Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TECHNIQUES AND CHALLENGES OF AUTOMATIC TEXT EXTRACTION IN COMPLEX IMAGES: A SURVEY

[1] C.P. SUMATHI, [2] T. SANTHANAM, [3] N. PRIYA

[1] Department of Computer Science, SDNB Vaishnav College For Women, Chennai, India
[2] Department of Computer Application, DG Vaishnav College For Men, Chennai, India
[3] Department of Computer Science, SDNB Vaishnav College For Women, Chennai, India

E-mail: [1] santsum@hotmail.com, [2] santhanam_dgvc@yahoo.com, [3] cmtsahaana@gmail.com

ABSTRACT

Text extraction in images and video has been developing rapidly since 1990s and is an important research field in content-based information indexing and retrieval, automatic annotation and structuring of images. Extraction of this information involves detection, localization, tracking, extraction, enhancement, and recognition of the text from a given image. However, variations of text due to differences in size, style, orientation, and alignment, as well as low image contrast and complex background make the problem of automatic text extraction extremely difficult and challenging job. A large number of techniques have been proposed to address this problem and the purpose of this paper is to classify and review these techniques, discuss the applications and performance evaluation, and to identify promising directions for future research.

Keywords : *Text Extraction, Text Detection, Text Localization, Text Tracking, Text Enhancement*

1. INTRODUCTION

Extracting text from images or videos is an important problem in many applications like document processing [1,2], image indexing, video content summary [3–5], video retrieval [6], video understanding [7], and since including. Usually, texts embedded in an image or a frame capture important media contexts such as player's name, title, date, story introduction, and since including. Therefore, the task can provide various advantages for annotating an image or a video and thus improves the accuracy of a content-based indexing system to search desired media contents. In addition, the information can be used for content filtering so that commercial programs can be found and removed out for video summary. Moreover, when analyzing video audios, the recognition result of text line can provide extra refinements for correcting the errors of speech recognition. Since 1990s, with rapid growth of available multimedia documents and increasing demand for information indexing and retrieval, much effort has been done on text extraction in images and videos. A larger number of approaches, such as region based, edge based, morphological based and texture based methods, have been proposed and already obtained impressive performance.

1.1 Text In Images

Content-based image indexing refers to the process of attaching labels to images based on their content. Image content can be divided into two main categories: *perceptual content* and *semantic content* [8]. Perceptual content includes attributes such as color, intensity, shape, texture, and their temporal changes, whereas semantic content means objects, events, and their relations. A number of studies on the use of relatively low-level perceptual content [9-13] for image and video indexing have already been reported. Studies on semantic image content in the form of text, face, vehicle, and human action have also attracted some recent interest [14-23].

Figures 1-3 show some examples of text in images. Text in images can be classified into caption text (Figure 2), which is artificially overlaid on the image, or scene text (Figure 3), which exists naturally in the image. Some researchers like to use the term '*graphics text*' for scene text, and '*superimposed text'* or '*artificial text*' for caption

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

text [24, 25]. It is well known that scene text is more difficult to detect and very little work has been done in this area. In contrast to caption text, scene text can have any orientation and may be distorted by the perspective projection. Moreover, it is often affected by variations in scene and camera parameters such as illumination, focus, motion, etc. Page layout analysis usually deals with document images (Figure 1). Readers may refer to papers on document segmentation/analysis [26, 27] for more examples of document images. Although images acquired by scanning book covers, CD covers, or other multi-colored documents have similar characteristics as the document images (Fig. 2)

1.2. Properties Of Text In Images

Texts usually have different appearance changes like font, size, style, orientation, alignment, texture, color, contrast, and background [28]. All the changes will make the problem of automatic text extraction become complicated and difficult. There are many researchers who have devoted themselves

Figure 1: Document Images

Figure 2: Caption Text Images

Figure 3: Scene Text Images

to investigating different methods for tackling the above problems[28-34]. Text in images can exhibit many variations with respect to the following properties:

Geometry:

. Size: Although the text size can vary a lot, assumptions can be made depending on the application domain.

