhute, B.B. Meshram "System Analysis and Design for Multimedia Indexing and Retrieval" *IJMA International journal of Multimedia and its Applications*, Vol 5, No.6, Dec 2013, ISSN 0975-5934.
- [53] A.N. Bhute, B.B. Meshram, "Multimedia Indexing and Retrieval" , *International Journal on International Journal Of Research In Computer Application & Management(IJRIME)*, Vol 3, Issue 4, DEC 2013; ISSN 2231-1009 PP. 16-55.
- [54] A.N. Bhute, B.B. Meshram "Content Based Image Indexing and Retrieval", *International Journal of Graphics and Image Processing*, Vol. 3, issue 4, Nov. 2013, ISSN 2249-5452, pp. 235-246
- [55] A.N. Bhute, B.B. Meshram "Automated Multimedia Information Retrieval Using Color and Texture Feature Techniques", *International Journal of Electronics Communication and Computer Engineering(IJECCE)*, Volume 3, Issue 5,Oct 2013,ISSN (Online): 2249-071X, ISSN (Print): 2278-4209, 2012, pp. 885-891.
- [56] A.N. Bhute, B.B. Meshram "Multimedia Indexing and Retrieval Techniques: A Review." *International Journal of Computer Applications IJCA*,Vol. 58, no. 3 November 2012, ISSN 0975-8887, pp. 35-43
- [57] [14]A.N. Bhute, B.B. Meshram "IVSS: Integration of Color Feature Extraction Techniques for Intelligent Video Search Systems" in proceeding of 4th International Conference on Electronics Computer Technology (ICECT 2012),April 6-8 2012, kanyakumari, India.
- [58] A.N. Bhute, B.B. Meshram "Techniques for Text, Image, Audio and Video Indexing and Retrieval" *International Conference on Network Infrastructure Management Systems Interface -2011*, VJTI, Matunga, Mumbai,16-17 Dec 2011, pp. 90-96
- [59] A.N. Bhute, B.B. Meshram "Intelligent Web Agent for Search Engines" in the proceeding of *International Conference on Trends and Advances in Computational Engineering (TRACE -2010)* Barkatullah University Institute Of Technology, Barkatullah University, Bhopal, India, Feb, 25th – 26th 2010. pp. 211-218
- [60] A.N. Bhute, B.B. Meshram "Decision Tree: Predicting Recurrence Pattern" in the proceeding of 4th *International Conference on Intelligent Systems & Networks (IISN-2010)* Computational Intelligence Laboratory (Ci-Lab), IST Klawad Yamuna Nagar, Haryana, Feb, 26th -28th 2010 pp. 215-219.
- [61] A.N. Bhute, B.B. Meshram "Performance Analysis of Ranking Algorithms" in the proceeding of *International conference on Green IT (GIT-2010)*, Sinhgad college of Management, Pune, India, Feb, 22nd -25th , 2010 pp. 197-202.
- [62] A.N. Bhute, B.B. Meshram "Novel Approaches for Performing Clustering and Classification Tasks Using Graphs Similarity Techniques: A Review" in the proceeding of *International Conference On Computing, Communication and Information Technology Applications (CCITA-2010)* at VLBJCOET, Coimbatore, India Jan 21st -23rd 2010,pp. 113-119, ISBN. 978-81-909789-0-3
- [63] Alan Eustace (September 2, 2011). "A fall spring-clean". *Official Google Blog*. Retrieved September 2, 2011.