Alignment: The characters in the caption text appear in clusters and usually lie horizontally, although sometimes they can appear as non-planar texts as a result of special effects. This does not apply to scene text, which can have various perspective distortions. Scene text can be aligned in any direction and can have geometric distortions (Fig. 4).

Inter-character distance: characters in a text line have a uniform distance between them.

Color: The characters in a text line tend to have the same or similar colors. This property makes it possible to use a connected component-based approach for text detection. Most of the research reported till date has concentrated on finding 'text strings of a single color (monochrome)'. However, video images and other complex color documents can contain 'text strings with more than two colors (polychrome)' for effective visualization, i.e., different colors within one word.

Motion: The same characters usually exist in consecutive frames in a video with or without movement. This property is used in text

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

tracking and enhancement. Caption text usually moves in a uniform way: horizontally or vertically. Scene text can have arbitrary motion due to camera or object movement.

Edge: Most caption and scene text are designed to be easily read, thereby resulting in strong edges at the boundaries of text and background.

Compression: Many digital images are recorded, transferred, and processed in a compressed format. Thus, a faster TIE system can be achieved if one can extract text without decompression.

1.3. Text Information Extraction (Tie)

A TIE system receives an input in the form of a still image or a sequence of images. The images can be in gray scale or color, compressed or un-compressed, and the text in the images may or may not move. The TIE problem can be divided into the following sub-problems: (i) detection, (ii) localization, (iii) tracking, (iv) extraction and enhancement, and (v) recognition (OCR)

Text detection refers to the determination of the presence of text in a given frame (normally text detection is used for a sequence of images).

Text localization is the process of determining the location of text in the image and generating bounding boxes around the text.

Text tracking is performed to reduce the processing time for text localization and to maintain the integrity of position across adjacent frames. Although the precise location of text in an image can be indicated by bounding boxes, the text still needs to be segmented from the background to facilitate its recognition. This means that the extracted text image has to be converted to a binary image and enhanced before it is fed into an OCR engine.

Text extraction is the stage where the text components are segmented from the background.

Text Enhancement of the extracted text components is required because the text region usually has low-resolution and is prone to noise. Thereafter, the extracted text images can be transformed into plain text using OCR technology.

1.4. Scope And Discussion

With rapid developments in text extraction techniques, several papers have been reported in the literature. Chen *et al.* [35] , Jung *et al.* [36] and Jing *et al.* [37] have presented comprehensive surveys of the text extraction approaches for images and videos in the recent years. This paper concentrate on the techniques and challenges proposed for text extraction in complex images and to summarize and discuss the recent progress in this research area.

The rest of the paper is organized as follows: Recent techniques for text extraction are reviewed in Section 2. Section 3 describes performance evaluation of text extraction methods. Section 4 includes a few new applications of text extraction techniques. Section 5 presents summary and conclusions.

2. TEXT EXTRACTION TECHNIQUES

Text extraction in images includes five stages, among which text detection and text localization are closely related and more challenging stages which attract the attention of most researchers. The goal of the two stages is to generate accurate bounding boxes of all text objects in images and video frames and provide a unique identity to each text. In this section, the recent techniques focused on text detection and localization are reviewed and then the results are discussed.

2.1 Region -Based Technique

Region-based methods use the properties of the color or gray-scale in a text region or their differences with the corresponding properties of the background. This method uses a bottom-up approach by grouping small components into successively larger components until all regions are identified in the image. A geometrical analysis is needed to merge the text components using the spatial arrangement of the components so as to filter out non-text components and mark the boundaries of the text regions.

Leon [37] presented a method for caption text detection. It included in a generic indexing system dealing with other semantic concepts which are to be automatically detected. To have a coherent detection system, the various object detection algorithms use a common image description. The author proposed the image description is a hierarchical region-based image model and

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

introduced the algorithm for text detection. This algorithm is divided into three phases:

1. *Text candidate spotting*: an attempt to separate text from background is done.
2. *Text characteristics verification*: where text candidate regions are grouped to discard those regions wrongly selected.
3. *Consistency analysis for output:* where regions representing text are modified to obtain a more useful character representation as input for an OCR.

This technique takes advantage of texture and geometric features to detect the caption text. Texture features are estimated using wavelet analysis and mainly applied for Text candidate spotting. In turn, Text characteristics verification is basically carried out relying on geometric features, which are estimated exploiting the region-based image model. Analysis of the region hierarchy provides the final caption text objects. The final step of Consistency analysis for output is performed by a binarization algorithm that robustly estimates the thresholds on the caption text area of support. Results are shown by step wise in the Figure 4.

Karin[38] approached a method for automatic text location and identification on colored book and journal cover. To reduce the amount of small variations in color, a clustering algorithm is applied in a preprocessing step. Two methods have been developed for extracting text hypotheses. One is based on a top-down analysis using successive splitting of image regions. The other is a bottom-up region growing algorithm. The results of both methods are combined to robustly distinguish between text and non-text elements. Text elements are binarized using automatically extracted information about text color. The binarized text regions can be used as input for a conventional OCR module. The proposed method is not restricted to cover pages, but can be applied to the extraction of text from other types of color images as well.

Debapratim [[39] described the bottom-up approach of Line Segmentation from handwritten text. In this method, first the picture is divided of into small squares of height 10 pixels and width 10 pixels. If 50% of the square box is filled up with black pixels then the total square is filled with black pixels. In this way graphically smooth image is found. Then, the height of the smoothed components is found. Depending on the height and the position information these smoothed blocks are joined to get the individual lines.

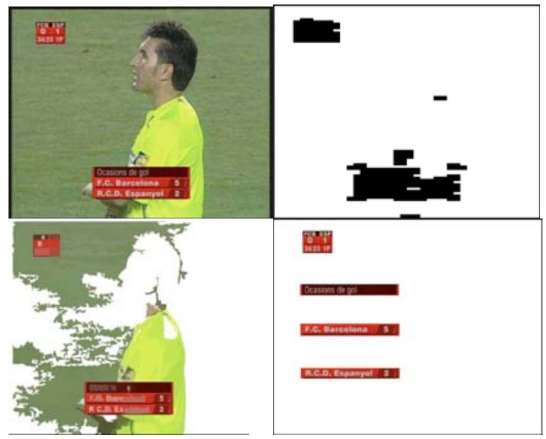

Figure 4: Stepwise Results of Text detection (Leon [37])

2.2. Edge Based Technique

Edges are a reliable feature of text regardless of color/intensity, layout, orientations, etc. Edge strength, density and the orientation variance are three distinguishing characteristics of text embedded in images, which can be used as main features for detecting text. Edge-based text extraction algorithm is a general-purpose method, which can quickly and effectively localize and extract the text from both document and indoor/ outdoor images.

Xiaoqing Liu[40] method consists of three stages: candidate text region detection, text region localization and character extraction. In the first stage, the magnitude of the second derivative of intensity as a measurement of edge strength is used, as this allows better detection of intensity peaks that normally characterize text in images. The edge density is calculated based on the average edge strength within a window. Considering effectiveness and efficiency, four orientations (0^o, 45^o, 90^o, 135^o) are used to evaluate the variance of orientations, where 0^o denotes horizontal direction, 90^o denotes vertical direction, and 45^o and 135^o are the two diagonal directions. Edge detector is carried out by using a multiscale strategy, where the multiscale images are produced by Gaussian pyramids after successively applying low-pass filter and down-sample the original image reducing the image in both vertical and horizontal directions. In the second stage, characteristics of clustering can be used to localize text regions. In the third stage, the

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

existing OCR engine where used. This can only deal with printed characters against clean backgrounds and cannot handle characters embedded in shaded, textured or complex backgrounds. Results are shown in the Figure 5.

(a) Original Image (b) Extracted Text

Figure 5: Object label image with different font sizes, colors and orientational alignments (a) Original images (b) Extracted text (Xiaoqing Liu[40])

Xin Zhang [41] used the color and edge features to extract the text from the video frame. In this work, two methods are combined, called color-edge combined algorithm, to remove text background. One of the combined methods is based on the exponential changes of text color, called Transition Map model, the other one uses the text edges of different gray level image. After removing complex background, text location is determined using the vertical and horizontal projection method. This algorithm is robust to the image with multilingual text. To improve the efficiency of this method, the edge feature is added to remove background and then edge detection is performed on each color image using Canny operator and some Morphology operation. Finally the background of text is removed with the help of Transition Map model.

2.3.Morphological Based Technique

Mathematical morphology is a topological and geometrical based approach for image analysis. It provides powerful tools for extracting geometrical structures and representing shapes in many applications. Morphological feature extraction techniques have been efficiently applied to character recognition and document analysis. It is used to extract important text contrast features from the processed images. The feature is invariant against various geometrical image changes like translation, rotation, and scaling. Even after the lighting condition or text color is changed, the feature still can be maintained. This method works robustly under different image alterations.

Jui-Chen Wu [42] presented a morphology-based text line extraction algorithm for extracting text regions from cluttered images. First of all, the method defines a novel set of morphological operations for extracting important contrast regions as possible text line candidates. In order to detect skewed text lines, a moment-based method is then used for estimating their orientation. According to the orientation, an x-projection technique can be applied to extract various text geometries from the text-analogue segments for text verification. However, due to noise, a text line region is often fragmented into different pieces of segments. Therefore, after the projection, a novel recovery algorithm is then proposed for recovering a complete text line from its pieces of segments. After that, a verification scheme is then proposed for verifying all extracted potential text lines according to their text geometries. In order to analyze the performance of this approach, an image database including 100 images was used for testing. After testing this method, these images have various appearance changes like contrast changes, complex backgrounds, lightings, different fonts, and sizes. Figure 6 shows the results of text line detection in different images with different alterations.

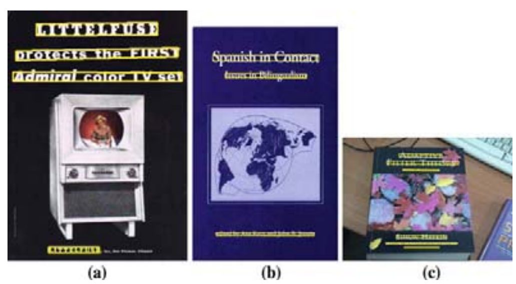

Figure 8: Results of text line detection

Rama Mohan[43] performed the edge detection operation using the basic operators of mathematical morphology. Using the edges the algorithm has tried to find out text candidate connected components. These components have been labeled to identify different components of the image. Once the components have been identified, the variance is found for each connected component considering the gray levels of those components. Then the text is extracted by selecting those connected components whose variance is less than some threshold value.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.4. Texture-Based Technique

Texture-based methods use the observation that text in images have distinct textural properties that distinguish them from the background. The techniques based on Gabor filters, Wavelet, FFT, spatial variance, etc. can be used to detect the textural properties of a text region in an image.

Chu Duc[44] presented a novel texture descriptor based on line-segment features for text detection in images and video sequences, which is applied to build a robust car license plate localization system. Unlike most of the existing approaches which use low level features (color, edge) for text / non-text discrimination, the aim is to exploit more accurate perceptual information. A – scale and rotation invariant - texture descriptor which describes the directionality, regularity, similarity, alignment and connectivity of group of segments are proposed. An improved algorithm for feature extraction based on local connective Hough transform has also been investigated.

Kwang [45] used a novel texture-based method for detecting texts in images. A support vector machine (SVM) is used to analyze the textural properties of texts. No external texture feature extraction module is used; rather, the intensities of the raw pixels that make up the textural pattern are fed directly to the SVM, which works well even in high-dimensional spaces. Next, text regions are identified by applying a continuously adaptive mean shift algorithm (CAMSHIFT) to the results of the texture analysis. The combination of CAMSHIFT and SVMs produces both robust and efficient text detection, as time-consuming texture analyses for less relevant pixels are restricted, leaving only a small part of the input image to be texture-analyzed. The performance criterion was the classification accuracy of the SVMs for text and nontext patterns rather than the overall text detection results. For this purpose, 100 training images were divided into two different classes of 70 training images and 30 validation images from which training patterns and validation patterns were collected, respectively. The SVMs were then trained using the training patterns and tested using the validation patterns.

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.5. Listing The Survey Of Text Extraction Using Different Approaches

A chronological listing of some of the published work on different approaches used for text extraction is presented in Table 1.

Technique	Author	year	Title	Method
Region-Based	Leon [37]	2009	Caption text extraction for indexing purposes using a hierarchical region-based image model	hierarchical region-based
	Debapratim[39]	2009	A Bottom-Up Approach of Line Segmentation from HandwrittenText	Bottom-Up Approach of Line Segmentation
	Karin[38]	2002	Identification of Text on Colored Book and Journal Covers	Clustering, Top-down successive splitting, bottom-up region growing
Edge Based	Xin Zhang [41]	2010	A Combined Algorithm for Video Text Extraction	Transition Map, Canny Operator
	Chowdhury[47]	2009	Robust Extraction of Text from Camera Images	A novel methodology is proposed to determine the discrete edges around the text boundaries
	Xiaoqing Liu[40]	2006	Multiscale Edge-Based Text Extraction From Complex Images	Multiscale Strategy, Clustering
Morphological Based	Rama Mohan[43]	2010	Text Extraction From Hetrogenous Images Using Mathematical Morphology	Thresholding using morphological operators
	Jui-Chen Wu[42]	2008	Morphology-based text line extraction	A novel set of morphological operations & an *x*-projection technique
	Yuming[48]	2008	Text String Extraction from Scene Image Based on Edge Feature and Morphology	Mathematical Morphology and Edge border ratio

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Texture Based	Chu Duc[44]	2009	Robust Car License Plate Localization using a Novel Texture Descriptor	Hough transform
	Bassem[46]	2006	A New Approach For Texture Features Extraction: Application For Text Localization In Video Images”	Hough Transform technique combined with an extremity segment’s neighborhood analysis
	Kwang [45]	2003	Texture-Based Approach for Text Detection in Images Using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm	SVM and CAMSHIFT

Table 1: Sample survey of Text Extraction of using different approaches

3. PERFORMANCE EVALUATION

Performance evaluation of text localization and detection using various approaches in TIE and highlights of several issues in these evaluation methods have been summarized by Jui-Chen Wu[42]

Recall Rate is the ratio of the number $Num_{Correct}$ of correct text detected by the algorithm to the total number Num_{actual} of actual text appearing in the test images, i.e.,

$$Recall = Num_{Correct} / Num_{actual}.$$

Precision Rate is the ratio of the number of text correctly detected by the algorithm to the total number
$Num_{Detected}$ of detected text ; that is,

$$Precision = Num_{Correct} / Num_{Detected}.$$

Several approaches were used for TIE, but in this paper recently presented ones, their comparison methodologies and results for selected four algorithms are listed in Table 2. Each algorithm was slightly modified for the sake of comparison and the ground truth data was generated, frame-by-frame, manually to include text box size, position, and orientation angle. The evaluation was performed pixel-by-pixel. After classifying all the pixels as *Correct Detection*, *False Alarm*, or *Missed Detection*, the recall and precision rates were calculated for all the algorithms, along with approximate processing times.

Technique	Recall rate (%)	Precision rate (%)
Leon [37]	86.3	86.5
Debapratim[39]	92(correctly detected)	
Xin Zhang [41]	90.74	49
Xiaoqing Liu[40]	96.6	91.8
Jui-Chen Wu[42]	95.3	99.4
Rama Mohan[43]	89	77.9
Chu Duc[44]	96.7	1.7
Bassem[46]	96	2.48

Table 2: Comparison Results

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. APPLICATIONS OF TEXT INFORMATION EXTRACTION

Text extraction from images finds many useful applications in document analysis, vehicle licence plate extraction, content based image retrieval, text- based image indexing, video content analysis, industrial automation etc and many applications have become realities in recent years [38]. Educational and training video and TV programs such as news contain mixed text-picture-graphics regions. Region classification is helpful in object-based compression, manipulation and accessibility. Also, text regions may carry useful information about the visual content.

However, as mentioned earlier, due to the variety of fonts, sizes, styles, orientations, alignment effects of uncontrolled illuminations, reflections, shadows, the distortion due to perspective projection as well as the complexity of image background, automatic localizing and extracting text is a challenging problem.

5. SUMMARY AND CONCLUSION

Text extraction in images, as an important research branch of content-based information retrieval and text-based image indexing, continues to be a topic of much interest to researchers. A large number of newly proposed approaches in the literature have contributed to an impressive progress of text extraction techniques Although many researchers have already investigated text localization, text detection and tracking for images is required for utilization in real applications (e.g., mobile handheld devices with a camera and real-time indexing systems). A text-image-analysis, is needed to enable a text information extraction system to be used for any type of image, including both scanned document images and real scene images through a video camera. Despite the many difficulties in using TIE systems in real world applications, the importance and usefulness of this field continues to attract much attention.

REFERENCES

[1] Jain, K., and Yu, B., "Document representation and its application to page decomposition," *IEEE Trans. Pattern Anal. Machine Intel*l., vol. 20, pp. 294–308, Mar. 1998.

[2] Jain, K., and Yu, B., "Automatic Text Location in Images and Video Frames", *Pattern Recognition*, 31 (12) 2055-2076, 1998.

[3] Jain, K., and Zhong, Y., "Page Segmentation using Texture Analysis", *Pattern Recognition*, 29 (5) 743-770, 1996.

[4] Chitrakala Gopalan and D. Manjula, "Contourlet Based Approach for Text Identification and Extraction from Heterogeneous Textual Images", *International Journal of Computer Science and Engineering* 2(4) pp.202-211,2008.

[5] Crandall, D., Antani, S., and Kasturi, R. "Robust Detection of Stylized Text Events in Digital Video", *Proceedings of International Conference on Document Analysis and Recognition*, pp. 865-869,2001.

[6] Serra, J.," Image Analysis and Mathematical Morphology" *New York:Academic*, 1982.

[7] Fan, K. C., Wang, L. S., and Wang, Y. K., "Page segmentation and identification for intelligent signal Processing," *Signal Process*., vol. 45, pp.329–346, 1995.

[8] Kim, H. K., "Efficient Automatic Text Location Method and Content-Based Indexing and Structuring of Video Database", *Journal of Visual Communication and Image Representation* 7 (4) 336-344,1996.

[9] Yu Zhong and Anil K. Jain, "Object Localization using Color, Texture, and Shape", *Pattern Recognition 33* 671-684,2000.

[10] Antani, S., Kasturi, R., and Jain, R., "A Survey on the Use of Pattern Recognition Methods for Abstraction, Indexing, and Retrieval of Images and Video", *Pattern Recognition* 35 945-965,2002.

[11] Flickner, M., Sawney H., "Query by Image and Video Content: The QBIC System", *IEEE Computer* 28 (9) 23-32,1995 .

[12] Zhang, H. J., Gong,Y., Smoliar, S. W., and Tan, S. Y., "Automatic Parsing of News Video", *Proc. of IEEE Conference on Multimedia Computing and Systems,* pp. 45-54,1994.

[13] Arnold W.M. Smeulders, Simone Santini, Amarnath Gupta, and Ramesh Jain, "Content-Based Image Retrieval at the End of the Early Years", *IEEE*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Transactions on Pattern Analysis and Machine Intelligence, 22 (12) 1349-1380,2000.

[14] Smith, M.A., and Kanade, T., "Video Skimming for Quick Browsing Based on Audio and Image Characterization", *Technical Report CMU-CS-95-186, Carnegie Mellon University*, 1995.

[15] Yang, M. H., Kriegman, D. J., and Ahuja, N., "Detecting faces in Images: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (1) 34-58,2002.

[16] Cui, N., and Huang, N., "Character Extraction of License Plates from Video", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 502 – 507,1997.

[17] Colombo, C., Bimbo, A. D., and Pala, P., "Semantics in Visual Information Retrieval", *IEEE Multimedia*, 6 (3) 38-53,1999.

[18] Sato, T., Kanade, T., Hughes, E. K., and Smith, M. A., "Video OCR for Digital News Archive", *Proc. of IEEE Workshop on Content based Access of Image and Video Databases*, pp. 52-60,1998.

[19] Atsuo Yoshitaka and Tadao Ichikawa, "A Survey on Content-based Retrieval for Multimedia Databases", *IEEE Transactions on Knowledge and Data Engineering*, 11 (1) 81-93,1999.

[20] Qi, W., Gu, L., Jiang, H., Chen, X., and Zhang, H., "Integrating Visual, Audio, and Text Analysis for News Video", *Proc. of IEEE International Conference on Image Processing*, , pp. 10-13,2000.

[21] Wactlar, H. D., Kanade, T., Smith, M. A., and. Stevens, S. M., "Intelligent Access to Digital Video: The Informedia Project", *IEEE Computer*, 29 (5) 46-52,1996.

[22] Rein-Lien, H., Abdel-Mottaleb, M., Jain, A. K., "Face Detection in Color Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (5) 696-706,2002.

[23] Strouthpoulos, C., Papamarkos, N., and Atsalakis, A.E., "Text Extraction in Complex Color Document", *Pattern Recognition*, 35 (8) 1743-1758,2002.

[24] Antani, S., Crandall, D.,. Narasimhamurthy, A., Mariano, V. Y., and Kasturi, R., "Evaluation of Methods for Detection and Localization of Text in Video", *Proc. of the IAPR workshop on Document Analysis Systems*, Rio de Janeiro, pp. 506-514,2000.

[25] Antani, S., "Reliable Extraction of Text From Video", *PhD thesis, Pennsylvania State University*, 2001.

[26] Jain, A. K., and Zhong, Y., "Page Segmentation using Texture Analysis", *Pattern Recognition*, 29 (5) 743-770,1996.

[27]. Tang, Y. Y., Lee, S. W., and Suen, C. Y., "Automatic Document Processing: A Survey", *Pattern Recognition*, 29 (12) 1931-1952,1996.

[28] Jung, K., Kim, K.I., Jain, A.K.: "Text information extraction in images and video: a survey". *Patt. Recognit.* **37**(5), 977–997,2004.

[29] Audithan, S., Chandrasekaran, RM., "Document Text Extraction from Document Images Using Haar Discrete Wavelet Transform", *European Journal of Scientific Research ISSN 1450-216X Vol.36 No.4 (2009)*, pp.502-512© EuroJournals Publishing, Inc. 2009

[30] Smith, M.A., Kanade, T.: "Video skimming for quick browsing based on audio and image characterization", *Technical Report CMU-CS-95–186, Carnegei Mellon University*, 1995

[31] Sato, T., Kanade, T., Hughes, E.K., Smith, M.A.: "Video OCR for digital news archive*", 1998 IEEE International Workshop on Content-Based Access of Image and Video Database*, pp. 52–60, Bombay India, 1998

[32] Lyu, M.R., Song, J., Cai, M.: "A comprehensive method for multilingual video text detection, localization, and extraction", *IEEE Trans. Circuits Syst. Video Technol.* **15**(2), 243–255 ,2005.

[33] Zhang, N., Tao, T., Satya, R.V., Mukheriee, A.: "Modified LZW algorithm for efficient compressed text retrieval", *In: Proceeding of International Conference on Information Technology, Coding and Computer*, pp. 224–228,2004.

[34] Hoogs, A., Mundy, J., Cross, G.: "Multi-modal fusion for video understanding", *In: Proceeding 30th Applied Imagery Pattern Recognition*, pp. 103–108 ,2001.

[35] Chen, D., Luettin, J., Shearer, K., "A survey of text detection and recognition in images and videos", *Institute Dalle Molled. Intelligence Perceptive (IDIAP) Research Report, IDIAP-RR* 00-38, 2000.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[36] Jung, K., Kim, K.I., and Jain, A.K., "Text Information extraction in images and video: a survey", *PatternRecognition*, pp. 977-997, 2004.

[37] Leon, M., Vilaplana, V., Gasull, A. and Marques, F., "Caption text extraction for indexing purposes using a hierarchical region-based image model," *IEEE ICIP 2009, El Cairo, Egypt*, 2009.

[38] Karin Sobottka, Horst Bunke and Heino Kronenberg,"Identification of Text on Colored Book and Journal Covers", *Document Analysis and Recognition*, 20-22 ,1999.

[39] Debapratim Sarkar, Raghunath Ghosh ,"A Bottom-Up Approach of Line Segmentation from Handwritten Text" ,2009

[40] Xiaoqing Liu and Jagath Samarabandu, " Multiscale Edge-Based Text Extraction From Complex Images" , *Multimedia and Expo, 2006 IEEE International Conference,2006.*

[41] Xin Zhang, Fuchun Sun, Lei Gu, "*A Combined Algorithm for Video Text Extraction", 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery,*2010.

[42] Jui-Chen Wu · Jun-Wei Hsieh · Yung-Sheng Chen,"Morphology-based text line extraction" *Machine Vision and Applications* 19:195–207 DOI 10.1007/s00138-007-0092-0,2008.

[43] Rama Mohan Babu, G., Srimaiyee, P. Srikrishna, A., "Text Extraction From Hetrogenous Images Using Mathematical Morphology" *Journal Of Theoretical And Applied Information Technology* © 2005 - 2010 Jatit

[44] Chu Duc Nguyen, Mohsen Ardabilian and Liming Chen ,"Robust Car License Plate Localization using a Novel Texture Descriptor", *Advanced Video and Signal Based Surveillance, 2009*

[45] Kwang In Kim, Keechul Jung, And Jin Hyung Kim ,"Texture-Based Approach For Text Detection In Images Using Support Vector Machines And Continuously Adaptive Mean Shift Algorithm" , *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 25, No. 12, 2003

[46] *Bassem Bouaziz, Walid Mahdi, Mohsen Ardabilain, Abdelmajid Ben Hamadou,"*A New Approach For Texture Features Extraction: Application For Text Localization In Video Images**"** *IEEE, ICME 2006*

[47] Chowdhury, S. P., Dhar,S., Das,A.K., Chanda,B., McMenemy,K.," Robust Extraction of Text from Camera Images**" ,** *10th International Conference on Document Analysis and Recognition, 2009.*

[48] Yuming Wang, Naoki Tanaka, **"**Text String Extraction from Scene Image Based on Edge Feature and Morphology", *Document Analysis Systems, 2008.*